Предисловие

Понятие колмогоровской сложности (или, как ещё говорят, алгоритмической энтропии) появилось в 1960-е годы на стыке теории алгоритмов, теории информации и теории вероятностей.

Идея Колмогорова, опубликованная им в знаменитой статье 1965 года [24], состояла в том, чтобы измерять количество информации, заключённой в индивидуальных конечных объектах (а не в случайных величинах, как в шенноновской теории информации). Оказалось, что это возможно (хотя лишь с точностью до ограниченного слагаемого).

На несколько лет раньше сходные идеи высказывал Р. Соломонов (см. [73] и другие его статьи, ссылки на которые даны в [38]). Он исходил из несколько другой мотивировки, пытаясь ввести понятие «априорной вероятности» — с какими вероятностями мы ожидаем появления тех или иных объектов, если ничего не знаем о порождающем их процессе? Оказалось, что вопросы эти тесно связаны. В самом первом приближении можно сказать так: простые объекты имеют большую априорную вероятность, а сложные — малую. (К сожалению, работы Соломонова приобрели известность в основном лишь после того, как Колмогоров упомянул их в своей статье.)

В 1965 году американский математик Г. Чейтин (тогда только что окончивший школу) представил к публикации две своих статьи [4] и [5], напечатанные в 1966 и 1969 годах. Во второй из них он дал то же определение алгоритмической сложности, что и Колмогоров.

Основные свойства колмогоровской сложности были исследованы в 1970-е годы. Левин (ученик Колмогорова) и Шнорр установили связь понятия сложности с алгоритмическим определением случайности (предложенным учеником Колмогорова шведским математиком Мартин-Лёфом). При этом они ввели некоторый новый вариант понятия сложности (так называемую монотонную сложность). Идеи Соломонова об априорной вероятности привели к (введённому Левиным и позднее Чейтиным) понятию префиксной сложности, которое оказалось не только философски важным, но и технически удобным. Были найдены применения понятия сложности как в теории алгоритмов, так и в теории вероятностей.

Интерес к колмогоровской сложности (заслуженный — это, безусловно, одно из базисных понятий теории алгоритмов, имеющее фундаментальное значение) возрос в последнее десятилетие, когда появилась монография Ли и Витаньи [38] (первое издание в 1993 году). Без преувеличения можно сказать, что практически всё известное о колмогоровской сложности вошло в эту книгу (в основной текст или в упражнения). Книга включает в себя подробный рассказ об истории области, ссылки на первые публикации и так далее.

Наша книга представляет собой учебник, не претендующий на полноту материала и исторических ссылок (мы рекомендуем обращаться к книге Ли и Витаньи). Мы пытались отобрать наиболее важные (философски и технически) факты о колмогоровской сложности и доступно о них рассказать. За редкими исключениями, мы почти ничего не говорим об истории вопроса. Как и в любом учебнике, многие утверждения формулируются без указания авторства (и это ни в коей мере не означает, что мы пытаемся приписать их себе).

В следующем разделе («О чём эта книга?») мы пытаемся кратко очертить круг вопросов, относящихся к колмогоровской сложности. Грубо говоря, он предназначен для тех, кто хочет понять, что это за область и стоит ли знакомиться с нею подробно. Далее мы переходим к систематическому изложению.

К сожалению, с названиями, обозначениями и терминологией в этой области большая путаница (отчасти и по вине авторов этой книги). Само основное понятие называется алгоритмической сложностью, колмогоровской сложностью, а также алгоритмической (или колмогоровской) энтропией. Мы будем считать все эти слова синонимами. Хуже с обозначениями: в оригинальной статье Колмогорова сложность слова x обозначалась $K(x)$. В других работах (например, у Ли и Витаньи) $K(x)$ обозначает префиксную сложность, а для исходного понятия используется обозначение $C(x)$. Чтобы избежать путаницы, мы вообще не будем использовать обозначение $K(x)$, а писать $\mathit{KS}(x)$ для исходного варианта сложности и $\mathit{KP}(x)$ для префиксной сложности. Есть и другие варианты понятия сложности; монотонную сложность (в её наиболее естественной форме) мы будем обозначать $\mathit{KM}(x)$, а сложность разрешения — $\mathit{KR}(x)$. (При этом некоторая путаница остаётся — у Ли и Витаньи монотонная сложность называется $\mathit{Km}(x)$, а обозначение $\mathit{KM}(x)$ используется для логарифма априорной вероятности, которые мы обозначаем $\mathit{KA}(x)$. Надеемся, это не собьёт читателя.)

Есть и другие, менее существенные, различия в названиях и обозначениях; для удобства в конце книги приведён список используемых обозначений и предметный указатель.

[Благодарности. Михаилу Вялому, Сергею Сальникову и Михаилу Устинову]

В. А. Успенский, Н. К. Верещагин, А. Шень

Оглавление

О чём эта книга?

Что такое колмогоровская сложность?

«На пальцах» это проще всего объяснить так. Существуют программы, которые сжимают файлы (для экономии места в архиве). Возможно, вы встречались с ними: наиболее распространённые называются `zip`, `gzip`, `compress`, `rar`, `arj` (есть и другие).

Применив такую программу к некоторому файлу (с текстом, данными, программой), мы получаем его сжатую версию (которая, как правило, короче исходного файла). По ней можно восстановить исходный файл (с помощью парной программы «декомпрессии»; часто сжатие и восстановление объединены в одну программу).

Так вот, в первом приближении колмогоровскую сложность файла можно описать как длину его сжатой версии. Тем самым файл, имеющий регулярную структуру и хорошо сжимаемый, имеет малую колмогоровскую сложность (в сравнении с его длиной). Напротив, плохо сжимаемый файл имеет сложность, близкую к длине.

Это описание весьма приблизительно и нуждается в уточнениях — как технических, так и принципиальных. Начнём с технического: вместо файлов (последовательностей байтов) мы будем рассматривать двоичные слова (конечные последовательности битов, то есть нулей и единиц). Длиной такого слова называется число знаков (так что слово 1001, скажем, имеет длину 4, а пустое слово имеет длину 0).

Более существенны другие отличия:

- Программы сжатия нет — мы рассматриваем лишь программу восстановления. Точнее, назовём *декомпрессором* произвольный алгоритм (программу), который получает на вход двоичные слова и выдаёт на выход также двоичные слова. Если декомпрессор D на входе x даёт y, мы пишем $D(x) = y$ и говорим, что x является *описанием* y при данном D (относительно данного D). Декомпрессоры мы также будем называть *способами описания*.

- Мы не требуем, чтобы декомпрессор был определён на всех словах. При некоторых x вычисление $D(x)$ может зацикливаться (не останавливаться) и не давать результата. Мы также не накладываем никаких ограничений на время работы D: на некоторых входах программа D может дать ответ лишь после очень долгой работы.

Используя терминологию теории вычислимых функций, можно сказать, что способ описания есть вычислимая функция из Ξ в Ξ, где Ξ — множество всех двоичных слов. Напомним, что с каждым алгоритмом D, входами и выходами которого являются двоичные слова, мы связываем функцию d, которая определена на слове x тогда и только тогда, когда D даёт результат на слове x; при этом значение $d(x)$ равно результату применения D ко входу x. Функции d, получаемые таким образом из всевозможных алгоритмов D, называются *вычислимыми функциями из Ξ в Ξ*. Обычно используют одну и ту же букву для обозначения алгоритма и вычисляемой им функции, и вместо $d(x)$ пишут $D(x)$. Мы также будем следовать этой традиции в тех случаях, когда это не вызывает путаницы.

Пусть фиксирован некоторый способ описания (декомпрессор) D. Для данного слова x рассмотрим все его описания, то есть все слова y, для которых $D(y)$ определено и равно x. Длину самого короткого из них и называют *колмогоровской сложностью* слова x при данном способе описания D:

$$KS_D(x) = \min\{\, l(y) \mid D(y) = x\}.$$

Здесь и далее $l(y)$ обозначает длину слова y. Индекс D подчёркивает, что определение зависит от выбора способа D. Минимум пустого множества, как обычно, считается равным $+\infty$, так что $KS_D(x)$ бесконечно для тех x, которые не входят в область значений функции D (не могут быть результатами декомпрессии, не имеют описаний).

Это определение кажется бессодержательным, поскольку для разных D получаются разные определения, в том числе абсурдные. Например, если D нигде не определён, то KS_D всегда бесконечно. Если $D(y) = \Lambda$ (пустое слово) при всех y, то сложность пустого слова равна нулю (поскольку $D(\Lambda) = \Lambda$ и $l(\Lambda) = 0$), а сложность всех остальных слов бесконечна.

Более осмысленный пример: если программа-декомпрессор копирует вход на выход и $D(x) = x$ при всех x, то каждое слово имеет единственное описание (самого себя) и $KS_D(x) = l(x)$.

Естественно, для любого слова x можно подобрать способ описания D, при котором x имеет малую сложность. Достаточно положить $D(\Lambda) = x$, и тогда $KS_D(x)$ будет равно нулю. Можно также подобрать способ описания, благоприятствующий целому классу слов: например, для слов из одних нулей хорош такой способ описания:

$$D(\mathrm{bin}(n)) = 000\ldots000 \quad (n \text{ нулей}),$$

где $\mathrm{bin}(n)$ — двоичная запись числа n. Легко проверить, что длина слова $\mathrm{bin}(n)$ примерно равна $\log_2 n$ (не превосходит $\log_2 n + 1$). Мы получаем, что для построенного способа описания сложность слова из n нулей близка к $\log_2 n$ (и много меньше длины слова, то есть n). Зато все другие слова (содержащие хотя бы одну единицу) имеют бесконечную сложность.

На первый взгляд кажется, что определение сложности настолько сильно зависит от выбора способа описания, что никакая общая теория невозможна.

Оптимальные способы описания

Способ описания тем лучше, чем короче описания. Поэтому естественно дать такое определение: способ D_1 *не хуже* способа D_2, если

$$KS_{D_1}(x) \leqslant KS_{D_2}(x) + c$$

при некотором c и при всех x.

В этом определении требует пояснения константа c. Мы соглашаемся пренебрегать увеличением сложности не более чем на константу. Конечно, можно сказать, что это лишает определение практического смысла, так как константа c может быть сколь угодно велика. Однако без такого «огрубления» обойтись не удаётся.

Пример. Пусть даны два произвольных способа описания D_1 и D_2. Покажем, что существует способ описания D, который не хуже их обоих.

В самом деле, положим

$$D(0y) = D_1(y)$$
$$D(1y) = D_2(y)$$

Другими словами, первый бит описания мы воспринимаем как номер способа описания, а остальную часть — как собственно описание. Ясно, что если y является описанием x при D_1 (или D_2), то $0y$ (соответственно $1y$) является описанием x при D, и это описание всего лишь на один бит длиннее. Поэтому

$$KS_D(x) \leqslant KS_{D_1}(x) + 1$$
$$KS_D(x) \leqslant KS_{D_2}(x) + 1$$

при всех x, так что способ D не хуже обоих способов D_1 и D_2.

Этот приём используется на практике. Например, формат zip-архива предусматривает заголовок, в котором указывается номер используемого метода сжатия, а затем идёт сам сжатый файл. При этом использование N различных способов описания приводит к тому, что начальные $\log_2 N$ битов (или около того) приходится зарезервировать для указания используемого способа.

Несколько обобщив эту же идею, можно доказать такую теорему:

Theorem 1 (Соломонова – Колмогорова). [intro-universal] *Существует способ описания D, который не хуже любого другого: для всякого способа описания D' найдётся такая константа c, что*

$$KS_D(x) \leqslant KS_{D'}(x) + c$$

для любого слова x.

Способ описания, обладающий указанным в теореме свойством, называют *оптимальным*.

$\lhd$ Напомним, что по нашему определению способами описания являются вычислимые функции. Программы, их вычисляющие, можно считать двоичными словами. Будем при этом предполагать, что по программе можно определить, где она кончается (такой способ записи по-английски называется self-delimiting; подходящего русского слова, пожалуй, нет, и мы будем называть такие программы самоограниченными). Если выбранный способ записи программ не таков, то его можно модифицировать. Например, можно продублировать каждый знак в записи программы (заменив 0 на 00 и 1 на 11), а в конце программы добавить группу 01.

Теперь определим новый способ описания D, положив

$$D(py) = p(y),$$

где p — произвольная программа (при выбранном способе записи), а y — любое двоичное слово. Другими словами, D читает слово слева направо, выделяя из него начало-программу. (Если таковой не оказывается, D делает что угодно, например, зацикливается.) Далее D применяет найденную программу (p) к остатку входа (y) и выдаёт полученный результат.

Покажем, что D не хуже любого другого способа описания P. Пусть p — программа, вычисляющая функцию P, причём записанная в выбранной нами форме. Если слово y является кратчайшим описанием слова x относительно P, то py будет описанием x относительно D (не обязательно кратчайшим). Поэтому при переходе от P к D длина описания увеличится не более чем на $l(p)$, и

$$KS_D(x) \leqslant KS_P(x) + l(p).$$

Видно, что константа (то есть $l(p)$) зависит лишь от выбора способа описания P, но не от x. $\triangleright$

По существу здесь используется тот же приём, что и в предыдущем примере, только вместо двух способов описания соединяются все мыслимые способы — каждый со своим префиксом. Именно так устроены так называемые «саморазархивирующиеся» архивы (self-extracting archives). Такой архив представляет собой исполняемый файл, в котором, однако, собственно программа занимает лишь небольшой начальный кусок. Эта программа загружается в память, после чего читает и разархивирует остаток архива.

Заметим, что построенный нами оптимальный способ описания на некоторых входах работает очень долго (поскольку бывают долго работающие программы), а на некоторых входах и вовсе не определён.

Колмогоровская сложность

Фиксируем некоторый оптимальный способ описания D и будем называть *колмогоровской сложностью* слова x значение $KS_D(x)$. В обозначении $KS_D(x)$ будем опускать индекс D и писать просто $KS(x)$.

Замена оптимального способа на другой оптимальный способ приводит к изменению сложности не более чем на константу: для любых оптимальных способов D_1 и D_2 найдётся такая константа $c(D_1, D_2)$, что

$$|KS_{D_1}(x) - KS_{D_2}(x)| \leqslant c(D_1, D_2)$$

при всех x. Это неравенство записывают как

$$KS_{D_1}(x) = KS_{D_2}(x) + O(1),$$

понимая под $O(1)$ произвольную ограниченную функцию от x.

Имеет ли смысл говорить о колмогоровской сложности конкретного слова x согласно этому определению, не указывая, какой оптимальный способ мы используем? Нет — легко понять, что подбором подходящего оптимального способа можно сделать сложность данного слова любой. Точно так же нельзя говорить, что слово x проще слова y (имеет меньшую сложность): выбирая тот или иной оптимальный способ, можно сделать любое из двух слов более простым.

Каков же тогда смысл колмогоровской сложности, если ни про какое конкретное слово ничего нельзя сказать?

Свойства оптимального способа, построенного при доказательстве теоремы Соломонова – Колмогорова, зависят от использованного способа записи программ (то есть от выбора языка программирования). Два таких способа приводят к сложностям, отличающимся не более чем на константу, которая есть длина интерпретатора одного языка программирования, написанного на другом языке. Можно надеяться, что при естественном выборе языков эта константа будет измеряться тысячами или даже сотнями. Тем самым, если мы говорим о сложностях порядка сотен тысяч (скажем, для текста романа) или миллионов (скажем, для ДНК), то уже не так важно, какой именно язык программирования мы выбрали.

Но тем не менее надо всегда помнить, что все утверждения о колмогоровской сложности носят асимптотический характер. Чтобы утверждение имело смысл, в нём должна идти речь о сложности не изолированного слова, а последовательности слов. Для теории сложности вычислений такого рода ситуация типична: скажем, оценки на сложность решения какой-либо задачи обычно также имеют асимптотический характер.

Сложность и информация

Колмогоровскую сложность слова x можно назвать также *количеством информации* в слове x. В самом деле, слово из одних нулей, которое может быть описано коротко, содержит мало информации, а какое-то сложное слово, которое не поддаётся сжатию, содержит много информации (пусть даже и бесполезной — мы не пытаемся отделить осмысленную информацию от бессмысленной, так что любая абракадабра для нас содержит много информации, если её нельзя задать коротко).

Если слово x имеет сложность k, мы говорим, что x содержит k битов информации. Естественно ожидать, что число битов информации в слове не превосходит его длины, то есть что $KS(x) \leqslant l(x)$. Так и оказывается (но только надо добавить константу, о чём мы уже говорили).

Theorem 2. [intro-length] *Существует такая константа c, что*

$$KS(x) \leqslant l(x) + c$$

для любого слова x.

$\lhd$ Мы уже говорили, что если $P(y) = y$ при всех y, то $KS_P(x) = l(x)$. В силу оптимальности найдётся такое c, что

$$KS(x) \leqslant KS_P(x) + c = l(x) + c$$

при всех x. $\rhd$

Утверждение этой теоремы обычно записывают так: $KS(x) \leqslant l(x) + O(1)$. Из него вытекает, в частности, что колмогоровская сложность любого слова конечна (то есть что любое слово имеет описание).

Вот ещё один пример свойства, которого естественно ожидать от понятия «количество информации»: при алгоритмических преобразованиях количество информации не увеличивается (точнее, увеличивается не более чем на константу, зависящую от алгоритма).

Theorem 3. [intro-transform] *Для любого алгоритма A существует такая константа c, что*

$$KS(A(x)) \leqslant KS(x) + c$$

для всех x, при которых $A(x)$ определено.

$\triangleleft$ Пусть D — оптимальный декомпрессор, используемый при определении колмогоровской сложности. Рассмотрим другой декомпрессор D':

$$D'(p) = A(D(p))$$

(D' применяет сначала D, а затем A). Если p является описанием некоторого x относительно D, причём $A(x)$ определено, то p является описанием $A(x)$ относительно D'. Взяв в качестве p кратчайшее описание x относительно D, находим, что

$$\mathit{KS}_{D'}(A(x)) \leqslant l(p) = \mathit{KS}_D(x) = \mathit{KS}(x),$$

а в силу оптимальности

$$\mathit{KS}(A(x)) \leqslant \mathit{KS}_{D'}(A(x)) + c \leqslant \mathit{KS}(x) + c$$

при некотором c и при всех x, что и требовалось доказать. $\triangleright$

Эта теорема гарантирует, что количество информации «не зависит от кодировки» — если мы, скажем, заменим в каком-то слове все нули на единицы и наоборот (или разбавим это слово нулями, добавив после каждой цифры по нулю), то полученное слово будет иметь ту же сложность, что и исходное (с точностью до константы), поскольку преобразования в ту и другую сторону выполняются некоторым алгоритмом.

Пусть x и y — два слова. Соединим их в одно, приписав y к x справа. Сколько битов информации будет иметь полученное слово? Естественно ожидать, что количество информации в нём не превосходит суммы количеств информации в x и в y. И действительно это так, правда, с некоторой поправкой.

Theorem 4. `[intro-pair]` *Существует такая константа c, что при любых x и y выполнено неравенство*

$$\mathit{KS}(xy) \leqslant \mathit{KS}(x) + 2\log \mathit{KS}(x) + \mathit{KS}(y) + c$$

$\triangleleft$ Попробуем для начала доказать теорему без добавочного члена $2\log \mathit{KS}(x)$, её ослабляющего. Это естественно делать так. Пусть D — оптимальный способ описания, используемый при определении колмогоровской сложности. Рассмотрим новый способ описания D'. Именно, если $D(p) = x$ и $D(q) = y$, будем считать pq описанием слова xy, то есть положим $D'(pq) = xy$. Тогда сложность слова xy относительно D' не превосходит длины слова pq, то есть $l(p) + l(q)$. Если в качестве p и q взять кратчайшие описания, то получится $\mathit{KS}_{D'}(xy) \leqslant \mathit{KS}_D(x) + \mathit{KS}_D(y)$, и остаётся воспользоваться оптимальностью D и перейти от D' к D (при этом возникает константа c).

Что неверно в этом рассуждении? Дело в том, что определение D' некорректно: мы полагаем $D'(pq) = D(p)D(q)$, но D' не имеет средств разделить p и q. Вполне может оказаться, что есть два разбиения слова на части, дающие разные результаты: $p_1q_1 = p_2q_2$, но $D(p_1)D(q_1) \neq D(p_2)D(q_2)$.

Есть два способа исправить эту ошибку. Первый, который мы и применим, состоит в том, чтобы перед словом p написать длину слова p в двоичной записи. При этом в этой двоичной записи мы удвоим каждый бит, а после неё напишем 01, чтобы алгоритм мог понять, где кончается двоичная запись и начинается само p. Более точно, обозначим через

$\mathrm{bin}(k)$ двоичную запись числа k, а через $\overline{x}$ результат удвоения каждого бита в слове x. (Например, $\mathrm{bin}(5) = 101$, а $\overline{\mathrm{bin}(5)} = 110011$.) Теперь положим

$$D'(\,\overline{\mathrm{bin}(l(p))}\,01pq) = D(p)D(q).$$

Это определение корректно: алгоритм D' сначала читает $\overline{\mathrm{bin}(l(p))}$, пока цифры идут парами. Когда появляется группа 01, он определяет $l(p)$, затем отсчитывает $l(p)$ цифр и получает p. Остаток входа есть q, после чего уже можно вычислить $D(p)D(q)$.

Величина $\mathit{KS}_{D'}(xy)$ оценивается теперь числом $2l(\mathrm{bin}(l(p))) + 2 + l(p) + l(q)$; двоичная запись числа $l(p)$ имеет длину не больше $\log_2 l(p) + 1$, поэтому D'-описание слова xy имеет длину не более $2\log_2 l(p) + 4 + l(p) + l(q)$, откуда и вытекает утверждение теоремы. $\triangleright$

Упомянутый нами второй способ исправления допущенной ошибки состоит в том, чтобы модифицировать определение сложности, потребовав, чтобы описания были самоограниченными (self-delimiting); мы обсуждаем его в главе 4.

Заметим также, что в теореме можно поменять местами p и q и доказать, что $\mathit{KS}(xy) \leqslant \mathit{KS}(x) + \mathit{KS}(y) + 2\log_2 \mathit{KS}(y) + c$.

Насколько неравенство теоремы 4 близко к равенству? Может ли $\mathit{KS}(xy)$ быть существенно меньше суммы $\mathit{KS}(x)+\mathit{KS}(y)$? В полном согласии с нашей интуицией это возможно, если x и y содержат много общего. Например, при $x = y$ мы имеем $\mathit{KS}(xy) = \mathit{KS}(xx) = \mathit{KS}(x) + O(1)$, поскольку xx алгоритмически получается из x и обратно (теорема 3).

Чтобы уточнить это наблюдение, мы определим понятие количества информации, которая содержится в x, но не содержится в y (для произвольных слов x и y). Эту величину называют *условной колмогоровской сложностью* x *при условии* y (говорят также «при известном y», «относительно y»). Её определение аналогично определению обычной (безусловной) сложности, но декомпрессор D имеет доступ не только к сжатому описанию, но и к слову y. Мы будем говорить об этом подробнее позже (глава 2), скажем лишь, что для условной сложности справедливо равенство

$$\mathit{KS}(xy) = \mathit{KS}(y) + \mathit{KS}(x|y) + O(\log n),$$

если x и y — слова сложности не более n. Читается оно так: количество информации в xy равно количеству информации в y плюс количество новой (отсутствующей в y) информации в x.

Разность $\mathit{KS}(x) - \mathit{KS}(x|y)$ естественно назвать количеством информации об x в y: эта разность показывает, насколько знание слова y упрощает описание слова x.

Понятие условной сложности позволяет придать смысл вопросу о том, сколько новой информации в ДНК одного организма по сравнению с ДНК другого: если d_1 — двоичное слово, кодирующее ДНК первого организма, а d_2 — двоичное слово, кодирующее ДНК второго, то искомая величина есть $\mathit{KS}(d_1|d_2)$. Аналогичным образом можно спрашивать, какой процент информации был потерян при переводе романа на другой язык: в этом случае нас интересует отношение

$$\mathit{KS}(\text{оригинал}|\text{перевод})/\,\mathit{KS}(\text{оригинал}).$$

Вопросы о количестве информации в различных объектах изучались и до алгоритмической теории информации, с помощью понятия шенноновской энтропии. Напомним её определение. Пусть случайная величина ξ принимает n значений с вероятностями $p_1, \ldots, p_n$.

Тогда её шенноновская энтропия определяется формулой

$$H(\xi) = \sum p_i(-\log_2 p_i).$$

Говорят, что исход, имеющий вероятность p_i, несёт в себе $(-\log_2 p_i)$ битов информации; тогда $H(\xi)$ можно интерпретировать как среднее количество информации в исходе случайной величины.

Чтобы применить понятие шенноновской энтропии, скажем, к оценке количества информации в данном тексте на русском языке, нужно включить этот текст в какой-то ансамбль текстов, считая его «типичным» значением некоторой случайной величины. Это имеет смысл для короткой поздравительной телеграммы, но для текста романа трудно вообразить себе подходящий ансамбль. Столь же сложно сделать это при оценке количества информации в ДНК: если считать ансамблем множество всех существовавших ДНК всех организмов, то число элементов этого ансамбля можно оценить сверху, скажем, числом 2^{1000}, и если считать их равновероятными (а какое ещё распределение мы можем выбрать?), получится абсурдно малое число — меньше тысячи битов.

Видно, что в этих ситуациях колмогоровская сложность является более адекватным языком, чем шенноновская энтропия.

Сложность и случайность

Вернёмся к неравенству $KS(x) \leqslant l(x) + O(1)$ (теорема 2). Для большинства слов данной длины это неравенство близко к равенству. В самом деле, справедливо такое утверждение:

Theorem 5. [intro-cardinality] *Пусть n — произвольное число. Тогда существует менее 2^n слов x, для которых $KS(x) < n$.*

$\lhd$ Пусть D — оптимальный способ описания, фиксированный при определении колмогоровской сложности. Тогда все слова вида $D(y)$ при $l(y) < n$ (и только они) имеют сложность менее n. А таких слов не больше, чем различных слов y, имеющих длину меньше n, которых имеется

$$1 + 2 + 4 + 8 + \ldots + 2^{n-1} = 2^n - 1$$

штук (слов длины k ровно 2^k). $\rhd$

Отсюда легко заключить, что доля слов сложности меньше $n - c$ среди всех слов длины n меньше $2^{n-c}/2^n = 2^{-c}$. Например, доля слов сложности менее 90 среди всех слов длины 100 меньше 2^{-10}.

Таким образом, большинство слов несжимаемы или почти несжимаемы. Это можно выразить и так: почти наверняка случайно взятое слово данной длины окажется (почти) несжимаемым. Рассмотрим следующий мысленный (или даже реальный) эксперимент. Бросим монету, скажем, 80000 раз и сделаем из результатов бросаний файл длиной в 10000 байтов (8 битов образуют один байт). Можно смело держать пари, что ни один существовавший до момента бросания архиватор не сумеет сжать этот файл более чем на 10 байтов. (В самом деле, вероятность этого меньше 2^{-80} для каждого конкретного архиватора, а число различных архиваторов не так уж велико.)

Вообще естественно считать случайными несжимаемые слова: случайность есть отсутствие закономерностей, которые позволяют сжать слово. Конечно, чёткой границы между случайными и неслучайными объектами провести нельзя. Глупо интересоваться, какие именно из восьми слов длины 3 (то есть из слов $000, 001, 010, 011, 100, 101, 110, 111$) случайны, а какие нет. Другой пример: начав со «случайного» слова длиной 10000, будем заменять в нём по очереди единицы на нули. В конце получится явно неслучайное слово (из одних нулей), но имеет ли смысл спрашивать, в какой именно момент слово перестало быть случайным? Вряд ли.

Естественно определить *дефект случайности* двоичного слова x как разность $l(x) - KS(x)$. Используя эту терминологию, можно сказать так: теорема 2 утверждает, что дефект случайности почти что неотрицателен (не меньше некоторой константы), а теорема 5 гарантирует, что для случайно взятого слова данной длины n (мы считаем все слова длины n равновероятными) вероятность иметь дефект больше d оценивается сверху числом $1/2^d$.

Теперь закон больших чисел (утверждающий, что у большинства двоичных слов данной длины n доля единиц близка к $1/2$) можно попытаться перевести на язык колмогоровской сложности так: у любого слова с малым дефектом случайности частота единиц близка к $1/2$. Из этого «перевода» вытекает исходная формулировка закона (поскольку мы уже знаем, что слова с малым дефектом случайности составляют большинство); как мы впоследствии убедимся, в некотором смысле эти формулировки равносильны.

Если мы хотим провести чёткую границу между случайными и неслучайными объектами, мы должны от конечных объектов перейти к бесконечным. Определение случайности для бесконечных последовательностей нулей и единиц, данное учеником Колмогорова шведским математиком Мартин-Лёфом, подробно обсуждается нами в главе 3. Впоследствии Левин и Шнорр нашли критерий случайности в терминах сложности: последовательность случайна тогда и только тогда, когда дефект случайности её начальных отрезков ограничен. (Правда, нужно использовать другой вариант колмогоровской сложности, так называемую *монотонную* сложность.)

Невычислимость KS и парадокс Берри

Прежде чем говорить о применениях колмогоровской сложности, скажем о препятствии, с которым сталкивается любое применение. Увы, функция KS невычислима: не существует алгоритма, который по данному слову x определяет его колмогоровскую сложность. Более того, не существует никакой нетривиальной (неограниченной) вычислимой нижней оценки для KS, как показывает следующая теорема.

Theorem 6. `[intro-nobound]` *Пусть k — произвольная (не обязательно всюду определённая) вычислимая функция из Ξ в $\mathbb{N}$. (Другими словами, вычисляющий её алгоритм применим к некоторым двоичным словам и даёт в качестве результатов натуральные числа.) Если k является нижней оценкой для колмогоровской сложности (то есть $k(x) \leqslant KS(x)$ для тех x, для которых $k(x)$ определено), то k ограничена: все её значения не превосходят некоторой константы.*

$\triangleleft$ Доказательство этой теоремы повторяет так называемый «парадокс Берри». Этот парадокс состоит в предложении рассмотреть

> *наименьшее число, которое нельзя определить фразой из не более чем тринадцати русских слов.*

Эта фраза как раз содержит тринадцать слов и определяет то самое число, которое нельзя определить, так что получается противоречие.

Следуя этой идее, мы можем искать *первое попавшееся двоичное слово, колмогоровская сложность которого больше некоторого* N. С одной стороны, это слово по построению будет иметь сложность больше N, с другой стороны, приведённое его описание будет коротким (оно включает в себя число N, но двоичная запись числа N гораздо короче самого́ N). Но как искать это слово? Для этого-то и нужна вычислимая нижняя оценка колмогоровской сложности.

Перейдём к формальному изложению доказательства. По условию функция k является вычислимой нижней оценкой колмогоровской сложности. Рассмотрим функцию $B(N)$, аргументом которой является натуральное число N, вычисляемую следующим алгоритмом: «Развернуть параллельно вычисления $k(0), k(1), k(2) \dots$ и проводить их до тех пор, пока не обнаружится некоторое x, для которого $k(x) > N$. Первое из таких x и будет результатом».

Если функция k ограничена, то теорема доказана. В противном случае функция B определена для всех N, и $k(B(N)) > N$ по построению. По предположению k является нижней оценкой сложности, так что и $\mathit{KS}(B(N)) > N$. С другой стороны, $B(N)$ эффективно получается по двоичной записи $\mathrm{bin}(N)$ числа N, поэтому

$$\mathit{KS}(B(N)) \leqslant \mathit{KS}(\mathrm{bin}(N)) + O(1) \leqslant l(\mathrm{bin}(N)) + O(1) \leqslant \log_2 N + O(1)$$

(первое неравенство следует из теоремы 3, а второе — из теоремы 2; $O(1)$ обозначает ограниченное слагаемое). Получается, что

$$N < \mathit{KS}(B(N)) \leqslant \log_2 N + O(1),$$

что при больши́х N приводит к противоречию. $\triangleright$

Применения колмогоровской сложности

Прежде всего оговоримся: речь пойдёт не о практических применениях, а о связи колмогоровской сложность с другими вопросами.

Бритва Оккама. Начнём с такого философского вопроса: что значит, что теория хорошо объясняет результаты эксперимента? Пусть имеется какой-то набор экспериментальных данных и разные теории, его объясняющие. Например, экспериментальными данными могут быть положения планет на небесной сфере. Их можно объяснять в духе Птолемея, рассматривая эпициклы и дифференты (и внося дополнительные поправки при необходимости), а можно ссылаться на законы современной механики. Как объяснить, чем современная теория лучше? Можно сказать так: современная теория позволяет вычислять положения планет с той же (и даже лучшей) точностью, имея меньше параметров. Условно говоря, достижение Кеплера состоит в том, что он нашёл более короткое описание для экспериментальных данных. Совсем грубо можно сказать, что экспериментаторы получают двоичные слова, после чего теоретики ищут для этих слов короткие описания (и тем самым верхние оценки на сложность), и лучше тот теоретик, у которого описание короче (оценка меньше).

Этот подход называют иногда «бритвой Оккама» по имени философа, который говорил, что не следует множить сущности без необходимости. Насколько такое толкование одобрил бы сам Оккам, сказать трудно.

Можно использовать ту же идею в более практической ситуации. Представим себе, что мы собираемся автоматически читать надписи на конвертах и ищем правило, отделяющее изображения нулей от изображений единиц. (Будем считать, что изображение дано в виде матрицы битов, записанной как двоичное слово.) У нас есть несколько тысяч образцов — картинок, про которые мы знаем, нуль это или единица. По этим образцам мы хотим сформулировать какое-то разумное разделяющее правило. Что означает в этом контексте слово «разумное»? Если мы просто перечислим все образцы того или другого типа, получится вполне действующее правило — действующее до тех пор, пока нам не принесут новый образец, — но оно будет слишком длинным. Естественно считать, что разумное правило должно иметь короткое описание, то есть по возможности меньшую колмогоровскую сложность.

Обоснование теории вероятностей. Речь идёт не о само́й математической теории вероятностей (которую можно рассматривать как часть теории меры и которая не требует никакого особого обоснования), а о механизме применения теории вероятностей. Представим себе, что мы бросаем монету тысячу раз (или, более реалистично, проверяем качество специального электронного датчика случайных чисел, дающего последовательность нулей и единиц). Если получится последовательность из тысячи нулей (или последовательность $0101010101\ldots$), то мы сочтём, что датчик плох. Но, собственно говоря, почему?

Часто говорят, что вероятность случайного появления тысячи нулей исчезающе мала (равна 2^{-1000}), если монета честная, и потому гипотеза честной монеты отвергается. С другой стороны, мы же не всегда бракуем датчик: возможна такая последовательность α из тысячи нулей и единиц, которая не вызовет у нас претензий к датчику. Заметим теперь, что вероятность появления последовательности α при бросании честной монеты также равна 2^{-1000} — так почему же мы в этом случае не возмущаемся? В чём разница между последовательностью из тысячи нулей и последовательностью α? Естественно усмотреть эту разницу в том, что последовательность из тысячи нулей имеет меньшую сложность, чем α.

Доказательство теорем теории вероятностей. Рассмотрим в качестве примера усиленный закон больших чисел. Он утверждает, что почти все бесконечные последовательности нулей и единиц (по равномерной бернуллиевой мере, соответствующей независимым бросаниям симметричной монеты) имеют предел частоты единиц, равный $1/2$.

Более подробно. Обозначим через Ω множество всех бесконечных последовательностей нулей и единиц. На этом множестве вводится равномерная бернуллиева мера. Это делается так: для каждого двоичного слова x рассмотрим множество Ω_x всех бесконечных продолжений этого слова. (Например, $\Omega_\Lambda = \Omega$.) Меру множества Ω_x положим равной $2^{-l(x)}$. (Например, мера множества Ω_{01}, состоящего из последовательностей, начинающихся на 01, равна $1/4$.)

Для каждой последовательности $\omega = \omega_0\omega_1\omega_2\ldots$ рассмотрим предел частоты единиц в начальных отрезках, то есть предел

$$\lim_{n\to\infty} \frac{\omega_0 + \omega_1 + \ldots + \omega_{n-1}}{n}.$$

Говорят, что последовательность удовлетворяет усиленному закону больших чисел, если этот предел существует и равен $1/2$. Например, последовательность $010101\ldots$ (с периодом 2) ему удовлетворяет, а последовательность $011011011\ldots$ (с периодом 3) — нет.

Усиленный закон больших чисел гласит, что множество последовательностей, не удовлетворяющих закону больших чисел, имеет меру нуль. Напомним: множество $A \subset \Omega$ имеет меру нуль, если для всякого $\varepsilon > 0$ можно указать последовательность слов $x_0, x_1, x_2, \ldots$, для которой

$$A \subset \Omega_{x_0} \cup \Omega_{x_1} \cup \Omega_{x_2} \cup \ldots$$

и сумма ряда

$$2^{-l(x_0)} + 2^{-l(x_1)} + 2^{-l(x_2)} + \ldots$$

(состоящего из мер множеств Ω_{x_i}) не превосходит ε.

Усиленный закон больших чисел можно доказать с помощью алгоритмической теории информации, используя понятие случайной по Мартин-Лёфу последовательности (о котором мы уже упоминали). Доказательство состоит из двух частей. Сначала мы доказываем, что всякая случайная по Мартин-Лёфу последовательность удовлетворяет закону больших чисел. Это можно сделать с использованием критерия случайности Левина – Шнорра (если предел частот не существует или не равен $1/2$, то некоторые начальные отрезки имеют сложность меньше, чем должно быть для случайной последовательности).

Вторая часть доказательства носит общий характер и не зависит от того, какой закон теории вероятностей рассматривается. Мы доказываем, что множество неслучайных (по Мартин-Лёфу) последовательностей имеет меру нуль. После этого ясно, что множество последовательностей, не удовлетворяющих закону больших чисел, содержится в множестве меры нуль и потому само имеет меру нуль.

Понятие случайной последовательности имеет и самостоятельный философский интерес. В начале XX века Рихард фон Мизес предложил положить это понятие (он называл его по-немецки Kollektiv) в основу теории вероятностей (в то время ещё не было идеи рассматривать теорию вероятностей в рамках теории меры). Подход Мизеса к определению случайной последовательности (он считал главным свойством так называемую «устойчивость частот») и его современные варианты мы рассматриваем в главе 9.

Нижние оценки в теории вычислительной сложности. Колмогоровская сложность оказалась полезным техническим средством для доказательства нижних оценок в теории сложности вычислений. Мы опишем её применение на модельном примере.

Рассмотрим следующую задачу: на ленте одноленточной машины Тьюринга написано некоторое слово x длины n. Требуется скопировать это слово справа от него, то есть получить на ленте слово xx. С середины 1960-х годов известно, что это требует времени, пропорционального n^2. Точнее говоря, можно доказать, что для любой машины Тьюринга M, решающей эту задачу для всех слов x, найдётся такое число $\varepsilon > 0$, что для любого n существует слово x длины n, копирование которого (с помощью M) длится дольше εn^2.

Имеется следующее интуитивное объяснение этого факта. Машина Тьюринга имеет ограниченное число внутренних состояний, то есть может запомнить (помимо ленты) ограниченное число битов. Кроме того, скорость передвижения её по ленте ограничена: за один шаг она может сместиться не более чем на одну клетку. Следовательно, скорость переноса информации с помощью машины Тьюринга (единица измерения: *бит · клетка/шаг*) ограничена константой, зависящей от числа состояний машины (пропорциональной логарифму числа состояний). Копирование слова x длины n требует переноса n битов информации, содержащихся в слове x, на расстояние в n клеток, и потому число шагов пропорционально n^2.

Это рассуждение можно сделать вполне строгим, используя понятие колмогоровской сложности. Видно сразу же, что трудными для копирования будут слова, содержащие много информации (сложность которых близка к максимуму, то есть к n).

Мы рассмотрим этот пример подробнее в главе 8.

Комбинаторный смысл утверждений о сложности. Мы приведём один пример такого рода. Существует неравенство для сложностей трёх слов и их комбинаций: [intro-triple]

$$2\,\mathit{KS}(xyz) \leqslant \mathit{KS}(xy) + \mathit{KS}(xz) + \mathit{KS}(yz) + O(\log n)$$

для любых трёх слов x, y, z длины не больше n.

Оказывается, что это неравенство имеет естественные интерпретации, в которых о сложности уже ничего не говорится. В частности, из него можно вывести такой геометрический факт.

Пусть имеется тело в трёхмерном координатном пространстве с осями OX, OY и OZ, имеющее объём V. Рассмотрим три его ортогональные проекции на плоскости OXY, OXZ и OYZ. Пусть S_{xy}, S_{xz} и S_{yz} — площади этих проекций. Тогда имеет место такое неравенство:

$$V^2 \leqslant S_{xy} \cdot S_{xz} \cdot S_{yz}.$$

А вот алгебраическое следствие того же неравенства: для произвольной группы G и её подгрупп X, Y и Z справедливо неравенство:

$$|X \cap Y \cap Z|^2 \geqslant \frac{|X \cap Y| \cdot |X \cap Z| \cdot |Y \cap Z|}{|G|}$$

(где $|H|$ обозначает число элементов в группе H).

Теорема о неполноте. Покажем, следуя Г. Чейтину, как можно использовать теорему 6, чтобы доказать известную теорему Гёделя о неполноте. Эта теорема гласит, что не все истинные утверждения достаточно богатой математической теории (скажем, формальной арифметики или теории множеств) можно доказать в этой теории.

Будем считать, что в нашей теории можно записать утверждения вида $\mathit{KS}(x) > n$ для любого двоичного слова x и любого натурального числа n. (Утверждение $\mathit{KS}(x) > n$ гласит, что выбранный декомпрессор D не даёт x ни на каком входе длины не более n; его легко записать в формальной арифметике и тем более в теории множеств.)

Будем перебирать все формальные выводы и искать среди них доказательства утверждений вида $\mathit{KS}(x) > n$ (для конкретных слов x и чисел n). Найдя такое утверждение, мы смотрим, каково в нём значение n (чем оно больше, тем лучше). Если n превосходит прежний рекорд, то мы запоминаем в таблице рекордов это n и соответствующее ему значение x_n. Теперь есть две возможности. Либо таблица рекордов будет расти неограниченно, либо начиная с некоторого времени новых рекордов не будет и какое-то утверждение $\mathit{KS}(X) > N$ так и останется непревзойдённым. Во втором случае целый класс истинных утверждений, а именно, все истинные утверждения вида $\mathit{KS}(x) > n$ при $n > N$, являются недоказуемыми. (Напомним, что по теореме 5 такие истинные утверждения существуют, и их бесконечно много.)

В первом случае мы получим вычислимую последовательность слов $x_0, x_1, x_2 \ldots$ и чисел $n_0 < n_1 < n_2 < \ldots$, для которых доказуемы утверждения $\mathit{KS}(x_i) > n_i$. Мы предполагаем, что в теории доказуемы лишь истинные утверждения, так что действительно $KS(x_i) > n_i$. Выбросив из этой последовательности повторения слов, можно считать, что все x_i различны (каждое слово может встречаться лишь конечное число раз, так как $n_i \to \infty$ при $i \to \infty$). После этого вычислимая функция k, для которой $k(x_i) = n_i$, будет неограниченной нижней оценкой для колмогоровской сложности, что противоречит теореме 6.

1. Простая колмогоровская сложность

1.1. Определение и основные свойства

[simpledf]

Напомним (данное во введении) определение колмогоровской сложности. Этот вариант определения сложности был предложен в статье Колмогорова [24]. Чтобы отличить его от других, определённую таким образом сложность называют *простой*, или *обыкновенной*, колмогоровской сложностью. Другие виды сложности (префиксная, монотонная) будут рассматриваться, начиная с главы 4, так что пока мы можем говорить о колмогоровской сложности (без уточняющего эпитета), не опасаясь путаницы.

Способом описания, или *декомпрессором*, мы называли произвольное вычислимое частичное отображение D из множества двоичных слов Ξ в себя. (Вычислимость отображения D означает, что есть алгоритм, который применим к словам из области определения отображения D и только к ним; результат применения алгоритма к слову x есть $D(x)$.) Если $D(y) = x$, говорят, что y является *описанием* x *при способе описания* D.

Для каждого способа описания D мы определяем *сложность относительно этого способа описания*, полагая её равной длине кратчайшего описания:

$$KS_D(x) = \min\{l(y) | D(y) = x\}.$$

При этом минимум пустого множества считается равным $+\infty$.

Говорят, что способ описания D_1 *не хуже* способа описания D_2, если найдётся такая константа c, что $KS_{D_1}(x) \leqslant KS_{D_2}(x) + c$ для всех слов x. (Краткая запись: $KS_{D_1}(x) \leqslant KS_{D_2}(x) + O(1)$.)

Способ описания называют *оптимальным*, если он не хуже любого другого способа описания. Теорема Колмогорова – Соломонова (с. 9) утверждает, что существуют оптимальные способы описания. Её доказательство (подробно изложенное во введении) проходит так. Выберем какой-либо универсальный язык программирования. Пусть U — интерпретатор этого языка: $U(p, x)$ есть результат работы программы p на входе x (программа и вход — двоичные слова). Далее мы полагаем

$$D(\hat{p}x) = U(p, x),$$

где вычислимое отображение $p \mapsto \hat{p}$ выбрано так, чтобы по слову $\hat{p}$ можно было определить p, а также место, где $\hat{p}$ кончается. (В этом случае слово $\hat{p}$ не может быть началом слова $\hat{q}$ при $p \neq q$, и гарантирует, что D корректно определено.) Тогда для любого способа описания D' имеем

$$KS_{D'}(x) \leqslant KS_D(x) + l(\hat{p}),$$

где p — программа, соответствующая способу описания D'. (В самом деле, если y есть описание x относительно D', то $\hat{p}y$ есть описание x относительно D.)

Мы фиксируем некоторый оптимальный способ описания и сложность слова x относительно этого способа описания обозначаем $KS(x)$ (без индекса).

Сравнение оптимального способа описания с тождественным (при котором слово является своим собственным описанием) показывает, что $KS(x) \leqslant l(x) + O(1)$ (с. 11).

Сравнивая оптимальный способ описания D со способом $y \mapsto A(D(y))$, где A — произвольная вычислимая функция, находим, что $KS(A(x)) \leqslant KS(x) + O(1)$ (невозрастание сложности при алгоритмических преобразованиях, с. 11).

Последнее свойство позволяет говорить не только о сложности слов, но и о сложности других «конструктивных объектов» (натуральных чисел, графов, перестановок, конечных множеств слов и т. п.) — любых объектов, которые можно естественным образом закодировать двоичными словами.

Поясним сказанное на примере натуральных чисел. Натуральное число n можно записать в двоичной системе — получится двоичное слово. Можно выбрать и другой способ представления натуральных чисел двоичными словами. Например, можно записать двоичные слова подряд (в порядке возрастания длин и в лексикографическом порядке для слов одной длины):

$$\Lambda, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, \ldots$$

и сопоставить их (в этом порядке) с натуральными числами $0, 1, 2, 3, \ldots$ (такое отождествление часто удобно, поскольку, в отличие от двоичной записи, оно является взаимно однозначным). Наконец, натуральное число n можно изображать словом из n единиц. Так вот, какой бы из этих способов ни выбрать для определения сложности натуральных чисел, мы получим функции сложности, отличающиеся не более чем на константу. В самом деле, существуют алгоритмы, переводящие натуральные числа из одной записи в другую, а применение алгоритма увеличивает сложность не более чем на константу.

Заметим, что колмогоровская сложность и так определена с точностью до ограниченного слагаемого, так что дополнительный произвол такого рода нам не страшен.

Отметим ещё, что двоичная запись натурального числа n имеет длину $\log n + O(1)$, и потому его сложность также не превосходит $\log n + O(1)$.

Другое простое применение свойства невозрастания сложности при алгоритмических преобразованиях: покажем, что добавление или удаление нуля или единицы в конце слова меняет его сложность не более чем на константу. В самом деле, функции $x \mapsto x0$, $x \mapsto x1$, а также функция, удаляющая последний бит слова, вычислимы.

Разумеется, то же самое верно для добавления бита в начале слова. Но изменение бита в произвольном месте слова может изменить его сложность более чем на константу. Например, если слово состоит из 2^n нулей, то его сложность не больше $KS(n) + O(1) \leqslant \log n + O(1)$. (Под $\log$ здесь и далее мы понимаем двоичный логарифм.) Заменив k-й нуль на единицу (при $k = 1, 2, \ldots, 2^n$), мы получим 2^n различных слов, поэтому сложность хотя бы одного из них не меньше n. (Как мы видели на с. 14, число слов сложности меньше n не превосходит числа всех описаний длины меньше n и потому меньше 2^n.)

Прибавление единицы к натуральному числу n меняет значение $KS(n)$ не более чем на константу, так что функция сложности, рассматриваемая как функция натурального аргумента, обладает «свойством Липшица» (это значит, что $|KS(m) - KS(n)| \leqslant c|m - n|$ для некоторого c и любых натуральных чисел m и n).

$\boxed{1}$ Докажите более сильное неравенство: $|KS(m) - KS(n)| \leqslant |m - n| + c$ (для некоторого c и для всех $m, n \in \mathbb{N}$) и даже $|KS(m) - KS(n)| \leqslant 2\log|m - n| + c$ (при $m \neq n$).

Вернёмся к утверждению о том, что число слов x с $KS(x) < n$ меньше 2^n. Заметим, что в этом утверждении нет никаких констант (что несколько необычно). Но всё равно зависимость от выбора оптимального способа описания в нём неявно присутствует: если мы заменим один способ описания на другой, то количество слов сложности меньше n по-прежнему будет меньше 2^n, но это уже будут другие слова!

$\boxed{2}$ [distribution-plain-complexity] Покажите, что число слов сложности меньше n заключено между 2^{n-c} и 2^n (при некотором c и всех n). [Указание: о верхней оценке 2^n мы уже говорили, а нижняя вытекает из того, что $KS(x) \leqslant l(x) + c$ при некотором c и потому все слова длины меньше $n - c$ имеют сложность меньше n.]

Покажите, что число слов сложности ровно n не превосходит 2^n, но может быть сильно меньше (например, таких слов может не быть вовсе для бесконечно многих n). [Указание: модифицируем оптимальный способ описания, добавив 0 или 11 ко всем описаниям и сделав их длину чётной.]

$\boxed{3}$ [average-plain-complexity] Покажите, что среднее арифметическое сложностей всех слов длины n равно $n + O(1)$. [Указание: ряд $\sum k/2^k$ сходится, а доля слов длины n, сложность которых равна $n - k$, равна примерно 2^{-k}.]

Продолжая эту тему (связь оценок сложности с оценками количества), докажем такое важное утверждение:

Theorem 7. [simple-upper] (а) *Множества $S_n = \{x \mid KS(x) < n\}$ образуют перечислимое семейство конечных перечислимых множеств, причём $|S_n| < 2^n$ при всех n. (Здесь $|S_n|$ — число элементов в множестве S_n.)*

(б) *Если V_n ($n = 0, 1, 2, \ldots$) — перечислимое семейство конечных перечислимых множеств, причём $|V_n| < 2^n$ при всех n, то найдётся такое c, что $KS(x) < n + c$ для любого n и для любого $x \in V_n$.*

Поясним, что такое перечислимое семейство конечных перечислимых множеств. Говорят, что множество (слов, натуральных чисел или иных конструктивных объектов) *перечислимо*, если существует алгоритм, который порождает элементы этого множества (программа, печатающая их один из другим, возможно с перерывами, и никогда не останавливающаяся). Повторения разрешаются (но это не важно, так как можно вычёркивать повторяющиеся элементы.) Если множество конечно, программа может с некоторого момента ничего не печатать (хотя наблюдатель не узнает, наступил этот момент или нет).

Например, множество всех n, при которых в десятичном разложении $\sqrt{2}$ встречается ровно n девяток подряд, перечислимо. Перечисляющий его алгоритм таков: вычисляем последовательные цифры числа $\sqrt{2}$, как только встретилась группа из n девяток подряд, окружённая не-девятками, печатаем n и продолжаем работу.

Перечислимость семейства множеств V_n означает, что множество пар $\{\langle n, x\rangle \mid x \in V_n\}$ перечислимо. Это, вообще говоря, более сильное свойство, чем перечислимость каждого из множеств V_n. Если семейство перечислимо, то каждое из множеств перечислимо (отбираем пары с данным первым членом), но обратное, вообще говоря, неверно. Например, если все множества V_n конечны (как у нас), то все они перечислимы, но это не гарантирует перечислимости семейства. (Можно проверить, что перечислимость семейства равносильна возможности по n алгоритмически получать алгоритм, перечисляющий V_n.) О понятии перечислимого множества можно прочесть в любом учебнике по теории вычислимости, например, в [77].

$\lhd$ Вернёмся к доказательству теоремы. Покажем, что множество пар $\{\langle n, x\rangle \mid x \in S_n\} = \{\langle n, x\rangle \mid \textit{KS}(x) < n\}$ (первыми компонентами пар являются натуральные числа, вторыми — двоичные слова) перечислимо. В самом деле, пусть D — оптимальный способ описания, использованный при определении $\textit{KS}$. Будем параллельно запускать D на всех двоичных словах, делая всё больше шагов на каждом слове и вовлекая всё новые и новые слова. (Например, на k-м этапе мы проводим по k шагов вычисления для первых k двоичных слов.) Как только одно из проводимых вычислений заканчивается, и обнаруживается, что $D(y) = x$, перечисляющий алгоритм выдаёт на выход пару $\langle l(y) + 1, x\rangle$ (в самом деле, мы установили, что сложность слова x меньше $l(y) + 1$, поскольку это слово имеет описание y). После этого он выдаёт пары $\langle l(y)+2, x\rangle, \langle l(y)+3, x\rangle \ldots$ (чередуя их с другими парами, которые нужно выдать).

Для знакомых с теорией алгоритмов всё это объяснение можно заменить одной строчкой

$$\textit{KS}(x) < n \Leftrightarrow \exists y\,(l(y) < n \wedge D(y) = x)$$

(множество в правой части перечислимо, так как график вычислимой функции D перечислим, а пересечение и проекция сохраняют перечислимость).

Более содержательно обратное утверждение. Пусть имеется перечислимое семейство конечных перечислимых множеств V_n, причём $|V_n| < 2^n$. Построим способ описания D_V следующим образом. Зарезервируем слова длины n для описания элементов множества V_n. Именно, будем считать k-е в лексикографическом порядке слово длины n описанием k-го элемента множества V_n (в порядке появления элементов множества V_n при перечислении). Поскольку $|V_n| < 2^n$, слов длины n хватит (естественно, при повторном появлении слова в перечислении мы не выделяем нового описания). Построенный таким образом способ описания D_V будет вычислимым. В самом деле, чтобы вычислить $D_V(y)$, мы находим порядковый номер слова y среди всех слов длины $l(y)$. Пусть он равен k. После этого мы запускаем алгоритм, перечисляющий множество пар $\langle n, x\rangle$, у которых $x \in V_n$, и ожидаем появления k различных пар, первая компонента которых равна $l(y)$. Вторая компонента последней из них и будет $D_V(y)$.

По определению $\textit{KS}_{D_V}(x) \leqslant n$ для любого $x \in V_n$. Сравнивая способ описания D_V с оптимальным, находим, что найдётся такое c, что $\textit{KS}(x) < n + c$ для любого $x \in V_n$. Теорема 7 доказана. $\rhd$

Интуитивный смысл доказанной только что теоремы можно объяснить так: она говорит, что утверждения «объектов определённого вида мало» (меньше 2^i) и «объекты этого вида просты» (имеют сложность меньше i) равносильны, если мы рассматриваем перечислимые семейства и измеряем сложность с точностью до аддитивной константы (а число элементов — с точностью до мультипликативной).

Эту же теорему можно сформулировать в других терминах. Пусть функция $f(x)$ определена на всех двоичных словах и принимает в качестве значений натуральные числа, а также специальный символ $+\infty$. Функция f называется *перечислимой сверху*, если существуют вычислимая функция $\langle x, k\rangle \mapsto F(x, k)$, определённая для всех слов x и всех натуральных чисел k, для которой

$$F(x, 0) \geqslant F(x, 1) \geqslant F(x, 2) \geqslant \ldots$$

и

$$f(x) = \lim_{k\to\infty} F(x, k).$$

при всех x. Значениями функции F также могут быть натуральные числа и $+\infty$. Наши требования гарантируют, что при любом k значение $F(x,k)$ является верхней оценкой для $f(x)$. Эта оценка уточняется с ростом k. При каждом x эта оценка в какой-то момент становится точной, но когда именно, мы можем не знать (если есть алгоритм, указывающий этот момент, то функция f вычислима).

Всякая вычислимая функция перечислима сверху.

Несложно проверить, что функция f перечислима сверху тогда и только тогда, когда множество

$$G_f = \{\langle x, n\rangle \mid f(x) < n\},$$

называемое иногда «надграфиком» функции f, перечислимо. (Это объясняет несколько странное название «перечислимая сверху».)

Убедимся в этом. Если функция f перечислима сверху и F — соответствующая функция двух аргументов, то

$$f(x) < n \Leftrightarrow \exists k\, F(x,k) < n.$$

Поэтому, вычисляя $F(x,k)$ параллельно для всех x и k, мы можем перечислять множество G_f. Обратно, если мы можем перечислять множество G_f, то $F(x,k)$ можно определить как наилучшую верхнюю оценку для f, которую мы можем дать после k шагов перечисления множества G_f. (В ходе этого перечисления мы устанавливаем, что $f(x) < n$ для некоторых x и n, тем самым получая верхние оценки для значения $f(x)$ при некоторых x; для остальных x мы полагаем $F(x,k) = +\infty$.)

Теперь мы можем переформулировать только что доказанную теорему 7 следующим образом:

Theorem 8. [simple-upper-reformulated] **(а)** *Функция KS перечислима сверху, причём* $|\{x \mid \mathit{KS}(x) < n\}| < 2^n$ *при всех* n.

(б) *Если функция KS′ перечислима сверху и* $|\{x \mid \mathit{KS}'(x) < n\}| < 2^n$ *при всех* n, *то найдётся такое* c, *что* $\mathit{KS}(x) < \mathit{KS}'(x) + c$ *для всех* x.

Отметим, что в пункте (б) этой теоремы можно ослабить оценку и написать, что $|\{x \mid \mathit{KS}'(x) < n\}| = O(2^n)$.

Таким образом, колмогоровскую сложность можно определить как минимальную (с точностью до константы) перечислимую сверху функцию K, для которой $|\{x \mid K(x) < n\}| = O(2^n)$.

Можно сделать ещё один шаг и избавиться от требования минимальности, получив следующее «аксиоматическое» определение колмогоровской сложности.

Theorem 9. [simple-axiom] *Пусть* K — *функция с натуральными значениями, определённая на всех двоичных словах. Пусть при этом:*

(а) K *перечислима сверху;* [аксиома перечислимости]

(б) *для любой вычислимой функции* A, *аргументами и значениями которой являются двоичные слова, выполнено неравенство* $K(A(x)) \leqslant K(x) + c$ *при некотором* c *и всех* x, *для которых* $A(x)$ *определено;* [аксиома невозрастания сложности]

(в) *количество слов* x, *для которых* $K(x) < n$, *заключено между* 2^{n-c_1} *и* 2^{n+c_2} *при некоторых* c_1, c_2 *и при всех* n; [аксиома калибровки]

Тогда $K(x) = \mathit{KS}(x) + O(1)$ *(другими словами, разница между* $K(x)$ *и* $\mathit{KS}(x)$ *ограничена константой, не зависящей от* x*).*

$\lhd$ Предыдущая теорема показывает, что $KS(x) \leqslant K(x) + O(1)$. Остаётся доказать, что $K(x) \leqslant KS(x) + O(1)$.

Лемма 1. Существует константа c и вычислимая последовательность конечных множеств двоичных слов

$$M_0 \subset M_1 \subset M_2 \subset \ldots$$

(говоря о вычислимости, мы имеем в виду, что множество задаётся списком своих элементов), в которой M_i содержит ровно 2^i слов и $K(x) \leqslant i + c$ для всех элементов $x \in M_i$ (при любом i).

Доказательство. Свойство (в) функции K гарантирует, что множество $A_i = \{x \mid K(x) < i + c\}$ содержит не менее 2^i элементов (при некотором c и при всех i). Свойство (а) гарантирует, что семейство множеств A_i перечислимо (как семейство перечислимых множеств). Если оставить от множества A_i только 2^i элементов, которые появляются при перечислении первыми, то получится множество B_i, которое содержит ровно 2^i элементов. При этом список элементов множества B_i можно получить алгоритмически по i (дождавшись появления 2^i элементов при перечислении множества A_i — наши предположения гарантируют, что это обязательно произойдёт). Правда, множества B_i не обязаны возрастать с ростом i, но это можно исправить, определив M_i индуктивно. Именно, M_0 полагаем равным B_0, а M_{i+1} есть объединение M_i с 2^i элементами B_{i+1}, не входящими в M_i. Лемма 1 доказана.

Лемма 2. Существует константа c, при которой $K(x) \leqslant l(x) + c$ для любого слова x (напомним, $l(x)$ — длина слова x).

Доказательство. Рассмотрим множества M_i из предыдущей леммы. Рассмотрим вычислимую функцию A, которая определена на объединении всех множеств M_i и отображает $M_{i+1} \setminus M_i$ на множество всех слов длины i. (В множестве $M_{i+1} \setminus M_i$ как раз 2^i слов, так что его можно взаимно однозначно отобразить на слова длины i.) По условию (б) мы знаем, что $K(A(y)) \leqslant K(y) + c'$ при некотором c' и всех x. Применяя это свойство для $y \in M_{i+1} \setminus M_i$, находим, что $K(x) < i + c$ для любого слова x длины i (при всех i и при некотором c, не зависящем от i). Лемма 2 доказана.

Теперь уже легко завершить доказательство теоремы. Пусть D — оптимальный способ описания. Если p — кратчайшее описание для слова x, то $K(x) = K(D(p)) \leqslant K(p) + O(1) \leqslant l(p) + O(1) = KS(x) + O(1)$. (Таким образом, мы дважды использовали свойство (б): один раз при доказательстве леммы 2 и второй раз только что, применяя его к вычислимой функции D.) $\rhd$

$\boxed{4}$ Покажите, что если в качестве описаний использовать слова в четырёхбуквенном алфавите (скажем, конечные последовательности цифр $0, 1, 2, 3$), то сложность (измеряемая как длина кратчайшего описания) будет равна половине обычной.

$\boxed{5}$ (Продолжение.) Сформулируйте и докажите аналогичное утверждение для n-буквенного алфавита.

$\boxed{6}$ `[complexity-bound]` Пусть $f\colon \mathbb{N} \to \mathbb{N}$ — всюду определённая вычислимая возрастающая функция, причём $\liminf f(n+1)/f(n) > 1$. Докажите, что если A_n — перечислимое семейство конечных перечислимых множеств, причём число элементов в A_n не превосходит $f(n)$ при всех n, то найдётся такая константа c, что $KS(x) \leqslant \log f(n) + c$ для любого n и для любого $x \in A_n$.

$\boxed{7}$ Покажите, что при некотором c для всякого слова x любой длины n существует слово y той же длины, отличающееся от x не более чем в одной позиции, для которого $KS(y) \leqslant n - \log n + c$. [Указание. Для произвольного k рассмотрим матрицу Хемминга — матрицу из нулей и единиц размером $k \times (2^k - 1)$, у которой все $2^k - 1$ столбцов различны и каждый содержит хотя бы одну единицу. Эта матрица задаёт линейное над полем $\mathbb{B} = \{0, 1\}$ отображение $\mathbb{B}^{2^k-1}$ в $\mathbb{B}^k$, обладающее таким свойством: в любом векторе длины $2^k - 1$ можно изменить не более одного бита и попасть в ядро этого отображения. Слово y надо брать из ядра такого отображения при $n = 2^k - 1$, и воспользоваться предыдущей задачей. Если число n не имеет вида $2^k - 1$, применяем те же рассуждения к части x' слова x, имеющей удобную (и достаточно большую) длину.]

1.2. Алгоритмические свойства

[simpleal]

Как мы видели, функция KS является перечислимой сверху, но не является вычислимой и даже не имеет вычислимых неограниченных нижних оценок (теорема 6, с. 15).

Заметим, что из этого вытекает, что никакой оптимальный способ описания не является всюду определённым (имеются слова, не являющиеся описаниями). В самом деле, если бы оптимальный способ описания D был всюду определён, то мы могли бы вычислить $KS_D(x)$, просто перепробовав все описания в порядке возрастания длин (до нахождения кратчайшего).

При этом возникает следующий любопытный парадокс. С точки зрения оптимальности наличие слов, не являющихся описаниями, явно невыгодно. Если $D(y)$ не определено, можно рассмотреть другой способ описания D', для которого $D'(y)$ равно некоторому слову z (а в остальном D' совпадает с D). При замене D на D' сложность слова z может уменьшиться, а сложность остальных слов не изменится. Тем не менее для оптимального способа описания всегда есть слова, на которых он не определён!

Формального противоречия тут нет (такое доопределение сохраняет вычислимость, лишь если его сделать в одной точке или в конечном числе точек), но наблюдение это любопытно (его сделал Ю. И. Манин в своей книжке «Вычислимое и невычислимое» [45] — той самой, в которой он обсуждал возможности квантовых компьютеров задолго до того, как их начали изучать).

Заодно мы показали, что область определения оптимального способа описания не может быть разрешимым множеством. (Множество слов называется *разрешимым*, если есть алгоритм, который по любому слову выясняет, принадлежит ли оно этому множеству.) В самом деле, если бы существовал алгоритм, выясняющий, определено ли $D(x)$ или нет, то D можно было бы продолжить до всюду определённого оптимального способа описания (положив, скажем $D(x) = 0$ для тех точек, где $D(x)$ было неопределённым).

Тем самым мы построили алгоритм с неразрешимой областью определения (это — центральный факт теории алгоритмов, см., например, [77]).

Вообще понятие колмогоровской сложности представляет интерес с точки зрения общей теории алгоритмов. Мы рассмотрим два вопроса такого рода — о простоте множества простых слов и о сложности больших чисел.

Простые слова и простые множества

Слово «простые» в этом разделе будет иметь совершенно разные значения в применении к словам и множествам. Говоря о простых словах, мы будем иметь в виду слова малой колмогоровской сложности. Понятие же простого множества, которое мы будем использовать, введено американским логиком Эмилем Постом и никакого отношения к колмогоровской сложности не имеет. (Трудно даже сказать, почему был выбран такой термин.)

Определение. Перечислимое множество A является *простым* (в смысле Поста), если его дополнение бесконечно, но не содержит бесконечного перечислимого подмножества.

Будем называть слово x «простым», если $KS(x) < l(x)/2$.

Theorem 10. *Множество всех «простых» слов является простым в смысле Поста.*

$\triangleleft$ Множество S всех «простых» слов перечислимо, поскольку функция KS вычислима сверху (используя всё более и более точные верхние оценки, мы рано или поздно сможем обнаружить и перечислить любое «простое» слово).

Как мы знаем, число слов сложности менее $n/2$ не превосходит $2^{n/2}$. Поэтому среди слов длины n «простые» слова составляют (ничтожное) меньшинство, так что дополнение к множеству S бесконечно.

Может ли дополнение к множеству S иметь бесконечное перечислимое подмножество? Нет. Предположим, что такое подмножество C существует, и покажем, что функция KS имеет неограниченную вычислимую нижнюю оценку. В самом деле, чтобы найти слово сложности больше t, достаточно перечислять множество C, пока не обнаружится слово c_t длины больше $2t$. Такое слово найдётся, поскольку C бесконечно; это слово должно иметь сложность больше t, иначе оно было бы «простым». Можно считать, что все c_t различны (отбраковывая уже использованные слова). Тогда функция, равная t на слове c_t, будет вычислимой неограниченной нижней оценкой для KS, а такое невозможно по теореме 6 (с. 15). $\triangleright$

Заметим, что в этой теорема граница $l(x)/2$ выбрана произвольно: с тем же успехом можно было бы называть слово «простым», если $KS(x) < l(x) - 1$ (или, скажем, $KS(x) < \log\log l(x)$).

Сложность больши́х чисел

Будем рассматривать сложность $KS(m)$ как функцию натурального числа m, отождествляя каждое двоичное слово с его порядковым номером. Легко понять, что предел $KS(m)$ при $m \to \infty$ равен бесконечности, так как для любого c лишь конечное число объектов могут иметь сложность меньше c. Но это стремление к бесконечности не является эффективно вычислимым: нет алгоритма, который по данному числу n указывал бы то N, начиная с которого колмогоровская сложность становится больше n. (Мы убедились в этом, когда говорили о вычислимых нижних оценках для KS, см. с. 15.)

В этом разделе мы изучим вопрос о скорости стремления KS к бесконечности более подробно. Для этого рассмотрим функцию

$$B(n) = \max\{m \in \mathbb{N} \mid KS(m) \leqslant n\}$$

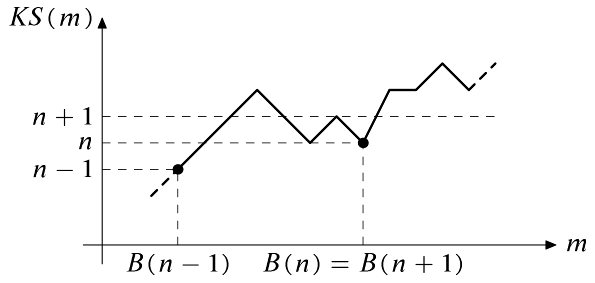

Рис. 1. К определению функции B: в точке $m = B(n-1)$ значение функции KS не превосходит $n-1$ (на рисунке показан случай равенства), а всюду правее оно больше $n-1$. В точке $m = B(n)$ значение функции KS не превосходит n (показан случай равенства), а правее оно больше n (на рисунке показан случай, когда оно даже больше $n+1$, и поэтому $B(n+1) = B(n)$). При $m \in (B(n-1), B(n)]$ значение $KS_{\geqslant}(m)$ равно n.

(наибольшее число, сложность которого не превосходит n). Эту функцию можно назвать «регулятором сходимости» $KS(m) \to \infty$ при $m \to \infty$, поскольку $K(x) > n$ при $x > B(n)$. (Формально говоря, для малых значений n может оказаться, что $KS(m) > n$ при любом натуральном m. В этом случае можно считать, например, что $B(n) = -1$.)

Функцию B можно назвать в каком-то смысле обратной к функции $KS_{\geqslant}(N) = \min\{KS(m) \mid m \geqslant N\}$. Функция $KS_{\geqslant}$ медленно растёт, принимая значение n на участке $(B(n-1), B(n)]$.

Медленный рост $KS_{\geqslant}$ соответствует быстрому росту функции B. О быстроте роста B говорит такой простой факт.

Theorem 11. *Пусть f — произвольная вычислимая функция с натуральными аргументами и значениями. Тогда $B(n) \geqslant f(n)$ для всех n, кроме конечного числа.*

(Заметим, что мы не требуем, чтобы $f(n)$ было определено при всех n. Говорится лишь, что для достаточно больших n, при которых $f(n)$ определено, выполнено неравенство $B(n) \geqslant f(n)$.)

$\lhd$ Алгоритмическое преобразование не увеличивает сложности, поэтому

$$KS(f(n)) \leqslant KS(n) + O(1) \leqslant \log n + c$$

для некоторого c и для всех n, при которых $f(n)$ определено. С другой стороны, $f(n) > B(n)$ влечёт $KS(f(n)) > n$ (по определению функции B). Получается, что для таких значений n выполняются неравенства

$$n < KS(f(n)) \leqslant \log n + c,$$

что возможно лишь для конечного числа значений n. $\rhd$

Определение $B(n)$ можно переформулировать так: пусть D — оптимальный способ описания, использованный при определении сложности; тогда $B(n)$ есть максимальное значение D на всех словах длины не больше n:

$$B(n) = \max\{D(x) \mid l(x) \leqslant n\}.$$

Напомним, что мы отождествляем слова с натуральными числами и считаем, что значениями D являются натуральные числа. При этом максимум пустого множества следует считать равным -1.

Можно вместо D рассмотреть любую вычислимую функцию d, аргументами которой являются слова, а значениями — натуральные числа, и рассмотреть функцию

$$B_d(n) = \max\{d(x) \mid l(x) \leqslant n \text{ и } d(x) \text{ определено}\}.$$

Рассматриваемая нами функция B является наибольшей среди них в следующем смысле:

Theorem 12. *Для всякой функции d найдётся такая константа c, что*

$$B_d(n) \leqslant B(n+c)$$

при всех n.

◁ В самом деле, значение $d(x)$ на слове x длины не более n имеет сложность не более $n+c$ для некоторой константы c, поскольку применение вычислимой функции d увеличивает сложность не более чем на константу, а $KS(x) \leqslant n + O(1)$. Другими словами, $d(x)$ — одно из чисел, сложность которых не больше $n+c$, и не превосходит наибольшего такого числа. ▷

Это (тривиальное) наблюдение окажется нам полезным в таком частном случае. Пусть M — некоторый алгоритм. Мы говорим, что умеем решать *проблему остановки* для алгоритма M и какого-то множества входных слов, если по любому входному слову x из этого множества мы можем определить, останавливается ли M на x или не останавливается.

Классический результат теории алгоритмов (с которого началась эта теория) состоит в том, что для некоторых алгоритмов проблема остановки (для данного алгоритма и произвольного входа) неразрешима.

Нас сейчас будет интересовать эта проблема для слов ограниченной длины. Итак, пусть фиксирован алгоритм M. Рассмотрим функцию $t(x)$, которая есть время работы алгоритма M на слове x. (Если алгоритм M не останавливается на данном x, то $t(x)$ не определено, так что область определения t совпадает с областью определения алгоритма M.) Тогда $B_t(n)$ есть наибольшее время работы алгоритма M на словах длины не более n. Зная число $B_t(n)$ или любое большее его число, мы можем решать проблему остановки для машины M и любого входа x длины не более n: надо подождать указанное время; если алгоритм не остановился, то он не остановится никогда.

Как мы видели, $B_t(n) \leqslant B(n+c)$ для некоторого c (зависящего от M), и зная $B(n+c)$ или любое большее число, мы можем решать проблему остановки для машины M и входов длины не более n. Сформулируем это замечание в виде теоремы:

Theorem 13. *Для всякого алгоритма M найдётся константа c и алгоритм A, который по любому n и по любому числу $t > B(n+c)$ выдаёт список всех слов длины не более n, на которых алгоритм M останавливается.*

Используя традиционную для теории алгоритмов терминологию, можно сказать, что согласно этой теореме проблема остановки для слов длины не более n *сводится* к задаче отыскания числа, большего $B(n+c)$.

Если в качестве M взять оптимальный декомпрессор D, то верно и обратное утверждение: зная n и умея решать проблему остановки D для слов длины не более n, мы можем составить список всех слов сложности не более n и тем самым определить $B(n)$. Заметим, что для решения проблемы остановки для слов длины не более n достаточно указать (помимо n) число слов такой длины, на которых D останавливается.

Это рассуждение можно продолжить, доказав такое утверждение:

Theorem 14. *Пусть $BB(n)$ — максимальное время работы оптимального декомпрессора на словах длины не более n. Тогда*

$$BB(n) \leqslant B(n+c) \quad \text{и} \quad B(n) \leqslant BB(n+c)$$

для некоторого c и всех n.

$\lhd$ Число $BB(n)$ имеет сложность не более $n + O(1)$, поскольку может быть получено по n и самому долго обрабатываемому описанию x длины не более n. Эта информация может быть записана в одном слове длины $n+1$, а именно, слове $0 \ldots 01x$ (в начале стоят $n - l(x)$ нулей). Поэтому $BB(n) \leqslant B(n+c)$ для некоторого c и всех n.

С другой стороны, по любому $t > BB(n)$ можно алгоритмически получить слово сложности больше n, если знать ещё и n (поскольку подождав t шагов для каждого описания длины не более n, можно обнаружить все слова сложности не более n, а затем взять первое слово, не попавшее в их число). Таким образом, для задания этого слова достаточно $KS(t)$ битов плюс $2\log n$ битов для самоограниченного описания n, а сложность его по построению больше n.

Поэтому при $t > BB(n)$ имеем $KS(t) \geqslant n - 2\log n - O(1)$. Отсюда следует, что $BB(n) \geqslant B(n - 2\log n - c)$. (Это немного меньше, чем нам требуется.)

Теперь мы можем дополнительно заметить, что для построения слова сложности более n при известном $t > BB(n)$ нам не обязательно знать само n, достаточно иметь верхнюю оценку n' для него, сделать t шагов работы декомпрессора для всех описаний длины не более n', а затем выбрать слово, отличное от всех обнаруженных результатов работы декомпрессора.

При этом значение $BB(n)$ (а также любое t, большее $BB(n)$) можно использовать в качестве верхней оценки для n, поскольку $B(n - 2\log n - c)$ заведомо больше n. Таким образом, по любому $t > BB(n)$ можно алгоритмически указать слово сложности более n, поэтому и само t имеет сложность не меньше $n - O(1)$. Отсюда получаем $BB(n) > B(n - O(1))$, что и требовалось. $\rhd$

Эта теорема показывает, что функцию B можно примерно (с точностью до $O(1)$-добавки в аргументе) охарактеризовать как «время обработки самого трудного входа» ограниченной длины. Родственная конструкция появлялась в теории сложности вычислений под именем Busy Beaver function («трудолюбивые бобры») — там для данного n рассматривалась машина Тьюринга с n состояниями и двухбуквенным алфавитом (единица и пробел), которая останавливается на пустом входе, напечатав максимально возможное число единиц (не обязательно подряд).

Вообще можно сказать, что знание любого из следующих объектов (а также значения параметра n) позволяет найти любой другой (возможно, для чуть меньшего значения n):

(а) список всех слов сложности не более n с указанием их сложностей;

(б) число таких слов;

(в) значение $B(n)$;

(г) значение $BB(n)$;

(д) список всех слов длины не более n, на которых определён оптимальный декомпрессор («проблема остановки» оптимального декомпрессора на словах длины не более n);

(е) число таких слов;

(ё) самый долго обрабатываемый оптимальным декомпрессором вход длины не более n;

(ж) таблица T_n, в которой указаны значения сложности $KS(x)$ для всех слов x длины n;

(з) первое в словарном порядке слово γ_n длины n, имеющее сложность не менее n (напомним, что такое существует, так как слов сложности менее n меньше, чем слов длины n).

Более точно, имеет место следующая теорема:

Theorem 15. [quasi-omega-words] *Все перечисленные объекты имеют сложность $n+O(1)$ и эквивалентны друг другу в следующем точном смысле: пусть X_n и Y_n — объекты, указанные в каких-либо двух пунктах из числа* (а)–(з). *Тогда найдётся константа c и алгоритм, позволяющий по X_n и n найти Y_{n-c}.*

$\triangleleft$ Проще всего доказать эквивалентность объектов (г), (д), (е) и (ё). Зная любой из них и число n, можно выяснить поведение оптимального декомпрессора на всех описаниях длины не более n, то есть выписать все его останавливающиеся вычисления на входах длины не более n. В самом деле, зная список (д), мы просто применяем оптимальный декомпрессор ко всем входам из этого списка (зная заранее, что вычисления остановятся). Зная (г), мы применяем оптимальный декомпрессор ко всем входам длины не более n, но даём ему проработать не более $BB(n)$ шагов (будучи уверенными, что не остановившиеся за это время вычисления никогда не остановятся). Зная самый долго обрабатываемый вход длины не более n, мы применяем к нему оптимальный декомпрессор и находим $BB(n)$, а дальше действуем, как в предыдущем случае. Наконец, зная число слов длины не более n, на которых оптимальный декомпрессор останавливается, мы применяем параллельно его ко всем словам длины не более n, пока не наберём нужное количество результатов (оставшиеся вычисления не закончатся никогда).

Обратно, зная n и поведение оптимального декомпрессора на всех описаниях длины не более n, легко найти любой из объектов (г), (д), (е) и (ё), а также, впрочем, и объекты (а)–(в). По транзитивности (почти очевидной) мы заключаем, что (г) – (ё) все эквивалентны.

Теперь докажем, что (а) – (в) эквивалентны между собой и эквивалентны (г) – (ё). Зная список всех объектов сложности не больше n, можно найти их число (переход от (а) к (б)). Перейти от (б) к (а) не так просто: зная число объектов сложности не больше n (и зная само n), можно дождаться появления всех таких объектов, и получить их список (в котором можно найти наибольшее число, получив (в)), но этот список будет без указания сложностей. Поскольку мы уже знаем, как от (г) перейти к (а), нам достаточно перейти от (в) к (г), то есть от $B(n)$ к $BB(n)$ (с изменением аргумента на $O(1)$). По существу этот переход уже обсуждался. Зная $B(n)$, мы знаем верхнюю оценку для $BB(n-c)$, и остаётся лишь подождать указанное время для всех входов длины не более $n-c$, чтобы узнать точное значение $BB(n-c)$. Итак, мы доказали эквивалентность всех объектов (а)–(ё).

Зная (а)–(ё), можно найти (ж) для чуть меньшего значения n — уменьшенного на константу c, для которой $KS(x) \leqslant l(x)+c$ при всех x. Ясно также, как от (ж) можно перейти к (з). Осталось показать, как от (з) перейти к (г). Как, зная первое в алфавитном порядке слово γ_n сложности не менее n среди слов длины n, найти $BB(n-O(1))$ или хотя бы получить верхнюю оценку для $BB(n-O(1))$? Это делается следующим образом.

Зная γ_n (и тем самым зная n), будем искать для всех предшествующих ему слов длины n их описания длины менее n; раз они существуют, то рано или поздно мы их найдём (пусть даже и не кратчайшие). Найдя их, рассмотрим максимальное время работы оптимального декомпрессора на этих (найденных нами) описаниях; пусть это время будет T. Мы хотим доказать, что $T > BB(n-c)$ для некоторой константы c, не зависящей от n. Допустим, что для данного c это неравенство не выполнено, то есть $T \leqslant BB(n-c)$. Докажем, что тогда c не превосходит некоторой константы (не зависящей от n). Рассмотрим самое долгоиграющее описание длины не более $n-c$ и обозначим через $n-c-d$ его длину. Зная это описание и $c+d$, мы можем найти n и $BB(n-c)$. Этой информации нам достаточно, чтобы найти слово γ_n, поскольку γ_n есть минимальное слово длины n, для которого после $BB(n-c)$ шагов не обнаруживается описания короче n. По построению слово γ_n имеет сложность не менее n, поэтому $n \leqslant KS(\gamma_n) \leqslant (n-c-d)+2\log(c+d)+O(1)$. Значит $(c+d)=O(1)$.

Тем самым завершено доказательство эквивалентности объектов (а)–(ж) (в указанном выше смысле). Докажем теперь, что сложность любого из них равна $n+O(1)$. Пусть X_n — любой из этих объектов. Как мы доказали, X_n можно получить по γ_{n+c} и n (число n равно $l(\gamma_{n+c})-c$ и потому может быть опущено), следовательно, $KS(X_n) \leqslant KS(\gamma_{n+c})+O(1) \leqslant n+O(1)$.

С другой стороны, пусть сложность X_n равна $n-d$. Слово γ_{n-c} по определению имеет сложность не меньше $n-c$ и может быть получено по кратчайшему описанию X_n длины $n-d$ и по d (заметим, что по этим данным можно восстановить n, сложив длину кратчайшего описания с d). Следовательно, $n-c \leqslant KS(\gamma_{n-c}) \leqslant (n-d)+2\log d+O(1)$, то есть $d \leqslant 2\log d + c + O(1)$ и $d = O(1)$. ▷

При доказательстве теоремы 15 мы предполагали, что фиксирован некоторый оптимальный декомпрессор (вообще говоря, все рассматриваемые в ней объекты зависят от его выбора). Это, однако, не играет роли, как показывает следующая задача:

$\boxed{8}$ Докажите, что утверждение теоремы 15 остаётся верным, если разрешить в разных пунктах (а)–(ж) использовать разные оптимальные декомпрессоры.

$\boxed{9}$ Покажите, что сложность всех объектов теоремы 15 становится равной $O(\log n)$, если релятивизовать определение сложности относительно $\mathbf{0}'$, то есть разрешить декомпрессору обращаться к оракулу для проблемы остановки.

Мы установили, что существует константа c и алгоритм A, который по слову γ_n решает проблему остановки оптимального декомпрессора для всех слов длины не больше $n-c$. Эта означает, что если у нас есть «оракул», который для каждого n указывает слово γ_n, то с его помощью можно решать проблему остановки. Вместо этого можно также использовать оракул для множества «несжимаемых» слов (тех слов x, для которых $KS(x) \geqslant l(x)$): умея проверять несжимаемость слова, мы можем найти γ_n перебором.

Как говорят, проблема остановки *сводится по Тьюрингу* к множеству несжимаемых слов. Отсюда очевидно следует, что она сводится по Тьюрингу к «надграфику» функции KS, то есть множеству $\{\langle x, k\rangle \mid KS(x) < k\}$. Как говорят, множество сжимаемых слов является *полным по Тьюрингу* перечислимым множеством (это означает, что к нему сводится по Тьюрингу проблема остановки).

$\boxed{10}$ Оценить сверху (хоть как-нибудь) число вопросов, которые придётся задать оракулу для множества $\{\langle x, k\rangle \mid KS(x) < k\}$, чтобы решить проблему остановки оптимального декомпрессора для всех слов длины не более n.

$\boxed{11}$ Докажите, что если f — вычислимая функция с натуральными аргументами и значениями, то найдётся такая константа c, что для всех n, для которых $f(B(n))$ определено, выполнено неравенство $B(n+c) \geqslant f(B(n))$. [Указание: число $f(B(n))$ имеет сложность не больше $n + O(1)$.]

$\boxed{12}$ Говорят, что множество U является *r-отделимым* [59], если всякое перечислимое множество V, не пересекающееся с U, можно отделить от U разрешимым множеством, то есть найдётся разрешимое множество R, содержащее V и не пересекающееся с U.

(а) Докажите, что надграфик функции KS, то есть множество пар $\{\langle x, k\rangle \mid KS(x) < k\}$, является r-отделимым множеством. [Указание: если этот график не пересекается с некоторым перечислимым множеством V, то вторые компоненты пар из V ограничены, иначе мы получили бы неограниченную нижнюю оценку для KS. Значит, V целиком содержится в некоторой полосе, а пересечение надграфика с этой полосой конечно.]

(б) Докажите, что если множество U_1 m-сводится к множеству U_2 (существует всюду определённая вычислимая функция f, для которой $U_1 = f^{-1}(U_2)$) и U_2 является r-отделимым, то U_1 также r-отделимо. [Указание. Если перечислимое V не пересекается с U_1, то $f(V)$ является перечислимым множеством, не пересекающимся с U_2. Если R отделяет $f(V)$ от U_2, то $f^{-1}(R)$ будет разрешимым множеством, отделяющим V от U_1.]

(в) Докажите, что существуют перечислимые множества, не обладающие свойством r-отделимости и потому не m-сводящиеся к надграфику функции KS. [Указание: существуют перечислимые неотделимые множества.]

Эта задача показывает, что с помощью понятия сложности можно построить перечислимое неразрешимое множество, не являющееся m-полным (надграфик функции KS).

[Здесь хорошо бы сослаться на результаты Посицельского, Мучника, Downey, Merkle и других, либо в специальной главе, либо просто ссылки куда-нибудь — если на специальную главу нет надежды, то хотя бы сформулировать основные результаты!!!]

Теорема 15 указывает среди всех объектов сложности n некоторый выделенный (с точностью до описанной эквивалентности). Это отчасти парадоксально: хотелось бы думать, что все «случайные» слова длины n (слова длины n и сложности примерно n) более или менее одинаковы (если бы какое-то из них выделялось чем-то особенным, то эту особенность можно было бы использовать, чтобы задать его короче, и тем самым слово было бы не случайным). Тем не менее мы указали некоторое специальное случайное слово γ_n, как же так? Разрешение этого парадокса состоит в том, что индивидуальные особенности этого слова могут быть использованы для его простого задания, но лишь с оракулом $\mathbf{0}'$.

Мы вернёмся ещё к этому вопросу, когда будем говорить о числе Ω в разделе 5.7, а также в разделе 16.3, говоря о двухчастных описаниях.

Отметим ещё, что хотя все объекты из теоремы 15 эквивалентны в указанном в ней смысле (просто получаются друг из друга), но по длине они отличаются радикально: от длины n в пунктах (б), (е), (ё) и (з) до длины $\log B(n)$ в пункте (в).

2. Сложность пары и условная сложность

[conditional]

2.1. Сложность пары

[conditp]

Мы уже говорили, что вычислимое кодирование позволяет говорить не только о сложности слов, но и о сложности других конструктивных объектов. Сейчас нас интересуют (упорядоченные) пары слов. Пару слов x, y можно кодировать, например, словом $[x, y] = \overline{x}01y$, где $\overline{x}$ означает слово x с удвоенными битами. Можно использовать и любое другое кодирование, важно только, чтобы оно было вычислимым и чтобы $[x, y] \neq [x', y']$, если $x \neq x'$ или $y \neq y'$. От одного такого кодирования можно алгоритмически переходить к другому, поэтому теорема 3 (с. 11) о невозрастании сложности при алгоритмическом преобразовании показывает, что смена кодирования изменяет сложность не более чем на константу.

Фиксируем какое-либо кодирование пар ($[x, y]$ обозначает слово, кодирующее пару слов x, y) и назовём сложностью пары слов x, y число $\mathit{KS}([x, y])$. Обозначение: $\mathit{KS}(x, y)$. Вот несколько очевидных свойств:

- $\mathit{KS}(x, x) = \mathit{KS}(x) + O(1)$;
- $\mathit{KS}(x, y) = \mathit{KS}(y, x) + O(1)$;
- $\mathit{KS}(x) \leqslant \mathit{KS}(x, y) + O(1)$; $\mathit{KS}(y) \leqslant \mathit{KS}(x, y) + O(1)$.

Следующая теорема оценивает сложность пары, если известны сложности её компонент:

Theorem 16. [condit-pair1]

$$\begin{aligned}
&\mathit{KS}(x, y) \leqslant \mathit{KS}(x) + 2\log \mathit{KS}(x) + \mathit{KS}(y) + O(1);\\
&\mathit{KS}(x, y) \leqslant \mathit{KS}(x) + \log \mathit{KS}(x) + 2\log\log \mathit{KS}(x) + \mathit{KS}(y) + O(1);\\
&\mathit{KS}(x, y) \leqslant \mathit{KS}(x) + \log \mathit{KS}(x) + \log\log \mathit{KS}(x) + 2\log\log\log \mathit{KS}(x) + \mathit{KS}(y) + O(1);\\
&\qquad\qquad \ldots
\end{aligned}$$

(Последовательность утверждений теоремы можно продолжать неограниченно. Кроме того, можно поменять местами x и y.)

$\triangleleft$ Это рассуждение (для первого из неравенств) по существу проводилось во введении (теорема 4 на с. 12); правда, там мы говорили о сложности конкатенации слов xy, а не о сложности пары. Повторим его для пар.

Назовём вычислимое отображение $x \mapsto \hat{x}$ множества двоичных слов в себя *беспрефиксным кодированием*, если ни для каких двух различных слов x и y слово $\hat{x}$ не является началом слова $\hat{y}$. (Отсюда, в частности, следует, что $\hat{x} \neq \hat{y}$ при $x \neq y$.) Смысл этого определения таков: любое слово вида $\hat{x}z$ можно однозначно разрезать на части и найти слова x и z.

Пример беспрефиксного кодирования: $x \mapsto \overline{x}01$, где $\overline{x}$ означает строку x с удвоенными битами. Здесь признаком окончания слова являются биты 01. Это кодирование не очень экономно (увеличивает длину вдвое). Более экономно такое кодирование:

$$x \mapsto \hat{x} = \overline{\mathrm{bin}(l(x))}\,01x$$

(здесь $\mathrm{bin}(l(x))$ — двоичная запись длины слова x). Для него

$$l(\hat{x}) = l(x) + 2\log l(x) + O(1).$$

Этот приём можно «итерировать»: начав с произвольного беспрефиксного кодирования $x \mapsto \hat{x}$, можно построить новое (также беспрефиксное) кодирование

$$x \mapsto \widehat{\mathrm{bin}(l(x))}x.$$

В самом деле, если слово $\widehat{\mathrm{bin}(l(x))}x$ является началом слова $\widehat{\mathrm{bin}(l(y))}y$, то одно из слов $\widehat{\mathrm{bin}(l(x))}$ и $\widehat{\mathrm{bin}(l(y))}$ является началом другого, и потому $\mathrm{bin}(l(x)) = \mathrm{bin}(l(y))$. Отсюда мы заключаем, что x есть начало y, а потом — что $x = y$. (Другими словами, сначала мы однозначно декодируем длину слова, пользуясь тем, что она записана в беспрефиксном коде, а потом уже однозначно определяем само слово.)

Такая итерация даёт беспрефиксное кодирование, при котором

$$l(\hat{x}) = l(x) + \log l(x) + 2\log\log l(x) + O(1),$$

затем

$$l(\hat{x}) = l(x) + \log l(x) + \log\log l(x) + 2\log\log\log l(x) + O(1),$$

и так далее.

Вернёмся к доказательству теоремы. Пусть D — оптимальный способ описания, используемый при определении сложности. Рассмотрим способ описания D', задаваемый так:

$$D'(\hat{p}q) = [D(p), D(q)],$$

где $\hat{p}$ — беспрефиксный код слова p, а квадратные скобки означают кодирование пар, использованное при определении сложности пары. Беспрефиксность кодирования $p \mapsto \hat{p}$ гарантирует корректность этого определения ($\hat{p}$ однозначно вычленяется из $\hat{p}q$).

Если p и q — кратчайшие описания слов x и y, то слово $\hat{p}q$ будет описанием слова $[x, y]$, и его длина как раз даёт оценку, указанную в теореме. (Чем более экономно беспрефиксное кодирование, тем лучше получается оценка; описанный выше итеративный метод построения беспрефиксных кодов даёт необходимые оценки.) $\triangleright$

Из теоремы 16 следует, что

$$KS(x, y) \leqslant KS(x) + KS(y) + O(\log n)$$

для слов x и y длины не более n. Как говорят, сложность пары не превосходит суммы сложностей её членов с точностью до логарифма (длин).

Возникает естественный вопрос: нельзя ли усилить оценку и доказать, что $KS(x, y) \leqslant KS(x) + KS(y) + O(1)$?

Следующее простое рассуждение показывает, что это невозможно.[condit-pair1-comment] В самом деле, из этой оценки вытекало бы, что $KS(x,y) \leqslant l(x)+l(y)+O(1)$. Рассмотрим какое-то N и всевозможные $n=0,1,2,\ldots,N$. Для каждого такого n имеется 2^n слов x длины n, а также 2^{N-n} слов y длины $N-n$. Комбинируя такие x и y, мы для данного n получаем 2^N пар $\langle x,y\rangle$, а всего (для всех n от 0 до N) получаем $(N+1)2^N$ пар. Если бы сложность всех этих пар $\langle x,y\rangle$ не превосходила $l(x)+l(y)+O(1)=N+O(1)$, то получилось бы $(N+1)2^N$ различных слов $[x,y]$, имеющих сложность не более $N+O(1)$, а таких слов, как мы знаем (теорема 7, с. 23), лишь $O(2^N)$.

$\boxed{13}$ [conditp-no-improvement] Докажите, что не найдётся такого c, что

$$KS(x,y) \leqslant KS(x)+\log KS(x)+KS(y)+c$$

при всех x и y. [Указание. Замените в правой части неравенства сложности на длины и подсчитайте количество пар.]

$\boxed{14}$ Дайте естественное определение сложности тройки слов. Покажите, что $KS(x,y,z) \leqslant KS(x)+KS(y)+KS(z)+O(\log n)$, если слова x,y,z имеют длину не больше n.

$\boxed{15}$ (а) Покажите, что если $x \mapsto \hat{x}$ — беспрефиксное кодирование, то

$$\sum_{x\in\Xi} 2^{-l(\hat{x})} \leqslant 1$$

(здесь Ξ — множество всех двоичных слов).

(б) Покажите, что если беспрефиксное кодирование увеличивает длину слова не более чем на $f(n)$ (где n — исходная длина), то есть $l(\hat{x}) \leqslant l(x)+f(l(x))$, то ряд $\sum_n 2^{-f(n)}$ сходится.

Эта задача объясняет, почему понадобился коэффициент 2 при логарифме в доказательстве теоремы 16 (с. 36): ряды

$$\sum\frac{1}{n^2},\quad \sum\frac{1}{n(\log n)^2},\quad \sum\frac{1}{n\log n(\log\log n)^2},\ldots$$

сходятся, в то время как ряды

$$\sum\frac{1}{n},\quad \sum\frac{1}{n\log n},\quad \sum\frac{1}{n\log n\log\log n},\ldots$$

расходятся.

$\boxed{16}$ Докажите, что все неравенства в теореме 16 перестанут быть верными, если заменить в них коэффициент 2 на 1 (но можно заменить его на $1+\varepsilon$ при любом $\varepsilon>0$). [Указание. Первое из неравенств рассматривалось в задаче 13.]

$\boxed{17}$ Докажите, что

$$KS(x,y) \leqslant KS(x)+\log KS(x)+KS(y)+\log KS(y)+O(1).$$

$\boxed{18}$ (Продолжение) Докажите, что верно даже более сильное неравенство:

$$KS(x, y) \leqslant KS(x) + KS(y) + \log(KS(x) + KS(y)) + O(1)$$

(где сумму под логарифмом можно заменить максимумом, так как они отличаются не более чем вдвое).

$\boxed{19}$ [complexity-added] Докажите, что $KS(x, KS(x)) = KS(x) + O(1)$. [Указание. Очевидно, что $KS(x, KS(x)) \geqslant KS(x) + O(1)$. С другой стороны, кратчайшее описание p слова x задаёт одновременно и x, и $KS(x)$, так что $KS(x, KS(x)) \leqslant l(p) + O(1) = KS(x) + O(1)$.]

2.2. Условная сложность

[condit-c] Посылая файл по электронной почте, можно сэкономить, если послать не сам файл, а его сжатый вариант (кратчайшее описание). Экономия может быть ещё больше, если получатель уже имеет старую версию файла — в таком случае достаточно описать внесённые изменения. Эти соображения приводят к следующему определению *условной сложности слова x при известном слове y*.

Назовём *способом условного описания* произвольную вычислимую функцию D двух аргументов (аргументы и значения функции D являются двоичными словами). Первый аргумент мы будем называть *описанием*, второй — *условием*. Если $D(y, z) = x$, мы говорим, что слово y является *описанием* слова x *при известном* z (ещё говорят «при условии z», «относительно z»). Сложность $KS_D(x|z)$ определяется как длина кратчайшего описания слова x при известном слове z:

$$KS_D(x|z) = \min\{l(y) \mid D(y, z) = x\}.$$

Говорят, что способ (условного) описания D_1 *не хуже* способа D_2, если найдётся такая константа c, что

$$KS_{D_1}(x|z) \leqslant KS_{D_2}(x|z) + c$$

для любых слов x и z. Способ (условного) описания D называется *оптимальным*, если он не хуже любого другого способа (условного) описания.

Theorem 17. [condit-universal] *Существует оптимальный способ условного описания.*

$\lhd$ Этот вариант теоремы Колмогорова – Соломонова доказывается точно так же, как и безусловный (см. теорему 1, с. 9).

Именно, фиксируем некоторый способ программирования для функций двух аргументов, при котором программы записываются в виде двоичных слов, и положим

$$D(\hat{p}y, z) = p(y, z),$$

где $p(y, z)$ обозначает результат применения программы p ко входам y и z, а $\hat{p}$ есть беспрефиксный код слова p. Теперь легко проверить, что если D' — произвольный способ условного описания, а p — соответствующая ему программа, то

$$KS_D(x|z) \leqslant KS_{D'}(x|z) + l(\hat{p}).$$

$\triangleright$

Как и прежде, мы фиксируем некоторый оптимальный способ D условного описания и опускаем индекс D в KS_D. Соответствующую функцию мы называем *условной колмогоровской сложностью*; как и безусловная, она определена с точностью до ограниченного слагаемого.

Вот несколько простых свойств условной колмогоровской сложности:

Theorem 18. [condit-basic]

$$\begin{aligned} KS(x|y) &\leqslant KS(x) + O(1); \\ KS(x|x) &= O(1); \\ KS(f(x,y)|y) &\leqslant KS(x|y) + O(1); \\ KS(x|y) &\leqslant KS(x|g(y)) + O(1). \end{aligned}$$

Здесь g и f — произвольные вычислимые функции (одного и двух аргументов); имеется в виду, что указанные в теореме неравенства выполнены, если $f(x,y)$ (соответственно $g(y)$) определено.

$\triangleleft$ Безусловный способ описания можно рассматривать и как условный (не зависящий от второго аргумента), отсюда получаем первое неравенство. Второе утверждение получается, если рассмотреть способ описания $D(p,z) = z$. Третье неравенство доказывается так: пусть D — оптимальный способ условного описания; рассмотрим другой способ

$$D'(p,y) = f(D(p,y),y)$$

и сравним его с оптимальным. Доказательство последнего неравенства аналогично, только нужно определить способ описания так:

$$D'(p,y) = D(p,g(y)).$$

$\triangleright$

$\boxed{20}$ Покажите, что условная сложность «непрерывна по второму аргументу»: $KS(x|y0) = KS(x|y) + O(1)$; $KS(x|y1) = KS(x|y) + O(1)$.

$\boxed{21}$ Покажите, что при фиксированном y функция $x \mapsto KS(x|y)$ отличается от KS не более чем на константу, зависящую от y (и не превосходящую $2\,KS(y) + O(1)$).

$\boxed{22}$ Докажите, что $KS([x,z]|[y,z]) \leqslant KS(x|y) + O(1)$ для любых x, y, z (квадратные скобки означают вычислимое кодирование пар).

$\boxed{23}$ [conditional-as-problem] Пусть фиксирован некоторый разумный язык программирования. (Формально говоря, нужно, чтобы соответствующая ему нумерация была главной, то есть чтобы была возможна эффективная трансляция программ с других языков [77].) Покажите, что условная сложность $KS(x|y)$ равна (с точностью до $O(1)$) минимальной сложности программы, которая даёт x на входе y. [Указание. Если D — оптимальный способ условного описания, то сложность программы, которая получается из D фиксацией первого аргумента p, не превосходит $l(p) + O(1)$. С другой стороны, если программа p переводит y в x, то $KS(x|y) = KS(p(y)|y) \leqslant KS(p) + O(1)$.]

К этой интерпретации условной сложности (как минимальной сложности объекта с определённым свойством) мы вернёмся в главе 13.

Многие свойства безусловной сложности легко переносятся и на условную. Вот некоторые из них (доказательства повторяют соответствующие рассуждения для безусловной сложности):

- Функция $\mathit{KS}(x|y)$ перечислима сверху (это означает, что множество троек $\langle x, y, n\rangle$, для которых $\mathit{KS}(x|y) < n$, перечислимо).
- При данных y и n множество тех слов x, для которых $\mathit{KS}(x|y) < n$, содержит менее 2^n элементов. Отсюда следствие:
- Для всякого y и для всякого n найдётся слово x длины n, сложность которого при известном y не меньше n.

$\boxed{24}$ Докажите, что для любых слов y и z и для любого n найдётся слово x длины n, у которого $\mathit{KS}(x|y) \geqslant n-1$ и $\mathit{KS}(x|z) \geqslant n-1$. [Указание: плохие слова каждого типа образуют менее половины.]

Theorem 19. `[condit-ub]` *Если $\langle x, y\rangle \mapsto K(x|y)$ — произвольная перечислимая сверху функция, причём для любых y и n множество*

$$\{x \mid K(x|y) < n\}$$

содержит менее 2^n элементов, то $\mathit{KS}(x|y) \leqslant K(x|y) + c$ при некотором c и при любых x и y.

В теореме о сложности пары (теорема 16, с. 36) также можно заменить сложность на условную:

Theorem 20. `[condit-pair2]`

$$\mathit{KS}(x, y) \leqslant \mathit{KS}(x) + 2\log \mathit{KS}(x) + \mathit{KS}(y|x) + O(1)$$

$\lhd$ Пусть D_1 — оптимальный способ безусловного описания, а D_2 — оптимальный способ условного описания. Построим способ описания D', положив

$$D'(\hat{p}q) = [D_1(p), D_2(q, D_1(p))].$$

Здесь $\hat{p}$ — беспрефиксный код слова p, а квадратные скобки обозначают вычислимое кодирование пар, используемое при определении сложности пары. Если p — кратчайшее описание слова x, а q — кратчайшее описание слова y при известном x, то слово $\hat{p}q$ будет описанием слова $[x, y]$ относительно D', откуда

$$\mathit{KS}(x, y) \leqslant \mathit{KS}_{D'}(x, y) + O(1) \leqslant l(\hat{p}) + l(q) + O(1).$$

Осталось заметить, что можно выбрать беспрефиксное кодирование так, чтобы $l(\hat{p})$ не превосходило $l(p) + 2\log l(p) + O(1)$ (см. доказательство теоремы 16 о сложности пары, с. 36). $\rhd$

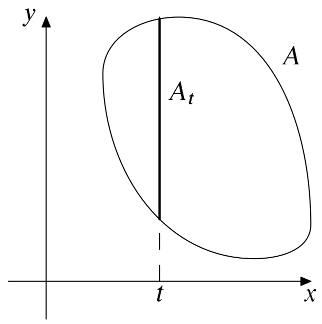

Рис. 2. Сечение A_t множества A простых пар.
[condit-c.1]

В этой теореме, как и раньше, можно заменить $2\log KS(x)$ на $\log KS(x) + 2\log\log KS(x)$ и так далее. Можно также добавочный член перенести на условную сложность, написав

$$KS(x, y) \leqslant KS(x) + KS(y|x) + 2\log KS(y|x) + O(1).$$

В доказательстве при этом следует заменить $D'(\hat{p}q)$ на $D'(\hat{q}p)$.

$\boxed{\textbf{25}}$ Докажите «неравенство треугольника»:

$$KS(x|z) \leqslant KS(x|y) + 2\log KS(x|y) + KS(y|z) + O(1)$$

для любых трёх слов x, y, z.

Если не вдаваться в тонкости различных вариантов оценки добавочного члена с логарифмом, результат теоремы 20 можно сформулировать так: для слов x и y длины не более n имеет место неравенство

$$KS(x, y) \leqslant KS(x) + KS(y|x) + O(\log n).$$

Оказывается, что это неравенство на самом деле представляет собой равенство (с той же логарифмической точностью).

Theorem 21 (Колмогорова – Левина). [condit-pair3]

$$KS(x, y) = KS(x) + KS(y|x) + O(\log n).$$

для слов x, y длины не больше n.

$\triangleleft$ В одну сторону неравенство уже доказано. Осталось доказать, что $KS(x, y) \geqslant KS(x) + KS(y|x) + O(\log n)$, если x и y — слова длины не более n.

Пусть x, y — произвольные слова длины не более n. Обозначим сложность $KS(x, y)$ пары $\langle x, y\rangle$ через a. Пусть A — множество всех пар слов, у которых сложность не больше a. Число элементов в множестве A не больше $O(2^a)$ (точнее, меньше 2^{a+1}), и пара $\langle x, y\rangle$ — один из них.

Для каждого слова t рассмотрим сечение A_t множества A:

$$A_t = \{u \mid \langle t, u\rangle \in A\}$$

(рис. 2). Сумма мощностей всех множеств A_t при всех t равна мощности множества A, то есть не превосходит $O(2^a)$. Поэтому больши́х сечений у множества A немного, что мы сейчас и используем.

Пусть $m = \lfloor \log_2 |A_x| \rfloor$, где x — первая компонента исходной пары. Другими словами, пусть число элементов в A_x заключено между 2^m и 2^{m+1}. Докажем, что $\mathit{KS}(y|x)$ не сильно превосходит m, а $\mathit{KS}(x)$ не сильно превосходит $a - m$. Начнём с первого.

Зная a, можно перечислять множество A. Если мы к тому же знаем x, то можно оставлять от A только пары, у которых первая координата равна x, и получить перечисление множества A_x. Чтобы задать y, помимо a и условия x, достаточно указать порядковый номер элемента y в этом перечислении. Для этого достаточно $m + O(1)$ битов, и вместе с a получается $m + O(\log n)$ битов. Заметим, что $a = \mathit{KS}(x, y)$ не превосходит $O(n)$, если x и y — слова длины не более n, а потому для задания a достаточно $O(\log n)$ битов. Итак,

$$\mathit{KS}(y|x) \leqslant m + O(\log n).$$

Перейдём ко второй оценке. Множество B всех t, для которых $|A_t| \geqslant 2^m$, содержит не более $2^{a+1}/2^m$ элементов (иначе сумма $|A| = \sum |A_t|$ была бы больше 2^{a+1}). Множество B можно перечислять, зная a и m. (В самом деле, надо перечислять пары в множестве A; как только найдётся 2^m пар с одинаковой первой координатой, эта координата помещается в перечисление множества B.) Тем самым исходное слово x (как и любой другой элемент множества B) можно задать, указав $(a - m) + O(\log n)$ битов ($a - m$ битов уходят на порядковый номер слова x в B, а $O(\log n)$ битов позволяют дополнительно указать a и m). Отсюда

$$\mathit{KS}(x) \leqslant (a - m) + O(\log n),$$

и остаётся сложить два полученных неравенства. $\triangleright$

Доказанную только что теорему можно рассматривать как перевод на язык колмогоровской сложности такого комбинаторного факта. Пусть дано множество A пар слов. Его мощность не превосходит произведения мощности проекции на первую координату и мощности наибольшего сечения A_t (первая координата равна t, вторая произвольна). Это соответствует неравенству $\mathit{KS}(x, y) \leqslant \mathit{KS}(x) + \mathit{KS}(y|x) + O(\log n)$. Обратное неравенство интерпретируется сложнее. Пусть дано множество A пар слов, а также числа p и q, для которых $|A| \leqslant pq$. Тогда можно разбить A на две части P и Q с такими свойствами: проекция P на первую координату содержит не более p элементов, а все сечения Q_x множества Q (первая координата равна x, вторая произвольна) содержат не более q элементов. (В самом деле, если отнести к P те сечения, которые содержат больше q элементов, то их число не превосходит p.) Подробнее об этом см. в главе 10.

Заметим, что на самом деле для доказательства важны не длины слов x и y, а их сложности. По существу мы доказали такой факт:

Theorem 22 (Колмогорова – Левина, вариант со сложностью). `[condit-pair3a]`

$$\mathit{KS}(x, y) = \mathit{KS}(x) + \mathit{KS}(y|x) + O(\log \mathit{KS}(x, y))$$

для любых слов x, y.

$\boxed{26}$ Оцените более точно константы в проведённом доказательстве и покажите, что

$$KS(x) + KS(y|x) \leqslant KS(x, y) + 3\log KS(x, y) + O(\log\log KS(x, y)).$$

$\boxed{27}$ Покажите, что в теореме Колмогорова – Левина члены порядка $O(\log n)$ неизбежны (причём в обе стороны): при любом n найдутся слова x и y длины не более n, для которых

$$KS(x, y) \geqslant KS(x) + KS(y|x) + \log n - O(1),$$

а также слова x и y длины не более n, для которых

$$KS(x, y) \leqslant KS(x) + KS(y|x) - \log n + O(1).$$

[Указание. В первом случае можно воспользоваться замечанием после теоремы 16 (с. 38). Во втором случае можно взять в качестве x число между $n/2$ и n, для которого $KS(x) = \log n + O(1)$, а затем в качестве y взять слово длины x, для которого $KS(y|x) = x + O(1)$.]

$\boxed{28}$ Покажите, что изменение одного бита в слове длины n меняет его сложность не более чем на $\log n + O(\log\log n)$.

$\boxed{29}$ `[number-of-descriptions]` Фиксируем некоторый (безусловный) способ описания D. Докажите, что для некоторой константы c и для всех n и k выполнено такое свойство: если какое-то слово x имеет 2^k описаний длины не более n, то $KS(x|k) \leqslant n - k + c$. [Указание. Пусть k фиксировано. Для каждого n рассмотрим слова x, имеющие не менее 2^k описаний длины не более n. Число таких слов (при данном k) не превосходит 2^{n-k}, и можно воспользоваться теоремой 19, с. 41.]

С помощью этой задачи можно доказать такое утверждение о безусловной сложности:

$\boxed{30}$ Пусть D — фиксированный (безусловный) способ описания. Тогда найдётся такое число c, что для любого слова x число кратчайших D-описаний слова x не превосходит c. [Указание. В условиях предыдущей задачи $KS(x) \leqslant n - k + 2\log k + O(1)$, поэтому при $KS(x) = n$ значение k ограничено.]

$\boxed{31}$ Докажите, что найдётся константа c с таким свойством: если для данных x и n вероятность события $KS(x|y) \leqslant k$ (для случайно взятого слова y длины n; все такие слова считаем равновероятными) не меньше 2^{-l}, то $KS(x|n, l) \leqslant k + l + c$. [Указание. Соединим каждое слово y длины n со всеми словами x, сложность которых относительно y не превосходит k, получим двудольный граф с $O(2^{n+k})$ рёбрами, и в нём число вершин x, из которых выходит не менее 2^{n-l} рёбер, есть $O(2^{k+l})$. Обратите внимание, что в $KS(x|n, l)$ не входит k — это не опечатка!]

Эта задача позволяет ответить на такой вопрос: чему в среднем равна сложность $KS(x|y)$ для данного x и случайно выбранного слова y данной длины n? Ясно, что $KS(x|y) \leqslant K(x|n) + O(1)$ (поскольку $n = l(y)$ восстанавливается по y). Оказывается, что для большинства слов y данной длины эта оценка точная:

$\boxed{32}$ Докажите, что найдётся такая константа c, что для любого слова x и для любых натуральных чисел n и d доля тех слов y длины n, у которых $KS(x|y) < KS(x|n) - d$ (среди всех слов длины n), не превосходит $cd^2/2^d$. Выведите отсюда, что среднее арифметическое $KS(x|y)$ по всем словам y данной длины n равно $KS(x|n) + O(1)$ (константа в $O(1)$ не зависит от x и n).

$\boxed{33}$ Докажите, что $KS(x) = KS(x|\,KS(x)) + O(1)$. [Указание. Пусть x имеет короткое описание q при известном $KS(x)$. Тогда для восстановления x достаточно задать q и разницу длин $KS(x) - l(q)$, и получается более короткое описание слова x, чем это возможно.]

$\boxed{34}$ [increasing-plain-complexity] Докажите, что (для некоторой константы c) для любого слова x и любого числа n найдётся такое слово y длины n, что

$$KS(xy) \geqslant KS(x|n) + n - c.$$

[Указание. Для данного n количество слов x, при которых $KS(xy) < k$ при всех y длины n, не превосходит $2^k/2^n$, и это свойство перечислимо, так что можно применить теорему 19 (с. 41).]

$\boxed{35}$ Докажите, что бесконечная последовательность $x_0x_1x_2\ldots$ нулей и единиц вычислима тогда и только тогда, когда величина $KS(x_0 \ldots x_{n-1}|n)$ (условная сложность её начальных отрезков при известной длине) ограничена сверху.

[Указание. Отметим в бесконечном двоичном дереве всех слов перечислимое множество S вершин (слов), у которых условная сложность при известной длине ограничена некоторой фиксированной константой. Все «горизонтальные» сечения множества S ограничены по мощности. Нам нужно вывести из этого, что любая бесконечная ветвь, целиком проходящая по S, вычислима. Будем предполагать, что S образует поддерево (оставив от него только те вершины, для которых путь из корня также лежит в S.) Фиксируем некоторую бесконечную ветвь ω в поддереве S и на каждом уровне n подсчитаем числа l_n и r_n вершин из S слева и справа от ω. Пусть L и R — верхние пределы чисел l_n и r_n, и пусть N — номер уровня, начиная с которого эти верхние пределы не превышаются. Зная L, R и N, можно вычислять сколь угодно большие начальные отрезки ветви ω: надо искать ветви, слева от которых есть не менее L элементов из S на некотором уровне (большем N) и справа есть не менее R элементов из S на некотором (возможно, другом, также большем N) уровне; как только такая ветвь нашлась, её начальный отрезок до нижнего из этих двух уровней совпадает с ω.]

$\boxed{36}$ Докажите, что в предыдущей задаче условие ограниченности $KS(x_0 \ldots x_{n-1}|n)$ можно заменить более слабым условием: $KS(x_0 \ldots x_{n-1}) \leqslant \log n + c$ для некоторого c и для всех n. [Указание: возникает перечислимое множество S слов (вершин бесконечного двоичного дерева), в котором число вершин на уровнях ниже N есть $O(N)$. Если бы числа вершин на всех уровнях были бы ограничены, то задача свелась бы к предыдущей. Этого можно добиться, оставив от множества только те вершины x, для которых существует продолжение длины $2l(x)$, все начала которого принадлежат S.]

$\boxed{37}$ Рассмотрим слова длины n, у которых сложность не меньше n. (Такие слова естественно назвать *несжимаемыми*.) (**а**) Покажите, что количество таких слов не меньше 2^{n-c} и не больше $2^n - 2^{n-c}$ (для некоторого c). (**б**) Покажите, что сложность количества несжимаемых слов длины n равна $n - O(1)$ (отсюда следуют утверждения предыдущего пункта!). (**в**) Покажите, что если слово x длины $2n$ является несжимаемым, то его половины x_1 и x_2 (длины n) имеют сложность $n - O(1)$. (г) Покажите, что если слово x длины n является несжимаемым, то любое его подслово длины k имеет сложность не меньше $k - O(\log n)$. (д) Покажите, что для любой константы $c < 1$ любое несжимаемое слово достаточно большой длины n содержит подслово из $\lfloor c \log_2 n \rfloor$ нулей. [Указание. (а) Всего описаний длины меньше n будет $2^n - 1$, но часть их них уходит на более короткие слова: любое слово длины $n - d$ при некотором d имеет сложность меньше n, что доказывает первое утверждение. Чтобы доказать второе, заметим, что слова длины n, начинающиеся с k нулей, можно задать $2 \log k + (n - k)$ битами. (б) Запишем число несжимаемых слов длины n в виде двоичного слова t; если это слово имеет длину $n - k$, то по t и $\log k$ битов можно восстановить и n, и список всех несжимаемых слов длины n, поэтому первое из несжимаемых слов будет иметь сложность меньше чем следует. (в) Если одну из половин можно описать короче, то и всё слово можно задать короче, начав с (беспрефиксного кода) разности между длиной и сложностью сжимаемой половины. (г) Если подслово имеет меньшую сложность, то и всё слово можно описать короче (задав подслово, его положение и остальные биты). (д) Подсчитаем число слов длины n, не содержащих подряд k нулей; рекуррентное соотношение показывает, что это число растёт с увеличением n примерно как геометрическая прогрессии, знаменатель которой есть наибольший действительный корень уравнения $x = 2 - (1/x^k)$, и можно оценить сложность таких слов.]

$\boxed{38}$ `[condit-c-simple-prefix]` Покажите, что (при некотором значении c) для любой бесконечной последовательности $x_0x_1x_2\ldots$ нулей и единиц найдётся бесконечно много значений n, для которых $KS(x_0x_1\ldots x_{n-1}) \leqslant n - \log n + c$.

Покажите, что найдётся последовательность $x_0x_1x_2\ldots$ и константа c, для которых $KS(x_0x_1\ldots x_{n-1}) \geqslant n - 2\log n - c$ при всех n.

[Указание. Ряд $\sum 1/n$ расходится, а ряд $\sum 1/n^2$ сходится. Подробнее см. теоремы 87 и 91.]

$\boxed{39}$ `[conditional-unconditional-def]` Для данного слова x длины n определим величины $d(x) = n - KS(x)$ и $d_c(x) = n - KS(x|n)$. Покажите, что они связаны неравенством

$$d_c(x) - 2\log d_c(x) - O(1) \leqslant d(x) \leqslant d_c(x) + O(1).$$

[Указание. Надо доказать, что если $KS(x|n) = n - d$, то $KS(x) \leqslant n - d + 2\log d + O(1)$. В самом деле, взяв условное описание p длины $n - d$ и добавив в его начало самоограниченную запись d, мы получим слово, по которому можно восстановить сначала d, потом n, и наконец x.]

(Интуитивный смысл разности между длиной слова и его сложностью как дефекта случайности обсуждается для разных видов сложности в главе 5 и в главе 16.)

2.3. Количество информации

`[conditi]`

Мы знаем (теорема 18), что условная сложность $KS(y|x)$ не превосходит безусловной сложности $KS(y)$ (с точностью до константы). Разность $KS(y) - KS(y|x)$ показывает, насколько знание слова x упрощает описание слова y. Поэтому её называют *количеством информации в слове x о слове y*. Обозначение: $I(x:y)$.

Теорема 18 показывает, что информация $I(x : y)$ неотрицательна (с точностью до константы): существует такое c, что $I(x : y) \geqslant c$ при всех x и y.

Вспомнив, что

$$KS(x, y) = KS(x) + KS(y|x) + O(\log KS(x, y)),$$

(теорема 22, с. 43), можно выразить условную сложность через безусловные: $KS(y|x) = KS(x, y) - KS(x) + O(\log KS(x, y))$. Тогда для информации получается выражение:

$$I(x : y) = KS(y) - KS(y|x) = KS(x) + KS(y) - KS(x, y) + O(\log KS(x, y)).$$

Отсюда сразу вытекает

Theorem 23 (симметрия информации). [condit-symm]

$$I(x : y) = I(y : x) + O(\log KS(x, y))$$

Эта теорема гарантирует, что разница между $I(x : y)$ и $I(y : x)$ логарифмически мала по сравнению с $KS(x, y)$. Как показывает следующая задача, эта разница может быть сравнима с самими значениями $I(x : y)$ и $I(y : x)$, если те много меньше $KS(x, y)$.

40 Покажите, что если x — слово длины n, для которого $KS(x|n) \geqslant n$, то $I(x : n) = KS(n) + O(1)$, в то время как $I(n : x) = O(1)$.

Симметрия (пусть и не полная, а лишь с точностью до логарифмического слагаемого) позволяет называть $I(x : y)$ (или $I(y : x)$) *взаимной информацией* слов x и y. Соотношения между взаимной информацией, условными сложностями и сложностью пары можно изобразить на символической картинке (рис. 3).

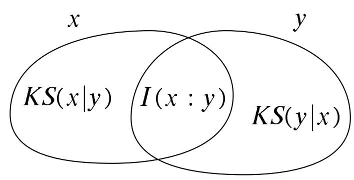

Рис. 3. Взаимная информация и условная сложность

[condit-i.1]

На ней показано, что слова x и y имеют $I(x : y) \approx I(y : x)$ битов общей информации. Добавив $KS(x|y)$ битов (информация, которая есть в x, но не в y, левая область), мы получаем

$$I(y : x) + KS(x|y) \approx (KS(x) - KS(x|y)) + KS(x|y) \approx KS(x)$$

битов (как и должно быть в слове x). Аналогичным образом центральная часть вместе с $KS(y|x)$ справа дают $KS(y)$. Наконец, все три области вместе складываются в

$$KS(x|y) + I(x : y) + KS(y|x) = KS(x) + KS(y|x) = KS(x|y) + KS(y) = KS(x, y)$$

(все равенства верны с точностью $O(\log n)$, если слова x и y имеют длину не больше n).

В некоторых случаях эту картинку можно понимать буквально. Возьмём, например, «несжимаемое» слово $r = r_1 \ldots r_n$ длины n, для которого $KS(r_1 \ldots r_n) \geqslant n$. Тогда сложность любого его подслова u равна $l(u)$ с точностью до $O(\log n)$. В самом деле, поскольку u является подсловом r, то $r = tuv$ для некоторых слов t и v. Тогда $l(r) = KS(r) \leqslant KS(t) + KS(u) + KS(v) \leqslant l(t) + l(u) + l(v) = l(r)$ (с логарифмической точностью) и потому все неравенства обращаются в равенства (с той же точностью).

Если теперь взять два перекрывающихся подслова x и y (рис. 4), то $KS(x)$ будет равно длине x, $KS(y)$ будет равно длине y (с точностью до $O(\log n)$).

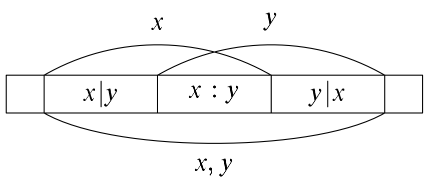

Рис. 4. Общая информация в перекрывающихся подсловах
[condit-i.2]

Сложность пары $KS(x, y)$ будет равна длине объединения отрезков (поскольку пара x, y отличается от этого объединения лишь информацией о длинах, которая требует $O(\log n)$ битов).

Следовательно, и условные сложности $KS(x|y)$, $KS(y|x)$, и общая информация $I(x : y)$ равны (с логарифмической точностью) длинам соответствующих отрезков.

Однако такая ситуация имеет место далеко не всегда. Как мы впоследствии убедимся (см. главу 11), далеко не для всяких двух слов x и y, имеющих большу́ю взаимную информацию $I(x : y)$, эту взаимную информацию можно «материализовать» в виде слова z, для которого $KS(z|x) \approx 0$, $KS(z|y) \approx 0$ и $KS(z) \approx I(x : y)$. (В нашем последнем примере таким z было пересечение подслов x и y.)

41 Докажите, что для любого слова x длины не больше n математическое ожидание количества общей информации в x и случайном слове длины n есть $O(\log n)$.

Рассмотрим теперь сложностные характеристики трёх слов. Здесь важным приёмом является *релятивизация* — перенесение свойств безусловной сложности на условную. Поясним это на примере.

Теорема о сложности пары (с. 36) утверждает, что $KS(x, y) \leqslant KS(x) + 2 \log KS(x) + KS(y) + O(1)$. Если все сложности заменить на условные (при известном z), получится неравенство

$$KS(x, y|z) \leqslant KS(x|z) + 2 \log KS(x|z) + KS(y|z) + O(1),$$

где под условной сложностью пары x, y при известном z понимается сложность её кода: $KS(x, y|z) = KS([x, y]|z)$. Как и для безусловной сложности пары, выбор кодирования несуществен (меняет сложность на $O(1)$).

Это неравенство доказывается точно так же, как и неравенство без z: мы соединяем описания p для x (при известном z) и q для y (при известном z) в единое описание $\hat{p}q$ относительно нового способа (условного) описания.

Поучительно исключить из этого неравенства условные сложности, заменив их на безусловные: $KS(x, y|z) = KS(x, y, z) - KS(z)$, а также $KS(x|z) = KS(x, z) - KS(z)$ и

$KS(y|z) = KS(y,z) - KS(z)$ (с логарифмической точностью). При этом мы приходим к такой теореме:

Theorem 24. `[condit-baseineq]`

$$KS(x,y,z) + KS(z) \leqslant KS(x,z) + KS(y,z) + O(\log n)$$

для слов x, y, z сложности не более n.

Это неравенство иногда называют *базисным*.

Релятивизацию можно применить и к теореме 21 (с. 42), связывающей сложность пары и условную сложность. Получается такое утверждение (мы формулируем его с логарифмической точностью):

Theorem 25. `[condit-pair4]`

$$KS(x,y|z) = KS(x|z) + KS(y|x,z) + O(\log n),$$

если слова x, y, z имеют сложность не больше n.

Здесь $KS(x|y,z)$ понимается как сложность x относительно кода пары $\langle y, z\rangle$, то есть как $KS(x|[y,z])$. Как и раньше, вычислимое кодирование пар может быть любым (сложность изменится не более чем на константу).

$\lhd$ Можно воспроизвести доказательство теоремы 21, заменив везде описания без условий на описания при известном z. При этом $KS(y|x)$ превратится в $KS(y|x,z)$. Можно сказать, что мы теперь имеем трёхмерное пространство с координатами x, y, z и проводим все прежние рассуждения одновременно во всех плоскостях, параллельных плоскости xy.

Если это кажется слишком сложным, можно просто выразить все условные сложности через безусловные: в левой части получим

$$KS(x,y|z) = KS(x,y,z) - KS(z),$$

а в правой

$$KS(x|z) + KS(y|x,z) = KS(x,z) - KS(z) + KS(y,x,z) - KS(x,z),$$

так что после сокращения левая и правая части становятся равными (с логарифмической точностью). (Отметим в скобках, что это более простое рассуждение даёт худшие значения констант в $O(\log n)$-обозначениях.) $\rhd$

$\boxed{42}$ Проверьте, что в теореме 25 достаточно потребовать, чтобы условные сложности $KS(x|z)$ и $KS(y|x,z)$ не превосходили n.

Можно «релятивизовать» определение информации, определив $I(x:y|z)$ как разность $KS(y|z) - KS(y|x,z)$. Как и для случая безусловной информации, эта величина неотрицательна (с точностью до $O(1)$). Выразив условную сложность через безусловную (с логарифмической точностью), можно записать неотрицательность $I(x:y|z)$ как

$$KS(y|z) - KS(y|x,z) = KS(y,z) - KS(z) - KS(y,x,z) + KS(x,z) \geqslant 0,$$

приходя к неравенству теоремы 24.

Вообще большинство известных равенств и неравенств, касающихся безусловных сложностей, условных сложностей и информации (с логарифмической точностью), являются прямыми следствиями теорем 21 и 24. Приведём два примера такого рода.

Независимые слова. Будем называть слова x и y «независимыми», если $I(x:y) \approx 0$. (Степень точности должна ещё уточняться, но мы предполагаем, что членами порядка $O(\log n)$ мы пренебрегаем, если n — максимальная из длин (или сложностей) используемых слов.)

Независимые слова можно рассматривать как аналог независимых случайных величин в теории вероятностей. Есть такая теорема: если случайная величина ξ независима с парой величин $\langle\alpha,\beta\rangle$, то она независима и с величинами α и β по отдельности.

Для колмогоровской сложности аналогичное утверждение (если слово x независимо с парой $\langle y,z\rangle$, то оно независимо и с её отдельными компонентами) естественно выражается неравенством

$$I(x:\langle y,z\rangle) \geqslant I(x:y)$$

(и аналогичным неравенством для z вместо y). Это неравенство справедливо (как обычно, с логарифмической точностью), и в этом легко убедиться, переписав его в терминах безусловных сложностей:

$$\mathit{KS}(x)+\mathit{KS}(y,z)-\mathit{KS}(x,y,z) \geqslant \mathit{KS}(x)+\mathit{KS}(y)-\mathit{KS}(x,y),$$

что после приведения подобных членов даёт базисное неравенство теоремы 24.

Сложность пар и троек. Напротив, для доказательства следующей теоремы (упоминавшейся на с. 19) полезно заменить безусловные сложности на условные:

Theorem 26. [condit-triple]

$$2\,\mathit{KS}(x,y,z) \leqslant \mathit{KS}(x,y)+\mathit{KS}(x,z)+\mathit{KS}(y,z)+O(\log n),$$

если x,y,z — слова сложности не более n.

$\triangleleft$ Перенося $\mathit{KS}(x,y)$ и $\mathit{KS}(x,z)$ в левую часть, и заменяя разности $\mathit{KS}(x,y,z)-\mathit{KS}(x,y)$ и $\mathit{KS}(x,y,z)-\mathit{KS}(x,z)$ на условные сложности $\mathit{KS}(z|x,y)$ и $\mathit{KS}(y|x,z)$, мы получаем такое неравенство:

$$\mathit{KS}(z|x,y)+\mathit{KS}(y|x,z) \leqslant K(y,z)+O(\log n).$$

Остаётся переписать правую часть в виде $\mathit{KS}(y)+\mathit{KS}(z|y)$, и заметить, что $\mathit{KS}(z|x,y) \leqslant \mathit{KS}(z|y)$ и $\mathit{KS}(y|x,z) \leqslant \mathit{KS}(y)$ (с логарифмической точностью). $\triangleright$

А можно вместо этого формально сложить два неравенства (базисное и для сложности пары):

$$\mathit{KS}(x,y,z)+\mathit{KS}(y) \leqslant \mathit{KS}(x,y)+\mathit{KS}(y,z)+O(\log n),$$
$$\mathit{KS}(x,y,z) \leqslant \mathit{KS}(y)+\mathit{KS}(x,z)+O(\log n),$$

после чего сократить $\mathit{KS}(y)$ в обеих частях и получить требуемое. (К сожалению, это доказательство, как и предыдущее, несимметрично.) Мы вернёмся к этому неравенству и его геометрическим следствиям в главе 10.

Различные сложностные характеристики трёх слов можно систематизировать следующим образом. Имеются семь основных таких характеристик (три сложности одиночных слов, три сложности пар и сложность всей тройки). Другие характеристики (условная сложность, информация) выражаются через них. Чтобы лучше представить себе, какие ограничения накладываются на эти семь сложностей, удобно перейти к новым координатам. Введём переменные $a_1, a_2, \ldots, a_7$, соответствующие семи областям на рис. 5.

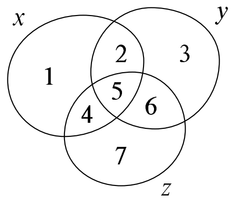

Рис. 5. Новые координаты $a_1, a_2, \ldots, a_7$.

[condit-i.3]

Будем считать, что

$$\begin{aligned}
KS(x) &= a_1 + a_2 + a_4 + a_5,\\
KS(y) &= a_2 + a_3 + a_5 + a_6,\\
KS(z) &= a_4 + a_5 + a_6 + a_7,\\
KS(x,y) &= a_1 + a_2 + a_3 + a_4 + a_5 + a_6,\\
KS(x,z) &= a_1 + a_2 + a_4 + a_5 + a_6 + a_7,\\
KS(y,z) &= a_2 + a_3 + a_4 + a_5 + a_6 + a_7,\\
KS(x,y,z) &= a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7.
\end{aligned}$$

Легко проверить, что эти равенства задают обратимое линейное преобразование семимерного пространства (каждому набору из семи сложностей соответствует единственное значение переменных $a_1, \ldots, a_7$).

Различные условные сложности и взаимные информации (с логарифмической точностью) выражаются через безусловные, поэтому можно их выразить и в новых координатах. Например, $I(x:y) = KS(x) + KS(y) - KS(x,y) = a_2 + a_5$, а $KS(x|y) = KS(x,y) - KS(y) = a_1 + a_4$.

Каков смысл наших новых координат? Легко проверить, что $a_1 = KS(x|y,z)$ (с логарифмической точностью). Аналогичный смысл имеют a_3 и a_7. Координата a_2 есть (с той же точностью) $I(x:y|z)$; аналогичный смысл имеют a_4 и a_6 (см. рис. 6). Отсюда, в частности, вытекает, что для любых слов x, y, z соответствующие значения координат $a_1, a_2, a_3, a_4, a_6, a_7$ неотрицательны (с точностью до $O(\log n)$, если слова x, y, z имеют сложность не больше n).

Ситуация с координатой a_5 сложнее. Её хотелось бы интерпретировать как «общую информацию, содержащуюся в трёх словах x, y, z». Иногда для неё используется обозначение $I(x:y:z)$. Однако смысл такого выражения далеко не очевиден, особенно если иметь в виду, что a_5 может быть отрицательным. Рассмотрим такой пример. Пусть x и y — две

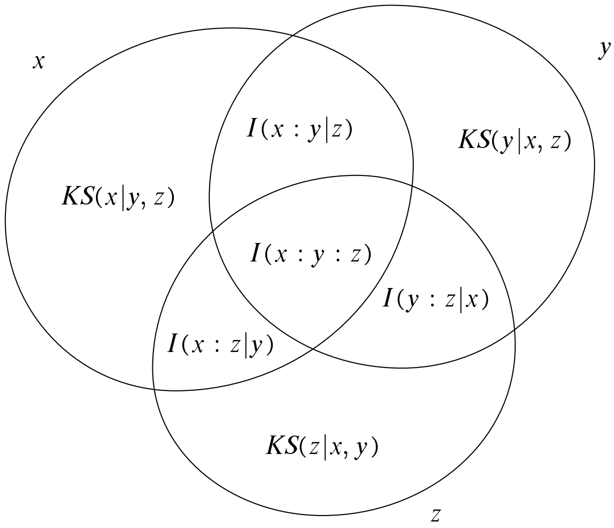

Рис. 6. Сложностной смысл новых координат.
[condit-i.5]

половины несжимаемого слова длины $2n$. Тогда $KS(x) = n$, $KS(y) = n$, $KS(x, y) = 2n$ и $I(x : y) = 0$ (с логарифмической точностью). Рассмотрим слово z длины n, которое является побитовой суммой x и y по модулю 2. Тогда любое из трёх слов x, y, z может быть восстановлено по двум другим, следовательно, сложности $KS(x, y), KS(y, z)$ и $KS(x, z)$ одинаковы и равны $2n$ (с логарифмической точностью), и сложность $KS(x, y, z)$ также равна $2n$. Сложность слова z равна n (она не больше n, так как его длина равна n, но и не меньше, так как иначе вместе с y не получилось бы пары сложности $2n$).

После этого можно вычислить значения $a_1, \ldots, a_7$ для этого примера (рис. 7):

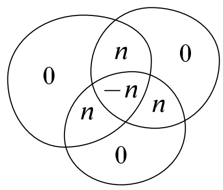

Рис. 7. Два независимых несжимаемых слова длины n и их сумма $\bmod 2$.
[condit-i.4]

Таким образом координата a_5 может быть отрицательной, однако суммы $a_5 + a_2$, $a_5 + a_4$ и $a_5 + a_6$, будучи взаимными информациями пар слов, должны быть неотрицательны. (В нашем примере эти суммы равны нулю.)

Этот пример соответствует простейшему способу «разделения секрета» z между двумя лицами: если одному человеку сообщить x, а другому y, то ни один из них по отдельности не будет ничего знать о z (поскольку $I(x : z) \approx 0$ и $I(y : z) \approx 0$), но вместе они могут восстановить z, побитово сложив x и y.

Можно проверить, что указанными неравенствами (неотрицательность всех a_i, кроме a_5, и неотрицательность трёх названных сумм) исчерпываются все линейные неравенства, имеющиеся для сложностных характеристик трёх слов. Мы вернёмся к этому вопросу в главе 10.

В качестве примера использования таких диаграмм докажем снова неравенство

$$2\,KS(x,y,z) \leqslant KS(x,y) + KS(x,z) + KS(y,z).$$

В наших новых переменных оно записывается как $a_2+a_4+a_5+a_6 \geqslant 0$ (это легко проверить в уме, подсчитав, сколько раз входит каждое из a_i в левую и правую часть неравенства). Теперь остаётся заметить, что, $a_2+a_5 \geqslant 0$, $a_4 \geqslant 0$ и $a_6 \geqslant 0$.

43 Докажите, что $I(xy:z) = I(x:z)+I(y:z|x)+O(\log n)$ для слов x,y,z сложности не более n. [Указание: это легко объяснить на диаграмме.]

Эта задача показывает, что информация в xy о z складывается из двух частей: информации в x о z и информации в y о z (при известном x). Можно рассматривать её как аналог равенства $KS(x,y) = KS(x)+KS(y|x)$, только теперь вместо сложности стоит информация о слове z. В качестве следствия получаем, что если xy независимо с z, то x независимо с z, а также y независимо с z при известном x. (Под независимостью понимается малость взаимной информации.) Симметричное рассуждение показывает, что x независимо с y, а также x независимо с z при известном y.

44 Покажите, что свойства «x и y независимы» и «x и y независимы при известном z» мало связаны друг с другом: любое из них может быть истинно, в то время как второе ложно.

45 Будем говорить, что слова x,y,z,t образуют *марковскую цепь* (по аналогии с соответствующим понятием теории вероятностей), если $I(x:z|y)$ и $I(\langle x,y\rangle : t|z)$ малы (разумеется, точный смысл определения зависит от выбора границы малости). Покажите, что в этом случае эти же слова в обратном порядке образуют марковскую цепь, то есть что $I(t:y|z)$ и $I(\langle t,z\rangle : x|y)$ малы. [Указание. $I(\langle x,y\rangle : t|z) = I(y:t|z) + I(x:t|y,z)$, и равенство нулю левой части означает обращение в нуль обоих слагаемых справа; второе слагаемое симметрично относительно обращения порядка.]

3. Случайность по Мартин-Лёфу

[random]

В этой главе мы прерываем изложение основных свойств колмогоровской сложности, чтобы определить другое базисное понятие алгоритмической теории информации — понятие случайной по Мартин-Лёфу, или «типичной», последовательности. (Материал этой главы не использует предыдущей и не используется в следующей; он понадобится в главе 5, где будет дан критерий случайности в терминах колмогоровской сложности.)

Мы начнём с напоминания основных фактов о мерах на последовательностях нулей и единиц.

3.1. Пространство Ω и меры

[randomcl]

Рассмотрим пространство Ω, элементами которого являются бесконечные последовательности нулей и единиц. Его называют *канторовским* пространством. Для каждого двоичного слова x мы рассмотрим множество Ω_x всех продолжений этого слова. Например, Ω_{00} есть множество всех последовательностей, начинающихся с двух нулей, а $\Omega_\Lambda = \Omega$ (здесь Λ обозначает пустое слово).

Множества вида Ω_x мы будем называть *интервалами*. Интервалы и всевозможные их объединения называют *открытыми* подмножествами пространства Ω. Эта топология соответствует стандартной метрике, в которой расстояние между двумя последовательностями тем меньше, чем больше у них совпадающее начало:

$$d(\omega, \omega') = 2^{-n},$$

где n — наименьшее число, для которого $\omega_n \neq \omega'_n$. (Через ω_n мы обозначаем член с номером n, начиная нумерацию с нуля: $\omega = \omega_0\omega_1\omega_2\ldots$)

46 Покажите, что это пространство гомеоморфно канторовскому множеству, которое получится, если из отрезка выбросить среднюю треть, из двух оставшихся частей также выбросить по средней трети и так далее.

Нас, однако, будет интересовать не столько топология, сколько мера. Семейство подмножеств пространства Ω называют *σ-алгеброй*, если оно замкнуто относительно конечных и счётных пересечений и объединений, а также перехода к дополнению.

Минимальную σ-алгебру, содержащую все множества Ω_x (и тем самым все открытые множества), называют алгеброй *борелевских* множеств.

Рассмотрим произвольную σ-алгебру, содержащую все интервалы. Пусть на ней задана функция μ, которая ставит в соответствие каждому множеству из σ-алгебры неотрицательное число, причём выполнено свойство *σ-аддитивности*:

> если множество A есть объединение конечного или счётного числа непересекающихся множеств $A_0, A_1, A_2, \ldots$, и все эти множества принадлежат σ-алгебре, на которой определена функция μ, то

$$\mu(A) = \mu(A_0) + \mu(A_1) + \mu(A_2) + \ldots$$

(в правой части стоит конечная сумма или сходящийся бесконечный ряд с неотрицательными членами).

Такие функции называют *мерами* на пространстве Ω, а $\mu(A)$ называют мерой множества A.

Если мера всего пространства Ω равна 1, то меру называют *распределением вероятностей*, элементы σ-алгебры называют *событиями*, а число $\mu(A)$ называют *вероятностью* события A.

Любая мера монотонна (если $A \subset B$, то $\mu(A) \leqslant \mu(B)$). В самом деле, $\mu(B) - \mu(A) = \mu(B \setminus A) \geqslant 0$.

Другое важное свойство мер — непрерывность: если множество B есть объединение возрастающей последовательности множеств

$$B_0 \subset B_1 \subset B_2 \subset \ldots,$$

то $\mu(B)$ равно пределу $\mu(B_i)$ при $i \to \infty$. (В самом деле, применяем счётную и конечную аддитивность к множествам $A_i = B_i \setminus B_{i-1}$.) Аналогичное свойство непрерывности верно и для убывающих последовательностей множеств.

Для каждой меры μ на Ω можно рассмотреть функцию p на двоичных словах, определённую равенством

$$p(x) = \mu(\Omega_x).$$

Эта функция принимает неотрицательные значения, при этом выполнено свойство *аддитивности*:

$$p(x) = p(x0) + p(x1)$$

для любого слова x. (В самом деле, интервал Ω_x есть непересекающееся объединение двух его половин Ω_{x0} и Ω_{x1}.)

Теория меры (теорема о продолжении меры по Лебегу) позволяет выполнить и обратный переход. Именно, для каждой функции p на двоичных словах, принимающей неотрицательные действительные значения и обладающей свойством аддитивности, можно построить меру μ, для которой $\mu(\Omega_x) = p(x)$ при всех x.

Получаемая при этом построении мера обладает дополнительным свойством: если $\mu(A) = 0$ и $B \subset A$, то B обязательно измеримо (откуда уже следует, что $\mu(B) = 0$). В дальнейшем мы будем рассматривать только меры, обладающие этим свойством.

Мы не приводим конструкции продолжения меры по Лебегу (она есть в любом учебнике по теории меры, например, [22] или [15]), но приведём используемое в ней явное описание множеств меры нуль.

Пусть фиксирована неотрицательная функция p на словах, обладающая свойством аддитивности. Будем называть число $p(x)$ мерой интервала Ω_x. Говорят, что множество $A \subset \Omega$ является *нулевым*, если для всякого $\varepsilon > 0$ можно найти конечное или счётное покрытие множества A интервалами с суммой мер не больше ε.

Другими словами, множество A имеет меру нуль, если существует функция $\langle \varepsilon, i \rangle \mapsto x(\varepsilon, i)$ (первый аргумент — положительное действительное число, второй — натуральное число, значениями являются двоичные слова), для которой

- $A \subset \Omega_{x(\varepsilon,0)} \cup \Omega_{x(\varepsilon,1)} \cup \Omega_{x(\varepsilon,2)} \ldots$ и
- $p(x(\varepsilon, 0)) + p(x(\varepsilon, 1)) + p(x(\varepsilon, 2)) + \ldots \leqslant \varepsilon$

при любом положительном ε. При этом возможность конечного покрытия мы учитываем, не требуя, чтобы $x(\varepsilon, i)$ было определено при всех ε и i (неопределённые значения пропускаются в объединении и в сумме).

Несколько простых наблюдений, которые нам пригодятся в дальнейшем:

- В определении нулевого множества можно было бы ограничиться рациональными значениями ε (или только значениями вида 2^{-k}).

- Всякое подмножество нулевого множества является нулевым.

- Конечное или счётное объединение нулевых множеств является нулевым. (В самом деле, чтобы найти покрытие объединённого множества с суммой мер меньше ε, достаточно соединить покрытия его частей с суммой мер меньше $\varepsilon/2, \varepsilon/4, \varepsilon/8, \ldots$)

- Пусть функция p такова, что каждая точка имеет меру нуль (это будет так, если для любой бесконечной последовательности $\omega = \omega_0\omega_1\omega_2\ldots$ предел мер $p(\omega_0\ldots\omega_n)$ при $n \to \infty$ равен нулю). Тогда любое конечное или счётное множество является нулевым.

Равномерная мера на пространстве Ω получится, если положить меру интервала Ω_x равной $2^{-l(x)}$:

$$p(x) = 2^{-n} \text{ для всех слов } x \text{ длины } n.$$

Возникающая при этом мера тесно связана с обычной мерой на действительной прямой (точнее, на отрезке $[0,1]$): мера множества $A \subset \Omega$ равна мере множества действительных чисел, которое получится, если каждую последовательность из A считать бесконечной двоичной дробью. (Замечание в скобках: соответствие между двоичными дробями и числами не совсем взаимно однозначно, так как двоично-рациональные числа имеют два представления. Например, $0{,}01111\ldots = 0{,}10000\ldots$ Но это затрагивает лишь счётное множество чисел, которое пренебрежимо с точки зрения теории меры.) В самом деле, числа, двоичные записи которых начинаются на данную строку x длины n, заполняют промежуток как раз длины 2^{-n}. Отсюда легко заключить, что для любого отрезка $I \subset [0,1]$ равномерная мера множества последовательностей, являющихся двоичными разложениями чисел из I, равна длине отрезка I.

С точки зрения теории вероятностей равномерная мера соответствует независимым бросаниям честной монеты. В самом деле, при независимых бросаниях честной монеты все возможные последовательности длины n равновероятны и имеют вероятность 2^{-n}. Множество Ω_x есть событие «полученная в результате бросаний последовательность начинается на x» и имеет меру $2^{-l(x)}$.

Можно рассматривать и несимметричную монету, по-прежнему предполагая бросания независимыми. Соответствующая мера (распределение вероятностей) называется *бернуллиевой*[nonuniform-bernoulli] с параметрами q, p (вероятности выпадения нуля и единицы; предполагается, что $p, q \geqslant 0$ и $p + q = 1$).

При этом вероятность появления последовательности, начинающейся на слово x, равна $q^u p^v$, где u и v — число нулей и единиц в этом слове. Другими словами, мы рассматриваем функцию

$$x \mapsto q^{u(x)} p^{v(x)}$$

где через $u(x)$ и $v(x)$ обозначено число нулей и единиц в слове x. Легко проверить, что эта функция аддитивна.

3.2. Усиленный закон больших чисел

[random-lln]

Чтобы проиллюстрировать все введённые понятия в действии, сформулируем и докажем *усиленный закон больших чисел.*

Пусть $p + q = 1$, $p, q \geqslant 0$. Рассмотрим множество A_p всех бесконечных последовательностей $\omega_0\omega_1\omega_2\ldots$ нулей и единиц, в которых предел частоты единиц существует и равен p, то есть

$$\lim_{n\to\infty} \frac{\omega_0 + \omega_1 + \ldots + \omega_{n-1}}{n} = p.$$

Theorem 27. [randomcl-large-numbers] *Множество A_p имеет меру 1 относительно бернуллиева распределения вероятности (с параметрами p и q).*

Другими словами, дополнение множества A_p, то есть множество тех последовательностей, в которых частота единиц не стремится ни к какому пределу или имеет предел, отличный от p, является нулевым.

$\lhd$ Мы докажем эту теорему для равномерной меры (то есть при $p = q = 1/2$) с помощью явного подсчёта. Для произвольного p доказательство будет кратко намечено в одном из упражнений (см. также раздел 9.6).

Сначала рассмотрим «допредельную» ситуацию, зафиксировав некоторое число n. Все последовательности из n нулей и единиц равновероятны, и нам нужно доказать, что большинство из них содержит примерно $n/2$ единиц. Пусть выбрана некоторая граница $\varepsilon > 0$. Посмотрим, сколько последовательностей имеет более $(1/2+\varepsilon)n$ единиц. Другими словами, мы должны просуммировать в треугольнике Паскаля часть n-й строки, которая начинается несколько правее середины. Всего в этой части не более n слагаемых и они убывают, поэтому сумму можно оценить как первое слагаемое, умноженное на n. (На самом деле нам не нужна бóльшая точность, и полиномиальные от n множители нам не повредят, так что ими мы сразу пренебрегаем. В частности, мы опускаем множитель n.)

Первое слагаемое равно

$$\frac{n!}{k!(n-k)!},$$

где k — ближайшее справа к $(1/2+\varepsilon)n$ целое число. Воспользуемся формулой Стирлинга:

$$m! = \sqrt{(2\pi + o(1))m}\left(\frac{m}{e}\right)^m,$$

где e — основание натуральных логарифмов. Отбрасывая полиномиальные множители и используя обозначения $u = k/n$, $v = (n-k)/n$, получаем

$$\frac{n!}{k!(n-k)!} \approx \frac{(n/e)^n}{(k/e)^k((n-k)/e)^{n-k}} = \frac{n^n}{k^k(n-k)^{n-k}} \approx$$
$$\approx \frac{n^n}{(un)^{un}(vn)^{vn}} = \frac{1}{u^{un}v^{vn}} = 2^{H(u,v)n},$$

где

$$H(u,v) = -u\log u - v\log v.$$

Число $H(u,v)$ называется *шенноновской энтропией* случайной величины с двумя значениями, принимаемыми с вероятностями u и v. (Подробно шенноновская энтропия обсуждается в главе 7.) На рисунке 8 показан соответствующий график (напомним, $v = 1-u$). Легко

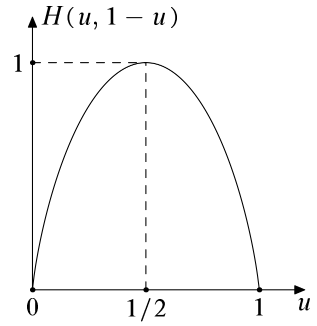

Рис. 8. Шенноновская энтропия как функция u.
[randomcl-entropy]

проверить, что величина $H(u,1-u)$ достигает максимума (равного 1) в единственной точке $u = 1/2$.

Возвращаясь к нашей задаче, мы получили, что число последовательностей длины n, у которых частота единиц больше $(1/2+\varepsilon)$, не превосходит $\mathrm{poly}(n)2^{H(1/2+\varepsilon,1/2-\varepsilon)n}$, то есть не превосходит $2^{cn+o(n)}$, где c — некоторая константа, меньшая единицы (зависящая от ε). Тем самым доля таких последовательностей экспоненциально убывает с ростом n. Точно такую же долю составляют последовательности, у которых доля единиц меньше $(1/2-\varepsilon)$.

Подведём промежуточный итог. Для любого фиксированного $\varepsilon > 0$ мы доказали такое утверждение:

Лемма. Доля последовательностей длины n, у которых частота единиц отклоняется от $1/2$ более чем на ε, среди всех последовательностей длины n не превосходит δ_n, где δ_n экспоненциально убывает с ростом n.

Эта лемма (если ограничиться утверждением о сходимости и не накладывать никаких ограничений на скорость сходимости к нулю) называется *законом больших чисел* (без слова «усиленный»). Нам же для продолжения доказательства важно знать, что ряд $\sum_n \delta_n$ сходится.

Мы хотим доказать, что множество $A_{1/2}$ тех последовательностей, у которых предел частоты единиц в начальных отрезках равен $1/2$, имеет меру 1. Другими словами, мы хотим показать, что его дополнение (обозначим его B) является нулевым множеством.

По определению предела B есть объединение по всем $\varepsilon > 0$ множеств B_ε, где B_ε есть множество тех последовательностей, у которых частота единиц бесконечное число раз бывает больше $1/2+\varepsilon$ или меньше $1/2-\varepsilon$.

Очевидно, достаточно рассматривать счётное число различных значений ε (скажем, только рациональные), а счётное объединение нулевых множеств есть нулевое множество. Осталось доказать, таким образом, что множество B_ε является нулевым при любом фиксированном ε.

Множество B_ε состоит из тех последовательностей, у которых есть сколь угодно длинные плохие начальные отрезки (если считать плохим двоичное слово, в котором частота единиц отличается от $1/2$ более чем на ε). Поэтому при любом N множество B_ε покрывается интервалами вида Ω_x, если взять все плохие x длины N и более. Но суммарная (равномерная) мера всех этих интервалов не превосходит

$$\delta_N + \delta_{N+1} + \delta_{N+2} + \ldots,$$

и эта сумма может быть сделана сколь угодно малой, так как ряд $\sum_i \delta_i$ сходится.

(Это рассуждение называется в теории вероятностей *леммой Бореля – Кантелли*. В общем виде эта лемма гласит, что если сумма мер множеств $A_0, A_1, \ldots$ конечна, то множество тех точек, которые принадлежат бесконечно многим множествам A_i, является нулевым.) $\triangleright$

Оценить число плохих слов длины n можно и не используя формулы Стирлинга. Будем рассматривать плохие слова, у которых частоты слишком велики (больше $1/2+\varepsilon$). Рассмотрим на множестве всех слов данной длины n два распределения вероятностей. Первое из них (назовём его L) — равномерное: все слова имеют вероятность 2^{-n}. Второе (назовём его S) отдаёт предпочтение единицам и соответствует n независимым бросаниям монеты, у которой вероятность единицы есть $p = 1/2+\varepsilon$. При этом вероятность слова из u нулей и v единиц равна $q^u p^v$ (где $q = 1/2-\varepsilon$ — вероятность нуля). Чем больше в слове x единиц, тем больше отношение вероятностей $S(x)/L(x)$. Легко подсчитать, что для всех плохих слов это отношение не меньше $2^n/2^{H(p,q)n}$, и потому суммарная L-мера всех плохих слов во столько же раз меньше их суммарной S-меры, которая не превосходит 1. Отсюда получаем, что доля плохих слов не больше $2^{H(p,q)n}/2^n$, то есть мы доказали лемму другим способом. Это доказательство технически проще (хотя до него труднее догадаться). Другое его достоинство в том, что с его помощью можно доказать усиленный закон больших чисел не только для равномерной меры, но и для произвольной бернуллиевой меры (для произвольного p).

47 [non-uniform-slln] Проведите подробно такое доказательство. [Указание. Пусть p_0 и q_0 — фиксированные положительные числа, причём $p_0+q_0=1$. Тогда выражение $-p_0\log p - q_0 \log q$, где p, q — произвольные положительные числа (и $p+q=1$), достигает минимума при $p=p_0$, $q=q_0$. См. также раздел 9.6, с. 241.]

Усиленный закон больших чисел часто формулируют таким образом: «в случайной (по равномерной мере) последовательности частота единиц стремится к $1/2$» (аналогично для других бернуллиевых мер). При этом слова «случайная последовательность» используются не буквально, а как оборот речи: выражение «случайная последовательность обладает свойством α» воспринимается как единое целое и означает, что множество всех последовательностей, не обладающих свойством α, является нулевым.

Возникает естественный вопрос: нельзя ли придать этому обороту буквальный смысл. Именно, выберем какую-либо меру на пространстве Ω, скажем, равномерную. Хотелось бы выделить в пространстве Ω некоторое подмножество, называемое множеством случайных последовательностей, причём так, чтобы для любого свойства α следующие два утверждения были бы эквивалентны:

- все случайные последовательности обладают свойством α;
- множество последовательностей, не обладающих свойством α, имеет меру нуль.

Другими словами, меру 1 должны иметь те и только те множества, которые содержат внутри себя все случайные последовательности. Ещё одна переформулировка: множество случайных последовательностей должно быть минимальным по включению множеством меры 1, а множество неслучайных последовательностей должно быть наибольшим по включению нулевым множеством.

Теперь уже ясно, что этот план невыполним: каждая точка пространства Ω образует одноэлементное множество, являющееся нулевым. Вместе с тем объединение этих одноэлементных множеств покрывает всё пространство.

В 1965 году ученик Колмогорова шведский математик Пер Мартин-Лёф обнаружил, что положение можно спасти, если рассматривать не все нулевые множества, а только «эффективно нулевые». Среди них есть наибольшее по включению, и потому можно определить понятие случайной последовательности так, чтобы свойство α выполнялось для всех случайных последовательностей тогда и только тогда, когда множество последовательностей, не обладающих этим свойством, является эффективно нулевым. В следующем разделе мы изложим конструкцию Мартин-Лёфа.

3.3. Эффективно нулевые множества

[randomml]

Пусть фиксирована некоторая мера на пространстве Ω; меру интервала Ω_x обозначаем $p(x)$.

Говорят, что множество $A \subset \Omega$ является эффективно нулевым (по данной мере), если по любому $\varepsilon > 0$ можно эффективно указать покрытие множества A последовательностью интервалов, суммарная мера которых не превосходит ε.

Это определение требует некоторых уточнений. Во-первых, мы будем рассматривать не все действительные ε, а только рациональные (иначе непонятно, в какой форме алгоритму можно подать на вход число ε). С другой стороны, надо уточнить, в какой форме алгоритм выдаёт последовательность интервалов. Вот одно из возможных уточнений:

Определение.[effective-null-set] Множество $A \subset \Omega$ называется *эффективно нулевым* (по данной мере), если существует вычислимая функция x двух аргументов (первый — положительное рациональное число, второй — натуральное число), значениями которой являются двоичные слова, причём

1) $A \subset \Omega_{x(\varepsilon,0)} \cup \Omega_{x(\varepsilon,1)} \cup \Omega_{x(\varepsilon,2)} \ldots$;

2) $p(x(\varepsilon,0)) + p(x(\varepsilon,1)) + p(x(\varepsilon,2)) + \ldots \leqslant \varepsilon$

при любом рациональном $\varepsilon > 0$. При этом мы не требуем, чтобы функция x была всюду определена; если $x(\varepsilon, i)$ не определено, то соответствующий член (в обоих условиях) пропускается.

$\boxed{\textbf{48}}$ Покажите, что определение не изменится, если предполагать, что алгоритм получает на вход рациональное $\varepsilon > 0$, а на выходе перечисляет некоторое множество двоичных слов (печатая их одно за другим с произвольными перерывами), которые образуют покрытие интервалами с суммарной мерой не больше ε.

$\boxed{49}$ Покажите, что определение не изменится, если мы будем рассматривать не все рациональные ε, а только числа вида 2^{-k} при натуральных k. Покажите, что определение не изменится, если заменить знак $\leqslant$ во втором условии на строгое неравенство.

$\boxed{50}$ Покажите, что определение не изменится, если требовать, чтобы для каждого ε функция $i \mapsto x(\varepsilon, i)$ была бы определена на некотором начальном отрезке натурального ряда (возможно, бесконечном).

$\boxed{51}$ Покажите, что определение не изменится, если заменить перечислимое семейство интервалов на разрешимое. [Указание: интервал можно разбить на несколько более мелких, поэтому можно считать, что длины интервалов в перечислении образуют невозрастающую последовательность, а в этом случае множество интервалов разрешимо.]

Приведём несколько примеров эффективно нулевых множеств относительно равномерной меры на Ω.

Рассмотрим одноэлементное множество, единственным элементом которого является последовательность из одних нулей. Оно является эффективно нулевым: чтобы найти покрытие из интервалов суммарной меры меньше данного $\varepsilon > 0$, возьмём натуральное k, для которого $2^k < \varepsilon$, и покрытие из единственного интервала $\Omega_{00\ldots0}$ (в индексе стоит слово из k нулей).

Формально говоря, $x(\varepsilon, 0) = 0^k$, где 0^k обозначает последовательность из k нулей, а k — наименьшее натуральное число, для которого $2^{-k} < \varepsilon$. Значения $x(\varepsilon, i)$ при $i \neq 0$ не определены.

В этом примере последовательность из одних нулей можно заменить на любую вычислимую последовательность нулей и единиц; нужно только вместо 0^k рассмотреть начальный отрезок этой последовательности длины k.

Но вычислимость существенна, как показывает следующая задача:

$\boxed{52}$ Покажите, что существует последовательность $\omega \in \Omega$, для которой одноэлементное множество $\{\omega\}$ не является эффективно нулевым. [Указание. Рассмотрим все вычислимые функции x, удовлетворяющие второму условию из определения эффективно нулевого множества. Их счётное число. Для каждой из них рассмотрим наибольшее множество A, удовлетворяющее первому условию этого определения (пересечение покрытий по всем ε). Это множество будет (эффективно) нулевым. Объединение счётного числа таких множеств будет нулевым, и потому можно взять ω вне этого объединения.

(Замечание в скобках. Утверждение этой задачи очевидно следует из теоремы Мартин-Лёфа о наибольшем эффективно нулевом множестве (теорема 28, с. 63)), которую мы докажем в этом разделе (и приведённое выше указание к задаче использует по существу тот же метод доказательства). Мы увидим, что множество $\{\omega\}$ является эффективно нулевым тогда и только тогда, когда последовательность ω «не случайна в смысле Мартин-Лёфа».)

Легко построить и невычислимую последовательность ω, для которой множество $\{\omega\}$ является эффективно нулевым. Возьмём любую последовательность $\omega = 0?0?0?0\ldots$ (нули чередуются с произвольными цифрами). Покажем, что множество $\{\omega\}$ является эффективно нулевым.

В самом деле, чтобы построить покрытие с суммарной мерой 2^{-n}, можно взять все конечные слова длины $2n$, в которых n нулей чередуется с n произвольными цифрами (как в ω). Таких слов 2^n, а мера каждого интервала 2^{-2n}, поэтому суммарная мера равна 2^{-n}.

На самом деле в этом примере мы доказали, что множество всех последовательностей, у которых на чётных местах нули, является эффективно нулевым. Тем самым и любое его подмножество (в том числе одноэлементное) является эффективно нулевым.

Возвращаясь к определению эффективно нулевого множества, заметим, что требования в его определении можно разделить в следующем смысле. Будем называть вычислимую функцию x корректной, если она удовлетворяет требованию (2) (про сумму мер). Напомним, что требование (1) определения означает, что множество A при любом рациональном $\varepsilon > 0$ лежит в объединении множеств

$$\Omega_{x(\varepsilon,0)} \cup \Omega_{x(\varepsilon,1)} \cup \Omega_{x(\varepsilon,2)} \ldots$$

Таким образом, корректная функция x «обслуживает» все множества, лежащие внутри

$$\bigcap_{\varepsilon>0} (\Omega_{x(\varepsilon,0)} \cup \Omega_{x(\varepsilon,1)} \cup \Omega_{x(\varepsilon,2)} \ldots) = \bigcap_{\varepsilon>0} \bigcup_{i} \Omega_{x(\varepsilon,i)}.$$

Мы видим, что каждой (вычислимой) корректной функции x соответствует некоторое эффективно нулевое множество (задаваемое только что написанной формулой), и что эффективно нулевыми являются все эти множества (для всех корректных функций), а также все их подмножества — и этим исчерпываются все эффективно нулевые множества.

Прежде чем сформулировать теорему Мартин-Лёфа, дадим определение ***вычислимой меры*** на пространстве Ω.

Действительное число α называют ***вычислимым***, если существует алгоритм, вычисляющий приближения к α с любой заданной точностью $\varepsilon > 0$. Точнее говоря, α вычислимо, если найдётся вычислимая функция $\varepsilon \mapsto a(\varepsilon)$, определённая на всех положительных рациональных числах и принимающая рациональные значения, для которой

$$|\alpha - a(\varepsilon)| < \varepsilon$$

при любом рациональном $\varepsilon > 0$.

53 Покажите, что определение не изменится, если дополнительно потребовать, чтобы выдаваемые приближения были «приближениями с недостатком», то есть чтобы $a(\varepsilon) < \alpha$ для всех ε. [Указание: потеряв вдвое в точности, можно любое приближение превратить в приближение с недостатком (или с избытком).]

54 Покажите, что сумма, разность, произведение и частное двух вычислимых действительных чисел вычислимы.

55 Покажите, что числа e (основание натуральных логарифмов) и π вычислимы.

56 Покажите, что элементарные функции (корень, синус, логарифм, экспонента и т. д.) сохраняют вычислимость, то есть что их значения в вычислимых точках вычислимы. (При этом мы предполагаем, естественно, что основание логарифмов и показательной функции вычислимо.)

Мера μ на пространстве Ω называется вычислимой, если меры всех интервалов являются вычислимыми действительными числами, и, более того, алгоритм приближения к $\mu(\Omega_x)$ можно эффективно указать по x. Более формально:

Определение. Мера μ на пространстве Ω называется *вычислимой*, если существует вычислимая функция $\langle x, \varepsilon\rangle \mapsto a(x, \varepsilon)$, определённая для всех слов x и всех положительных рациональных чисел ε, для которой

$$|\mu(\Omega_x) - a(x, \varepsilon)| < \varepsilon$$

для любых x и ε.

Вообще говоря, это определение не предполагает, что мера всего пространства равна единице, но на практике оно нам понадобится только для таких μ (то есть для распределений вероятностей).

Theorem 28. [randomml-main-theorem] *Пусть μ — вычислимая мера на Ω. Тогда существует наибольшее по включению эффективно нулевое относительно меры μ множество.*

Переформулировка: объединение всех эффективно нулевых (относительно данной вычислимой меры) множеств является эффективно нулевым множеством.

$\triangleleft$ Как мы говорили, каждой корректной функции x соответствует эффективно нулевое множество. Таких множеств счётное число, и любое эффективно нулевое множество содержится в одном из них, поэтому мы немедленно заключаем, что объединение всех эффективно нулевых множеств является нулевым множеством.

Проблема лишь в том, чтобы доказать, что оно является *эффективно* нулевым. Для этого мы будем перечислять все корректные функции, а затем применим эффективный вариант теоремы о счётном объединении нулевых множеств.

По техническим причинам нам будет удобно изменить определение корректной функции. Именно, мы будем называть вычислимую функцию x корректной, если всякая конечная частичная сумма ряда

$$p(x(\varepsilon, 0)) + p(x(\varepsilon, 1)) + p(x(\varepsilon, 2)) + \ldots$$

строго меньше ε. (Напомним, что $p(x) = \mu(\Omega_x)$.) Это требование чуть более сильное (если все частичные суммы ряда строго меньше ε, то его сумма не превосходит ε, но не наоборот). Но легко понять, что определение эффективно нулевого множества не изменится, так как там всегда можно заменить ε, скажем, на $\varepsilon/2$.

В дальнейшем, говоря о корректных функциях (они понадобятся нам только в этом доказательстве), мы имеем в виду это усиленное требование корректности.

Следующая лемма утверждает, что можно эффективно перечислять все корректные функции x.

Лемма. Существует вычислимая (частичная) функция трёх аргументов

$$\langle q, \varepsilon, i\rangle \mapsto X(q, \varepsilon, i)$$

(q и i — натуральные числа, ε — положительное рациональное число), которая при любом фиксированном q даёт корректную функцию X_q двух аргументов и все корректные (вычислимые) функции двух аргументов могут быть получены таким образом.

Доказательство леммы. Расположим все программы для функций двух аргументов (корректных и некорректных) в вычислимую последовательность, и «программой номер q» будем называть q-й член этой последовательности.

Мы определим $X(q, \varepsilon, i)$ как результат применения программы номер q к входам ε, i, если выполны некоторые условия; если нет, то $X(q, \varepsilon, i)$ не определено. Условия гарантируют, что все X_q регулярны и что регулярные функции остались нетронутыми.

Чтобы вычислить $X(q, \varepsilon, i)$, мы параллельно запускаем программу номер q на всех парах

$$(\varepsilon, 0), (\varepsilon, 1), \ldots,$$

сначала делая один шаг первого вычисления, потом два шага первых двух вычислений и т. д. Как только одно из вычислений заканчивается и выдаёт некоторое слово в качестве результата, мы приостанавливаем этот процесс и начинаем проверку корректности. Проверка эта состоит в том, что для всех обнаруженных слов z мы начинаем всё с большей и большей точностью вычислять соответствующие значения $p(z)$, пока не убедимся, что сумма этих значений меньше ε, — точнее говоря, что сумма текущих приближений меньше ε на величину, меньшую суммарной погрешности приближений к $p(z)$. Здесь важно, что мера μ вычислима, так что мы можем вычислять приближения к $p(z)$ с любой заданной точностью для любого слова z.

Возможно, эта проверка никогда и не закончится (такое возможно, если сумма мер уже обнаруженных интервалов не меньше ε), и мы так и не «вернёмся из прерывания».

Теперь $X(q, \varepsilon, i)$ определяется как результат применения программы номер q к входу (ε, i), если этот результат был получен и прошёл проверку корректности в ходе описанного нами процесса.

Если программа номер q на самом деле вычисляет корректную функцию, то все тесты корректности успешно завершатся и X_q будет совпадать с этой функцией. С другой стороны, во всех слу- чаях функция X_q окажется корректной: даже если для некоторого ε программа номер q (применённая к этому ε и ко всем $i = 0, 1, 2, \ldots$) даёт интервалы с слиш- ком большой суммой мер, то все интервалы, начиная с некоторого момента, будут заблокированы, а пропущенные интервалы будут иметь сумму мер меньше ε. Лемма доказана.

57 Объясните, почему нам понадобилось изменить определение корректности: где не проходит доказательство при старом определении? [Указание: если ряд содержит конечное число ненулевых членов, сумма которых в точности равна ε, мы этого никогда не узнаем.]

Продолжаем доказательство теоремы Мартин-Лёфа. Пусть X — функция из леммы. При каждом $q = 0, 1, 2, \ldots$ рассмотрим эффективно нулевое множество Z_q, соответствующее корректной функции X_q. Мы уже знаем, что всякое эффективно нулевое множество содержится в одном из Z_q; остаётся доказать, что объединение $Z_0 \cup Z_1 \cup \ldots$ является эффективно нулевым множеством.

Это делается по той же схеме, что и доказательство теоремы о счётном объединении нулевых множеств. Именно, чтобы найти покрытие суммарной меры не больше ε для $\cup_q Z_q$, мы объединим покрытие размера $(\varepsilon/2)$ для Z_0, покрытие размера $(\varepsilon/4)$ для Z_1, и так далее.

Формально говоря, мы рассматриваем функцию $x(\varepsilon, i)$, определённую так:

$$x(\varepsilon, [q, k]) = X(q, \varepsilon/2^{q+1}, k),$$

где $[q, k]$ обозначает номер пары q, k при каком-то взаимно однозначном кодировании пар натуральных чисел натуральными числами. (Это кодирование должно быть вычислимым, но в остальном может быть произвольным.) $\triangleright$

Теперь мы можем дать определение случайной по Мартин-Лёфу последовательности. Пусть фиксирована некоторая вычислимая мера μ на пространстве Ω.

Определение. Последовательность ω называется *случайной по Мартин-Лёфу* относительно меры μ, если она не содержится в наибольшем эффективно нулевом множестве относительно этой меры.

Переформулировка: последовательность случайна по Мартин-Лёфу, если она не содержится ни в каком эффективно нулевом множестве.

Ещё одна переформулировка: последовательность ω случайна по Мартин-Лёфу, если множество $\{\omega\}$ не является эффективно нулевым.

Отступление о терминологии. Понятие случайной по Мартин-Лёфу последовательности формализует интуитивную идею «типической» (или «типичной») последовательности. Неформально говоря, последовательность типична, если она не обладает никакими особенными свойствами (в целом для последовательностей нехарактерными). Подобным образом понимается слово «типичный» в жаргонном выражении «Онегин — типичный представитель лишних людей» (или, если взять более современный пример, «Вова — типичный лох»). «Особенное» свойство — это свойство, которым обладает лишь пренебрежимо малая часть рассматриваемых объектов. Скажем, свойство последовательности «начинаться с нуля» не является особенным, так как им обладает половина последовательностей. А свойство «каждый второй член равен нулю» является особенным.

Эта неформальная идея получает уточнение в конструкции Мартин-Лёфа: особенным свойством считается принадлежность эффективно нулевому множеству, и таким образом типическими последовательностями становятся последовательности, не принадлежащие никакому эффективно нулевому множеству, то есть случайные по Мартин-Лёфу.

Было бы правильно использовать для таких последовательностей термин «типические», оставив слово «случайные» для интуитивного понятия, допускающего различные уточнения (одним из которых является типичность по Мартин-Лёфу). Однако попытки ввести новую, более правильную, терминологию часто приводят лишь к увеличению путаницы (это замечание авторы самокритично относят и к своим попыткам такого рода). Тем более что путаница и так велика, и в разных текстах слова «случайная последовательность» имеют разный смысл.

В этой книге мы будем говорить «случайная по Мартин-Лёфу» («МЛ-случайная», или «типическая») последовательность, оставив слово «случайная» (без уточнений) для интуитивного представления о случайности.

Следующее утверждение, несмотря на всю его очевидность, полезно осознать (и прочувствовать его парадоксальность):

Theorem 29. `[randomml-null-criterion]` *Множество $A \subset \Omega$ является эффективно нулевым тогда и только тогда, когда все его элементы не случайны по Мартин-Лёфу («нетипичны»).*

В частности, множество всех нетипичных последовательностей является (наибольшим) эффективно нулевым множеством, а множество всех типических последовательностей имеет меру 1.

$\triangleleft$ В самом деле, элементы эффективно нулевого множества нетипичны по определению; с другой стороны, если все элементы множества A нетипичны, то A содержится в наибольшем эффективно нулевом множестве и потому является эффективно нулевым. $\triangleright$

Парадокс здесь в том, что свойство множества «быть нулевым» скорее означает, что у него «мало элементов», чем ограничивает природу этих элементов. Всякая точка на отрезке (или всякая последовательность в Ω) образует нулевое множество, какой бы она ни была.

С другой стороны, получается, что свойство «быть эффективно нулевым» можно сформулировать в терминах ограничений на элементы множества — запрещается включать в него случайные по Мартин-Лёфу последовательности. А неслучайных (по Мартин-Лёфу) элементов можно включать сколько угодно.

В частности, вспомним, что всякая вычислимая последовательность образует эффективно нулевое одноэлементное множество (по равномерной мере). Отсюда сразу же следует такое утверждение:

Theorem 30. [computable-not-random] *Множество всех вычислимых последовательностей нулей и единиц является эффективно нулевым подмножеством Ω относительно равномерной меры.*

Любопытно отметить, что по существу это наблюдение было сделано ещё до Мартин-Лёфа, при изучении конструктивного варианта математического анализа («пример Заславского» [84]). Там речь шла о действительных числах, а не о последовательностях нулей и единиц.

В следующем разделе мы продолжим обсуждение свойств последовательностей, МЛ-случайных по равномерной мере. А сейчас мы приведём любопытный критерий МЛ-случайности последовательности, приведённый со ссылкой на Соловея (R. Solovay) в [7].

Theorem 31. [solovay-criterion] *Последовательность ω не МЛ-случайна по вычислимой мере μ тогда и только тогда, когда существует вычислимая последовательность интервалов с конечной суммой мер, покрывающая эту последовательность бесконечное число раз, то есть вычислимая последовательность слов $x_0, x_1, x_2, \ldots$, для которой*

$$\sum_i \mu(\Omega_{x_i}) < \infty$$

и $\omega \in \Omega_{x_i}$ при бесконечно многих i.

$\triangleleft$ Если последовательность ω не МЛ-случайна, то множество $\{\omega\}$ имеет покрытия с суммой мер не больше ε для любого положительного рационального ε. Соединим такие покрытия для $\varepsilon = 1, 1/2, 1/4, 1/8, \ldots$ Соединённое покрытие имеет сумму мер не больше 2 и покрывает ω бесконечно много раз (на каждом уровне хотя бы по разу).

Напротив, пусть есть некоторое покрытие точки ω интервалами, соответствующими словам $x_0, x_1, x_2, \ldots$, и сумма мер этих интервалов не превосходит какой-то границы c, которую удобно считать рациональным числом. Чтобы найти покрытие точки ω с суммой не больше ε, рассмотрим множество тех точек из Ω, которые покрыты не менее N раз, где целое положительное N выбрано так, чтобы c/N было меньше ε. Легко понять, что это множество представимо в виде объединения непересекающихся интервалов, и эти интервалы можно перечислять алгоритмически, зная N и глядя на последовательность $x_0, x_1, x_2, \ldots$. Тем самым множество $\{\omega\}$ является эффективно нулевым, а последовательность ω — неслучайной по Мартин-Лёфу. $\triangleright$

Замечание. Эта теорема является «конструктивизацией» леммы Бореля – Кантелли (если сумма мер множеств $A_0, A_1, \ldots$ конечна, то множество точек, принадлежащих бесконечно многим из A_i, является нулевым), и приведённое нами рассуждение является конструктивизацией доказательства леммы Бореля – Кантелли. Однако тут нужна осторожность: для конструктивизации годится не всякое доказательство. Обычное доказательство этой леммы (ряд сходится, значит, хвосты его стремятся к нулю и покрывают интересующее нас множество) не годится, так как мы не умеем по ε эффективно находить хвост, меньший ε.

3.4. Свойства случайных по Мартин-Лёфу последовательностей

[randomun]

Усиленный закон больших чисел также даёт пример эффективно нулевого множества (относительно равномерной меры).

Theorem 32. [randomml-lln] *Множество последовательностей нулей и единиц, для которых $1/2$ не является пределом последовательности частот, является эффективно нулевым по равномерной мере.*

$\lhd$ Достаточно показать, что при любом рациональном $\varepsilon > 0$ множество тех последовательностей, у которых частота бесконечное число раз становится больше $1/2 + \varepsilon$ (или меньше $1/2 - \varepsilon$), является эффективно нулевым. Для этого заметим, что оценка меры, приведённая при доказательстве закона больших чисел в предыдущем разделе (теорема 27, с. 57), была эффективной: покрытие состояло из продолжений всех достаточно длинных слов с большим отклонением частоты, а его мера эффективно оценивалась сверху остаточным членом в сумме бесконечно убывающей геометрической прогрессии. $\rhd$

Переформулируем доказанное утверждение как свойство индивидуальных МЛ-случайных последовательностей:

Theorem 33. [random-lln-effective] *Пусть $\omega = \omega_0\omega_1\ldots$ — случайная в смысле Мартин-Лёфа последовательность относительно равномерной меры. Тогда*

$$\lim_{n\to\infty} \frac{\omega_0 + \omega_1 + \ldots + \omega_{n-1}}{n} = \frac{1}{2}.$$

Аналогичное утверждение справедливо и для неравномерных бернуллиевых мер. Пусть даны положительные числа p, q, для которых $p + q = 1$, причём p и q вычислимы. Рассмотрим бернуллиеву меру с параметрами q, p (соответствующую последовательности независимых испытаний с вероятностью успеха p, см. с. 56). Легко проверить, что эта мера вычислима (поскольку вычислимы p и q).

Theorem 34. *Любая МЛ-случайная относительно бернуллиевой меры с вычислимыми параметрами q, p последовательность имеет предел частот, равный p.*

$\lhd$ В самом деле, оценка вероятностей последовательностей с большими уклонениями частот от p (производимая с помощью сравнения меры с другой мерой, в которой p сдвинуто, см. задачу 47, с. 59), даёт явную оценку и покрытие, поэтому получается эффективно нулевое множество. $\rhd$

Ещё несколько свойств типичности (МЛ-случайности) по равномерной мере:

Theorem 35. *Пусть ω — типическая по равномерной мере последовательность. Тогда последовательность полученная из ω удалением, добавлением или изменением конечного числа членов, также является типической.*

$\triangleleft$ Достаточно доказать, что дописывание в начало последовательности нуля или единицы, а также удаление первого её члена не сказывается на типичности.

В самом деле, пусть последовательность ω нетипична, то есть образует эффективно нулевое множество: по всякому ε можно построить покрытие интервалами с суммой мер не больше ε. Припишем к словам, задающим эти интервалы, нуль в начале. Получится покрытие для последовательности 0ω вдвое меньшей меры. Это рассуждение показывает, что если ω не типична, то и 0ω не типична. (Аналогично для 1ω.)

С другой стороны, если у всех слов, задающих интервалы покрытия, удалить первый бит, то получится семейство интервалов вдвое большей меры, покрывающее последовательность ω', получающуюся удалением из ω первого бита. Значит, ω' не типична. $\triangleright$

58 Покажите, что замена всех нулей на единицы и наоборот сохраняет типичность последовательности (по равномерной мере).

Типичность также сохраняется при переходе к вычислимой подпоследовательности, как показывает следующая задача.

59 Пусть $n_0, n_1, n_2, \ldots$ — вычислимая последовательность различных натуральных чисел ($n_i \neq n_j$ при $i \neq j$). Покажите, что если последовательность $\omega = \omega_0\omega_1\omega_2\ldots$ типична, то и её подпоследовательность

$$\omega|n = \omega_{n_0}\omega_{n_1}\omega_{n_2}\ldots$$

типична. [Указание. Из каждого интервала Ω_x в покрытии для $\omega|n$ получается конечное семейство интервалов, у которых члены с номерами $n_0, n_1, \ldots, n_{i-1}$ совпадают с битами слова x (здесь i — длина слова x), а остальные биты любые. Общая мера этих интервалов равна мере интервала Ω_x.]

Мы продолжим эту тему (о выборе подпоследовательности из случайной последовательности) в главе о частотном подходе к определению случайности, предложенном фон Мизесом (глава 9, с. 228).

60 Пусть последовательность ω типична (МЛ-случайна) по равномерной мере. Разделим её на блоки по две цифры и заменим блоки 00 на нули, а блоки 01, 10 и 11 — на единицы. Докажите, что полученная последовательность типична по бернуллиевой мере с параметрами $1/4, 3/4$. [Указание. Описанное преобразование задаёт отображение Ω в себя. Прообраз любого открытого множества U при этом открыт, и равномерная мера этого прообраза равна $(1/4, 3/4)$-мере множества U.]

61 (Продолжение.) Докажите, что любая типическая по $(1/4, 3/4)$-мере последовательность может быть получена описанным способом из последовательности, типической по равномерной мере. [Указание. Для любого открытого множества $B \subset \Omega$ рассмотрим множество B' всех тех последовательностей ω, у которых $F^{-1}(\{\omega\}) \subset B$ (те последовательности, которые не имеют прообраза вне B, или дополнение к образу дополнения B). Образ компактного множества компактен, поэтому B' открыто. Проверьте, что если B есть объединение перечислимого семейства интервалов, то и B' имеет такой вид, и бернуллиева мера B' не превосходит равномерной меры B. См. также доказательство более общего утверждения (теорема 99, с. 154).

Интересно понять, какова может быть сложность МЛ-случайной (по равномерной мере) последовательности с точки зрения теории вычислимых функций. Мы уже видели, что такая последовательность не может быть вычислимой. Она также не может быть характеристической функцией перечислимого множества.

Theorem 36. *Пусть A — перечислимое множество натуральных чисел. Тогда его характеристическая последовательность $a_0a_1a_2\ldots$, где $a_i = 0$ при $i \notin A$ и $a_i = 1$ при $i \in A$, не является МЛ-случайной по равномерной мере.*

$\lhd$ Пусть k — произвольное натуральное число. Будем перечислять множество A, следя за первыми k членами его характеристической последовательности. По мере перечисления среди этих членов появляются всё новые и новые единицы и мы получаем всё новые и новые варианты k-битового начала характеристической последовательности (какой-то из них окажется окончательным, но у нас нет способа узнать об этом). Но так или иначе этих вариантов не больше $k+1$ штук (число единиц может меняться от 0 до k). И если мы все эти слова включим в покрытие, то общая мера интервалов покрытия будет не больше $(k+1)/2^k$ и таким образом может быть сделана сколь угодно малой. (Заметим, что в определении эффективно нулевого множества как раз требуется перечислять элементы покрытия, не указывая момента окончания такого перечисления.) $\rhd$

Возникает вопрос, можно ли вообще в каком-то смысле указать конкретную случайную по Мартин-Лёфу последовательность. Следующий результат (приводимый для читателей, знакомых с началами теории вычислимых функций, о которых можно прочесть, например, в [77]), показывает, что случайную последовательность можно найти в классе $\Sigma_2 \cap \Pi_2$ арифметической иерархии (другое описание последовательностей этого класса — вычислимые относительно $\mathbf{0}'$ последовательности).

Theorem 37. *Существует вычислимая относительно $\mathbf{0}'$ последовательность, МЛ-случайная по равномерной мере.*

$\lhd$ Достаточно показать, что для любого перечислимого множества слов $x_0, x_1, \ldots$, у которого $\sum 2^{-l(x_i)} < 1/2$, существует вычислимая относительно $\mathbf{0}'$ последовательность, для которой ни одно из слов x_i не является началом. (Максимальное эффективно нулевое множество имеет покрытие с суммой мер меньше $1/2$, и всякая последовательность вне этого покрытия будет МЛ-случайной.)

Посмотрим, как интервалы Ω_{x_i} распределяются между двумя половинами множества Ω (какие из них начинаются на нуль, а какие на единицу).

В сумме мера тех и других не превосходит $1/2$, значит, мера интервалов, приходящихся на одну из половин, не больше $1/4$. Правда, наблюдая за последовательностью x_i, мы не можем с уверенностью выбрать эту половину (вдруг потом объявится короткое слово, которое попадёт в эту половину).

Но если у нас есть оракул для $\mathbf{0}'$, то с его помощью такой выбор можно сделать (поскольку превышение меры над $1/4$ есть перечислимое событие). Соответствующую половину разделим на две четверти и посмотрим, в какой из них мера интервалов покрытия будет не больше $1/8$. И так далее. Тем самым мы получим вычислимую (относительно $\mathbf{0}'$) последовательность нулей и единиц с таким свойством: над любым её началом интервалы покрытия заполняют не более половины всех продолжений. В частности, никакое начало этой последовательности не может войти в покрытие, что нам и требовалось. $\triangleright$

(Другая конструкция последовательности с таким свойством приведена в разделе 5.7 на с. 146.)

По существу доказательство этой теоремы представляет собой релятивизованный вариант такой задачи:

$\boxed{62}$ `[schnorr-nonuniversal]` Пусть имеется вычислимая последовательность слов $x_0, x_1, x_2, \ldots$, причём сумма ряда

$$\sum_i 2^{-l(x_i)}$$

меньше 1 и является вычислимым действительным числом. Тогда существует вычислимая последовательность нулей и единиц, для которой ни одно из слов x_i не является началом.

[Указание. Пусть сумма этого ряда меньше рационального числа S, меньшего единицы. Тогда можно по индукции построить вычислимую последовательность $\omega_0\omega_1\omega_2\ldots$ с таким свойством: доля множества $U = \cup\Omega_{x_i}$ среди продолжений слова $\omega_0 \ldots \omega_k$ составляет менее S.]

Последняя задача связана с модификацией определения случайности по Мартин-Лёфу, предложенной Шнорром [66]. А именно, в определении эффективно нулевого множества можно дополнительно потребовать, чтобы для каждого $\varepsilon > 0$ ряд из мер покрывающих интервалов (сумма которого не должна превосходить ε) вычислимо сходился. Это означает, что для каждого $\varepsilon > 0$ и $\delta > 0$ можно алгоритмически указать конечное число членов ряда $\sum_i p(x(\varepsilon, i))$, которые все определены и приближают сумму всего ряда с ошибкой не более δ. (Поскольку члены ряда неотрицательны, это эквивалентно тому, что его сумма является вычислимым действительным числом, которое можно указать по ε.) Эффективно нулевые множества, для которых выполнено это дополнительное условие, будем называть *эффективно нулевыми по Шнорру*. (Шнорр называет их *total rekursive Nullmenge*, см. определение 8.1 в [66], в отличие от эффективно нулевых в смысле Мартин-Лёфа, которые называются у Шнорра *rekursive Nullmenge*, см. там же определение 4.1.)

$\boxed{63}$ Покажите, что если в определении эффективно нулевого множества требовать, чтобы суммарная мера покрывающих интервалов *в точности равнялась* ε, то получится определение, эквивалентное определению Шнорра. (Можно также вместо суммарной меры покрывающих интервалов говорить о мере их объединения.)

Задача 62 показывает, что для любого эффективно нулевого по Шнорру множества существует вычислимая последовательность, ему не принадлежащая. (Для простоты мы ограничиваемся случаем равномерной меры, хотя это и не существенно.) Поскольку вычислимая последовательность образует одноэлементное эффективно нулевое по Шнорру множество, отсюда следует, что среди эффективно нулевых по Шнорру множеств нет наибольшего (другими словами, объединение нулевых по Шнорру множеств не является нулевым по Шнорру). Тем не менее можно назвать последовательность, не входящую ни в одно нулевое по Шнорру множество, *случайной по Шнорру* (или *типической по Шнорру*).

Полученный класс последовательностей оказывается более широким: как показывает следующая задача (и результаты главы 5), существуют случайные по Шнорру последовательности, не являющиеся случайными по Мартин-Лёфу.

$\boxed{\textbf{64}}$ Покажите, что существует случайная по Шнорру последовательность $\omega = \omega_0\omega_1\omega_2\ldots$, у которой сложность начальных отрезков растёт логарифмически, то есть $\mathit{KS}(\omega_0\ldots\omega_{n-1}) = O(\log n)$.

[Указание. В задаче 62 мы видели, как строить вычислимую последовательность, не принадлежащую заданному эффективно нулевому по Шнорру множеству. Если в какой-то момент мы захотим ввести в этом построение другое эффективно нулевое по Шнорру множество, это вполне возможно, надо только выбрать достаточно малое покрытие (исходя из имеющегося на данный момент запаса). Так можно вводить эффективно нулевые множества одно за другим. Это не даст вычислимой случайной по Шнорру последовательности (которой быть не может), поскольку нам нужна дополнительная информация о том, какие алгоритмы задают эффективно нулевые по Шнорру множества, а какие нет. Но если вводить новые алгоритмы постепенно, когда построенный кусок последовательности уже длинен, то эта дополнительная информация будет логарифмической по сравнению с длиной последовательности.]

Мы вернёмся к определению случайности по Шнорру в разделе 9.8, где будет дана его переформулировка в терминах вычислимых мартингалов.

$\boxed{\textbf{65}}$ `[schnorr-solovay]` Покажите, что последовательность ω не является случайной по Шнорру тогда и только тогда, когда существует вычислимая последовательность слов $x_0, x_1, \ldots$, для которой ряд $\sum_i p(x_i)$ вычислимо сходится и среди x_i имеется бесконечное много начал последовательности ω (аналог критерия Соловея для случайности по Шнорру, теоремы 31). [Указание. В данном случае работает и стандартное доказательство леммы Бореля – Кантелли.]

4. Априорная вероятность и префиксная сложность

[prefix]

4.1. Вероятностные машины и полумеры на $\mathbb{N}$

[prefix-pp] Рассмотрим алгоритм (машину, программу) с датчиком случайных битов. Такой алгоритм содержит операции

$$b := \textit{random};$$

при которых переменной b присваивается нуль или единица с равными вероятностями. (Встретив такую команду, мы прерываем выполнение алгоритма, бросаем честную монету и результат бросания записываем в переменную b.) Такие алгоритмы называют *вероятностными*.

Результат работы вероятностного алгоритма зависит не только от его входа, но и от результатов бросаний монеты (случайных битов). Таким образом, при данном входе выход (результат работы) алгоритма является случайной величиной.

Более формально вероятность того или иного выхода вероятностного алгоритма A определяется так. На пространстве Ω всех бесконечных последовательностей нулей и единиц рассматривается равномерная бернуллиева мера. При этом мера множества Ω_u всех бесконечных продолжений данного конечного слова u равна $2^{-l(u)}$.

Для данного входа x и последовательности $\omega \in \Omega$ через $A(x, \omega)$ обозначим результат работы A на входе x, если в качестве случайных битов берутся биты из последовательности ω (каждый вызов *random* берёт новый бит). Значение $A(x, \omega)$ может быть не определено при некоторых x и ω. Для каждого возможного выхода y алгоритма рассмотрим множество $\{\omega \mid A(x, \omega) = y\}$. Это множество измеримо (и даже открыто в смысле естественной топологии в Ω); оно является объединением интервалов Ω_z для всех слов z, которые являются результатами бросаний монеты к моменту появления выхода y. Мера этого множества и называется *вероятностью выхода y при входе x*.

В этом разделе мы будем рассматривать вероятностные машины без входа, выходами которых являются натуральные числа. Пример: машина бросает монету, пока не появится единица, и выдаёт на выход число нулей перед первой единицей. Для этого алгоритма распределение вероятностей на возможных выходах такое: вероятность p_i появления числа i равна $2^{-(i+1)}$. В самом деле, p_0 есть вероятность того, что первое бросание монеты дало единицу, p_1 — вероятность того, что первое бросание дало нуль, а второе — единицу и так далее.

В данном конкретном случае сумма ряда $\sum p_i$ равна единице: вероятность того, что машина так ничего и не напечатает (так случится, если всё время будут выпадать нули) равна нулю. Но для других алгоритмов эта сумма может быть и меньше единицы.

Итак, каждой вероятностной машине без входа, выдающей натуральные числа на выходе, соответствует последовательность $p_0, p_1, \ldots$, где p_i есть вероятность появления числа i. Какие последовательности действительных чисел $p_0, p_1, \ldots$ могут получиться таким образом? Одно условие очевидно: $\sum p_i \leqslant 1$. Но, конечно, это условие не является достаточным: вероятностных машин счётное число, а возможных последовательностей $p_0, p_1, \ldots$ — континуум.

Начнём с более простого вопроса: какова может быть вероятность остановки вероятностной машины (без входа)? Чтобы сформулировать ответ, введём понятие перечислимого снизу действительного числа.

Говорят, что действительное число α *перечислимо снизу*, если оно есть предел вычислимой неубывающей последовательности рациональных чисел.

$\boxed{66}$ Докажите, что если действительное число α вычислимо (существует алгоритм, который по любому рациональному $\varepsilon > 0$ указывает приближение к α с погрешностью не более ε), то α перечислимо снизу. [Указание. Последовательность приближений снизу можно переделать в возрастающую.]

$\boxed{67}$ Докажите, что действительное число α вычислимо тогда и только тогда, когда оба числа α и $-\alpha$ перечислимы снизу.

Вот эквивалентное определение перечислимости снизу: число α перечислимо снизу, если множество всех рациональных чисел, меньших α, перечислимо.

В самом деле, если α есть предел неубывающей вычислимой последовательности $a_0 \leqslant a_1 \leqslant a_2 \leqslant \ldots$, то перечисляя все числа, меньшие какого-либо из a_i, мы перечислим все числа, меньшие α. Напротив, умея перечислять все рациональные числа, меньшие α, мы можем составить из них последовательность. Затем эту пследовательность надо сделать неубывающей, выбрасывая из неё члены, меньшие уже встречавшихся.

Понятие перечислимого снизу действительного числа даёт ответ на поставленный нами вопрос:

Theorem 38. [enumerable-machines] **(а)** *Пусть M — произвольная вероятностная машина (без входа). Тогда вероятность её остановки есть перечислимое снизу действительное число.*

(б) *Всякое перечислимое снизу действительное число есть вероятность остановки некоторой вероятностной машины.*

$\triangleleft$ (а) Вероятность p_n остановки машины в течение n (или менее) шагов есть некоторое рациональное число: за n шагов можно сделать не более n бросаний монеты, поэтому каждый исход имеет вероятность, кратную $1/2^n$, и вероятность остановки также кратна $1/2^n$.

Число p_n можно найти, моделируя работу машины при всех вариантах бросаний. С ростом n оно возрастает (точнее, не убывает) и стремится к вероятности остановки машины (без ограничения числа шагов).

(б) Пусть q — произвольное перечислимое снизу число. Это означает, что $q = \lim q_n$ для некоторой вычислимой последовательности

$$q_0 \leqslant q_1 \leqslant q_2 \leqslant \ldots$$

рациональных чисел. Построим машину, вероятность остановки которой равна q. Машина бросает монету и воспринимает полученные биты $b_0, b_1, b_2, \ldots$ как последовательные знаки двоичного числа $\beta = 0{,}b_0b_1b_2\ldots$. Параллельно она вычисляет рациональные числа $q_0, q_1, q_2, \ldots$ и останавливается, как только имеющейся у неё информации достаточно, чтобы утверждать, что $\beta < q$. Другими словами, она останавливается, как только

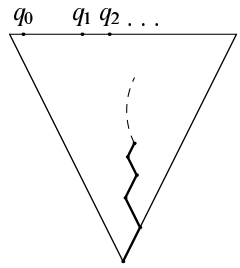

Рис. 9. Сравнение $\beta = 0{,}b_0b_1\dots$ и $q = \lim q_i$.

[prefixpp.1]

$0{,}b_0b_1\dots b_i111\dots$ (текущая верхняя оценка для числа β) оказывается меньше q_i (текущей нижней оценки для числа q). Символически этот процесс изображён на рис. 9.

Построенная машина останавливается тогда и только тогда, когда $\beta < q$. Убедимся в этом. Если число β меньше q, то машина остановится. В самом деле, числа q_i стремятся к q, а верхние оценки для числа β стремятся к β, и потому в какой-то момент q_i превысит текущую верхнюю оценку. С другой стороны, в случае остановки $\beta < q$ по построению.

Таким образом, вероятность остановки есть вероятность события $\beta < q$, то есть длина промежутка $[0, q)$, то есть q (число β, составленное из случайных битов, равномерно распределено на отрезке $[0, 1]$). $\triangleright$

Вернёмся к вопросу о том, каким может быть распределение вероятностей на выходах вероятностной машины. Нам понадобится следующее определение. Говорят, что последовательность действительных чисел $p_0, p_1, p_2, \dots$ *перечислима снизу*, если существует вычислимая всюду определённая функция p двух натуральных аргументов с рациональными значениями (разрешается также дополнительное значение $-\infty$), для которой

$$p(i, 0) \leqslant p(i, 1) \leqslant p(i, 2) \leqslant \dots$$

и

$$p_i = \lim_{n \to \infty} p(i, n)$$

для любого i.

Можно сказать, что в этом определении мы требуем перечислимости снизу чисел p_i «равномерно по i». Можно было бы определить перечислимость снизу и по-другому, как показывает следующая теорема:

Theorem 39. *Последовательность $p_0, p_1, p_2 \dots$ перечислима снизу тогда и только тогда, когда множество пар $\langle r, i\rangle$, где i — натуральное число, а r — рациональное число, меньшее p_i, перечислимо.*

$\triangleleft$ Перечислимость множества означает, что существует алгоритм, перечисляющий его элементы (в произвольном порядке, с произвольными промежутками; алгоритм может никогда не завершать работу, даже если множество конечно).

Предположим, что последовательность $p_0, p_1, p_2, \dots$ перечислима снизу и $p_i = \lim_n p(i, n)$. Алгоритм перебирает все пары $\langle r, i\rangle$, возвращаясь к каждой паре бесконечное число раз. При n-м обращении к паре $\langle r, i\rangle$ мы сравниваем r с $p(i,n)$; если оказывается, что $r < p(i,n)$, то пара $\langle r, i\rangle$ включается в перечисление. Ясно, что таким образом мы перечислим все нужные пары и только их ($r < \lim_n p(i,n)$ тогда и только тогда, когда $r < p(i,n)$ для некоторого n).

Напротив, пусть свойство $r < p_i$ перечислимо. Возьмём алгоритм, перечисляющий такие пары. Чтобы вычислить $p(i,n)$, сделаем n шагов перечисления; отберём все пары $\langle r, i\rangle$ с данным i и возьмём в них максимальное r. Положим $p(i,n)$ равным этому r (если пар с нужным значением i не окажется, то $p(i,n) = -\infty$). Легко понять, что с ростом n пар становится больше и значение $p(i,n)$ может только возрасти, а предел $\lim_n p(i,n)$ равен p_i (поскольку в перечислении появляются все рациональные числа, меньшие p_i). $\triangleright$

Теперь можно дать обещанную характеризацию распределений вероятностей, соответствующих вероятностным машинам.

Theorem 40. [semimeasure-machine] *(а) Пусть M — произвольная вероятностная машина без входа, возможными выходами которой являются натуральные числа. Пусть p_i — вероятность появления числа i на выходе. Тогда последовательность p_i перечислима снизу и $\sum_i p_i \leqslant 1$.*

(б) Пусть $p_0, p_1, \ldots$ — перечислимая снизу последовательность неотрицательных действительных чисел, причём $\sum_i p_i \leqslant 1$. Тогда существует вероятностная машина M, для которой вероятность появления числа i на выходе равна p_i.

$\triangleleft$ Первая часть доказательства (что вероятности перечислимы снизу) проходит точно так же, как и раньше, только вместо единственной вероятности остановки мы оцениваем снизу вероятность остановки с ответом i (для каждого i).

Вторая часть на самом деле тоже мало меняется. Раньше у нас выделялась всё бóльшая часть пространства под область остановки. Теперь область остановки будет поделена на счётное число подобластей; для каждого i есть область остановки с ответом i. Возрастающие приближения снизу к p_i составляют требования к размеру соответствующей области. Наша задача — распределять пространство между всеми требованиями. Это можно делать самым простым способом, выделяя каждому новому требованию кусок отрезка (слева направо) в соответствии с увеличением его заказа. Таким образом отрезок (не обязательно весь!) распределяется между различными возможными выходами. Заметим, что область остановки с выходом i не является связной (она состоит из отдельных промежутков, число которых увеличивается по ходу процесса).

Параллельно с распределением отрезка по областям мы порождаем биты случайного числа β, и как только становится ясно, в чью область попадёт β, соответствующее число выдаётся на выход.

Более формально можно сказать это следующим образом. Пусть $p_i = \lim_n p(i,n)$ в соответствии с определением перечислимости снизу. Без ограничения общности можно считать, что все $p(i,n)$ неотрицательны (заменим отрицательные на нуль), и что для каждого n лишь конечное число значений $p(i,n)$ отличны от нуля (положим $p(i,n) = 0$ при $i > n$). Будем откладывать от нуля слева направо отрезки, помечая каждый натуральным числом. При этом мы хотим, чтобы к n-му шагу суммарная длина всех отрезков, помеченных числом i, равнялась $p(i,n)$. Поэтому при увеличении n на единицу мы смотрим, насколько увеличились значения $p(i,n)$ при разных i, и добавляем отрезки недостающего размера с нужными пометками.

За пределы отрезка $[0,1]$ мы не выйдем, поскольку $p(i,n) \leqslant p_i$ и $\sum p_i \leqslant 1$.

Вероятностная машина действует следующим образом: мы останавливаемся с выходом i, если полученные к данному моменту случайные биты $b_0 b_1 \ldots b_k$ таковы, что отрезок на действительной прямой, состоящий из чисел, двоичные записи которых начинаются на $b_0 b_1 \ldots b_k$, содержится во внутренности одного из отрезков, помеченного числом i. (Внутренность отрезка $[u,v]$ есть интервал (u,v).) Легко проверить, что выход i появится тогда и только тогда, когда β принадлежит внутренности одного из отрезков, помеченных числом i, а общая мера этого множества (объединения соответствующих интервалов) равна как раз p_i. $\triangleright$

Будем называть *перечислимой снизу полумерой* на $\mathbb{N}$ любую последовательность чисел p_i, обладающую указанными в теореме свойствами. (Иногда мы будем также писать $p(i)$ вместо p_i.) Таким образом, у нас есть два определения перечислимых снизу полумер: (1) распределения вероятностей на выходах вероятностных машин; (2) перечислимые снизу ряды с неотрицательными членами и суммой не больше 1. Только что доказанная теорема утверждает, что эти два определения эквивалентны.

Слово «полумера» выглядит довольно странно (особенно учитывая обиходное выражение «ограничиться полумерами»), но другого употребительного термина нет. Если не требовать перечислимости, можно назвать *полумерой* на $\mathbb{N}$ любую функцию $i \mapsto p_i$, для которой $\sum_i p_i \leqslant 1$. Такие функции можно рассматривать как распределения вероятностей на множестве $\mathbb{N} \cup \{\perp\}$, где $\perp$ — специальный символ «неопределённости». При этом вероятность числа i есть p_i, а вероятность $\perp$ есть $1 - \sum_i p_i$. Но мы будем рассматривать лишь перечислимые снизу полумеры (если обратное не оговорено явно).

Мы определяли (перечислимые снизу) полумеры на множестве натуральных чисел. Но эти определения без всяких изменений переносятся и на двоичные слова (или любые другие конструктивные объекты). Например, если мы рассмотрим вероятностную машину, выходами которой являются двоичные слова, то она задаёт некоторую перечислимую снизу полумеру на множестве двоичных слов.

Важное замечание: в главе 5 мы будем рассматривать другое понятие полумеры на пространстве конечных и бесконечных двоичных последовательностей. Это понятие будет соответствовать машинам, которые порождают выход бит за битом и не обязаны останавливаться. Но пока что появление на выходе слова x означает, что машина напечатала все биты слова x и остановилась, так что двоичные слова ничем не отличаются в этом смысле от натуральных чисел.

4.2. Наибольшая полумера

[prefix-m] Будем сравнивать полумеры на $\mathbb{N}$ с точностью до мультипликативной константы. Назовём перечислимую снизу полумеру m *максимальной*, или *наибольшей*, если для любой другой перечислимой снизу полумеры m' выполнено равенство $m'(i) \leqslant cm(i)$ для некоторого c и для всех i. (Термин «наибольшая» более точен, поскольку речь идёт именно о наибольших элементах в некотором частично упорядоченном множестве, но чаще говорят о максимальной перечислимой полумере.)

Theorem 41. [max-semi-N] *Существует наибольшая перечислимая снизу полумера на* $\mathbb{N}$.

$\triangleleft$ Мы должны построить вероятностную машину M с таким свойством: для любого другой машины M' вероятность появления произвольного числа i на выходе машины M меньше такой же вероятности для M' не более чем в константу раз.

Этого можно добиться так: пусть машина M сначала случайно выберет вероятностную машину (вероятности выбрать ту или иную машину могут быть любыми, важно только, чтобы всякая вероятностная машина появлялась с положительной вероятностью), а потом моделирует выбранную машину. Если вероятность того, что машина M' будет выбрана, равна p, то вероятность $m(i)$ появления i на выходе машины M не меньше $p \cdot m'(i)$, где $m'(i)$ — вероятность появления i на выходе машины M'. Поэтому можно положить $c = 1/p$.

Случайный выбор можно реализовать, например, так. Перенумеруем все машины каким-либо естественным образом; получится последовательность $M_0, M_1, M_2, \ldots$, включающая все вероятностные машины. Машина M бросает монету, ожидая появления первой единицы. Далее она моделирует работу машины M_i, где i — число нулей перед первой единицей. $\triangleright$

Поучительно провести параллельно доказательство этой теоремы на языке перечислимых снизу рядов. Мы должны доказать, грубо говоря, что существует «самый плохо сходящийся» перечислимый снизу ряд, члены которого мажорируют (с точностью до умножения на константу) любой другой сходящийся перечислимый снизу ряд. (Точнее говоря, нас интересуют ряды, не просто сходящиеся, а ряды с суммой не больше единицы, но это не имеет значения, так как всё равно мы разрешаем умножение на константу.)

Идея доказательства проста: возьмём все перечислимые снизу ряды с неотрицательными членами и суммой не больше 1, и сложим их с коэффициентами, которые образуют сходящийся ряд. Суммарный ряд, очевидно, будет наибольшим (с точностью до умножения на константу). Вопрос только в том, как сделать, чтобы получился перечислимый снизу ряд.

Произвольная перечислимая снизу полумера задаётся вычислимой функцией $p\colon \langle i, n\rangle \mapsto p(i, n)$. Таких функций счётное число (поскольку алгоритмов счётное число). Расположим их в последовательность $p^{(0)}, p^{(1)}, p^{(2)}, \ldots$ и рассмотрим функцию

$$p(i, n) = \sum_{k=0}^{n} \lambda_k p^{(k)}(i, n),$$

где λ_k — вычислимая последовательность рациональных чисел с $\sum_k \lambda_k \leqslant 1$ (например, $\lambda_k = 2^{-k-1}$). Определённая таким образом функция p возрастает с ростом n (при фиксированном i), поскольку сумма включает всё больше членов и сами члены растут. При этом

$$\lim_{n\to\infty} p(i, n) = \sum_k \lambda_k \lim_{n\to\infty} p^{(k)}(i, n)$$

для любого i, то есть построенная полумера есть действительно сумма всех полумер с коэффициентами λ_k.

Это рассуждение, однако, содержит пробел. Проблема в том, что функция $p(i, n)$ должна быть вычислимой. Поэтому недостаточно просто расположить функции, задающие полумеры, в последовательность, нужно ещё, чтобы эта последовательность была вычислимой (как функция трёх аргументов). Мы не можем просто написать подряд все программы и сказать, что $p^{(k)}$ — это функция, вычисляемая k-й программой (для функций двух натуральных аргументов с рациональными значениями). Может случиться, что k-я программа не задаёт перечислимую снизу полумеру (вычисляет не всюду определённую функцию, или эта функция не монотонна по второму аргументу, или сумма пределов больше единицы).

Положение спасает такая

Лемма. Всякую программу P для функции двух натуральных аргументов с рациональными значениями (и, возможно, дополнительным значением $-\infty$) можно алгоритмически преобразовать в программу P', которая вычисляет всюду определённую функцию того же типа, задающую некоторую перечислимую снизу полумеру. При этом, если сама программа P задавала полумеру, то новая полумера (соответствующая P') равна старой.

Доказательство леммы. Пусть P — данная нам программа (не обязательно всюду определённая). Для начала положим $P'(i,n)$ равным максимальному из чисел, которые получаются за n шагов при вычислении значений $P(i,0),\ldots,P(i,n)$ (если за n шагов ни одно из вычислений не даёт результата или все результаты отрицательны, то $P'(i,n)=0$). Это уже гарантирует всюду-определённость P', неотрицательность и монотонность по второму аргументу. При этом, если (для данного i) значения $P(i,n)$ определены при всех n, неотрицательны и не убывают по n, то $\lim_n P'(i,n)=\lim_n P(i,n)$.

Единственное, чего нам ещё недостаёт: нужно, чтобы $\sum p'_i \leqslant 1$, где $p'_i=\lim_n P'(i,n)$. Для начала заметим, что можно полагать $P'(i,n)$ равным нулю при всех $n<i$. (Ясно, что такое преобразование не нарушит монотонности и не изменит предела $\lim_n P'(i,n)$.) Сумма $P'(i,n)$ по всем i при фиксированном n теперь будет конечной, и её можно вычислить. Нам нужно, чтобы эта сумма не превосходила единицы. Чтобы этого добиться, принудительно скорректируем P': перестанем увеличивать P' с ростом n, как только такое увеличение сделает сумму слишком большой. (Сперва мы корректируем все значения при $n=0$, потом при $n=1$ и так далее.) Лемма доказана.

Используя лемму, мы можем расположить все перечислимые снизу полумеры в вычислимую последовательность и затем сложить их с коэффициентами, получив наибольшую полумеру и тем самым завершив другое доказательство теоремы 41.

Фиксируем некоторую наибольшую перечислимую снизу полумеру, которую будем обозначать m. Значение $m(i)$ этой полумеры на числе i мы будем называть *априорной вероятностью* числа i. Смысл слов «априорная вероятность» такой. Пусть имеется некоторое устройство — чёрный ящик, при включении которого на выходе появляется некоторое натуральное число. Не зная ничего об устройстве ящика, мы хотим до его включения (*a priori*, как сказали бы философы) оценить сверху вероятность появления на выходе числа i. Так вот, если ящик считать вероятностной машиной, а в качестве оценки выбрать число $m(i)$, то мы не сильно занизим оценку (максимум — в константу раз). Иногда априорную вероятность называют *универсальной полумерой* на $\mathbb{N}$.

Как мы увидим, априорная вероятность числа i тесно связана с его сложностью. Грубо говоря, она тем больше, чем число проще. Это утверждение имеет вполне точный смысл: мы вскоре покажем, что несколько модифицированная сложность (так называемая «префиксная» сложность) числа i равна $-\log m(i)$.

4.3. Префиксные машины

[prefix-ma] Префиксная сложность отличается от обычной тем, что мы рассматриваем «самоограниченные описания»: декодирующей машине не сообщается, где кончается описание, а она решает это сама. Эту идею можно уточнять разными (и не эквивалентными) способами. Мы обсудим их подробно далее, а сейчас приведём формальные определения.

Пусть f — функция, аргументами которой являются двоичные слова. Будем говорить, что функция f является *префиксно корректной*, если выполнено такое свойство:

$$(f(x) \text{ определено}) \text{ и } (x \text{ — начало } y) \;\Rightarrow\; f(y) \text{ определено и равно } f(x).$$

Theorem 42. [prefix-corr-opt] *Среди префиксно корректных способов описания имеется оптимальный.*

$\lhd$ Напомним, что способом описания мы называем вычислимую функцию, аргументами и значениями которой являются двоичные слова. Теперь мы рассматриваем не все такие функции, как раньше, а лишь префиксно корректные, и сравниваем соответствующие меры сложности (сложность по-прежнему определяется как длина кратчайшего описания). Теорема утверждает, что в классе префиксно корректных способов описания имеется оптимальный *для этого класса* способ D; это означает, что для любого *префиксно корректного* способа D' имеет место неравенство $\mathit{KP}_D(x) \leqslant \mathit{KP}_{D'}(x) + O(1)$.

Здесь мы пишем KP вместо KS, чтобы подчеркнуть, что мы рассматриваем только префиксно корректные способы описания, хотя для данного способа D определение $\mathit{KP}_D(x)$ ничем не отличается от $\mathit{KS}_D(x)$.

Раньше (для обычной сложности) оптимальный способ описания строился так:

$$D(\hat{p}y) = p(y),$$

где $\hat{p}$ — какой-либо беспрефиксный код слова p (например, $\hat{p} = \overline{p}01$, где в $\overline{p}$ удвоен каждый бит слова p), а $p(y)$ — результат применения программы p к входу y (слово p понимается как программа в некотором универсальном языке программирования).

Будет ли D префиксно корректным? Легко понять, что нет, поскольку среди программ p есть всякие, в том числе и не префиксно корректные. Если какая-то программа p некорректна и, скажем, $p(0) = a$ и $p(00) = b$ при $a \neq b$, то $D(\hat{p}0) = a$ и $D(\hat{p}00) = b$, что противоречит требованию корректности для D.

Поэтому мы принудительно «скорректируем» программы и вместо $p(y)$ будем рассматривать $[p](y)$, определяемое так:

(1) Параллельно применяем программу p ко всех словам. Как только какое-то вычисление заканчивается, мы выписываем его аргумент и результат. Получаем последовательность пар $\langle y_i, z_i\rangle$, у которых $z_i = p(y_i)$.

(2) Эта последовательность прореживается с учётом требования корректности. Удобно использовать такую терминологию: слова y и y' *совместны*, если одно из них является началом другого (равносильная формулировка: если оба они являются началом некоторого третьего слова). Прореживание состоит в следующем: если какая-то пара $\langle y_i, z_i\rangle$ противоречит одной из предыдущих пар $\langle y_j, z_j\rangle$ при $j < i$, то она выбрасывается. Слово «противоречит» означает, что y_i совместно с y_j, но $z_i \neq z_j$. (Заметим, что можно было бы сравнивать очередную пару только с невыброшенными парами $\langle y_j, z_j\rangle$, но это не важно.)

(3) Все эти построения не зависят от входа y. Вычисляя $[p](y)$, мы ожидаем появления (невыброшенной) пары $\langle y_i, z_i\rangle$, у которой y_i является началом y. Как только такая пара появляется, мы выдаём результат z_i в качестве $[p](y)$.

Легко понять, что какова бы ни была программа p, функция $y \mapsto [p](y)$ является префиксно корректной. В самом деле, пусть $[p](y) = z$. Это значит, что в «прореженной» последовательности встретилась пара $\langle y_i, z\rangle$, где y_i является началом слова y. Пусть теперь y является началом слова y'. При вычислении $[p](y')$ мы также можем воспользоваться парой $\langle y_i, z\rangle$. Более ранние (невыброшенные) пары $\langle y_j, z_j\rangle$ нам либо не подойдут (если y_j не совместно с y_i, то y_j не может быть началом слова y'), либо дадут тот же результат (если y_j совместно с y_i, то $z_i = z_j$).

Если программа p с самого начала была префиксно корректной, то её корректировка ничего не меняет, то есть $[p](y) = p(y)$ при всех y. В самом деле, выброшенных пар не будет, и значение $[p](y)$ совпадёт со значением p на самом слове y или на каком-то его начале (что одно и то же в силу корректности p).

Осталось проверить, что построенный способ описания D является префиксно корректным и оптимальным (в классе префиксно корректных).

В самом деле, мы должны сравнить $D(\widehat{p_1}y_1)$ и $D(\widehat{p_2}y_2)$ в случае, когда $\widehat{p_1}y_1$ является началом $\widehat{p_2}y_2$. В этом случае $\widehat{p_1}$ и $\widehat{p_2}$ (будучи началами слова $\widehat{p_2}y_2$) совместны, и по свойствам беспрефиксного кодирования $p_1 = p_2$. Следовательно, y_1 есть начало y_2, и можно воспользоваться корректностью $[p_1]$ (или, что то же самое, $[p_2]$).

Корректность D доказана. Оптимальность вытекает из того, что $[p](y) = p(y)$ для корректного p, и потому переход от произвольного корректного D' (с программой p) к D увеличивает сложность не более чем на длину слова $\hat{p}$. $\triangleright$

Фиксировав произвольным образом оптимальный префиксно корректный способ описания, мы опускаем индекс и говорим о *префиксной сложности* $KP(x)$ слова x. Как обычно, надо иметь в виду, что замена способа описания приводит к изменению функции сложности (но только на ограниченное слагаемое).

Существует и другой вариант определения префиксной сложности, в котором требование корректности заменяется на другое. Будем называть функцию, аргументами которой являются двоичные слова, *беспрефиксной*, если никакие два слова в её области определения не сравнимы. Если беспрефиксная функция определена на каком-то слове, то она уже не может быть определена ни на его началах, ни на его продолжениях.

Будем рассматривать теперь только беспрефиксные способы описания, то есть беспрефиксные вычислимые функции, аргументами и значениями которых являются двоичные слова. Для них справедлива теорема, аналогичная теореме 42:

Theorem 43. `[prefix-opt]` *Среди беспрефиксных способов описания имеется оптимальный.*

$\triangleleft$ Доказательство следует той же схеме, но «корректировать» программы следует иначе. Для каждой программы p построим беспрефиксную функцию $y \mapsto \{p\}(y)$ следующим образом:

(1) Как и раньше, параллельно применяем p ко всем входам, получая последовательность пар $\langle y_i, z_i\rangle$, у которых $z_i = p(y_i)$.

(2) Эта последовательность прореживается: если в какой-то паре $\langle y_i, z_i\rangle$ слово y_i совместно с каким-то из предыдущих y_j (при $j < i$), то эта пара выбрасывается.

(3) Далее, имея вход y, мы ожидаем появления (невыброшенной) пары $\langle y_i, z_i\rangle$ с первым членом $y_i = y$. Второй член этой пары и будет $\{p\}(y)$.

Легко проверить, что функция $y \mapsto \{p\}(y)$ является беспрефиксной и что описанное преобразование оставляет нетронутыми беспрефиксные функции.

Далее рассуждение в точности повторяет доказательство теоремы 42. $\triangleright$

Теперь можно фиксировать какой-либо оптимальный беспрефиксный способ описания и рассмотреть соответствующую функцию сложности (обозначим её KP').

Какая же из мер сложности KP и KP' является «правильной»? Это скорее дело вкуса. Мы увидим вскоре (см. раздел 4.5), что на самом деле эти меры совпадают с точностью до $O(1)$ (и совпадают с минус логарифмом априорной вероятности), так что скорее следует спрашивать не о том, какая мера сложности лучше, а о том, какое определение более естественно. Это тем более вопрос вкуса. Авторам представляется, что более естественно определение с префиксно-корректными способами описания (не зря мы его привели первым). Вместе с тем в некоторых рассуждениях (например, при доказательстве теоремы о сложности пары в разделе 4.6) использование второго определения позволяет дать простое и короткое доказательство.

Насколько свойства префиксной сложности (KP, KP') отличаются от свойств обычной? Что из ранее известных свойств остаётся в силе?

- Прежде всего отметим очевидное свойство:

$$KS(x) \leqslant KP(x) + O(1) \quad \text{и} \quad KS(x) \leqslant KP'(x) + O(1),$$

  поскольку беспрефиксные и префиксно корректные способы описания, используемые при определении префиксной сложности, являются частным случаем способов описания из определения KS.

- Мы видели, что $KS(x) \leqslant l(x)+O(1)$ (достаточно рассмотреть тождественный способ описания). Теперь это рассуждение не проходит, так как тождественный способ описания не является ни беспрефиксным, ни префиксно корректным. Как мы покажем в разделе 4.5, это не случайно: для префиксной сложности аналогичное неравенство неверно.

- [prefix-complexity-length] Тем не менее можно дать оценки префиксной сложности в терминах длины. (Мы сделаем это для KP', конструкции для KP аналогичны.) Докажем, что $KP'(x) \leqslant 2l(x) + O(1)$. В самом деле, способ описания D, определённый формулой

$$D(\overline{x}01) = x$$

  ($\overline{x}$ получается удвоением всех битов в x), является беспрефиксным. Выбирая более экономное беспрефиксное кодирование $\hat{x}$ вместо $\overline{x}01$, можно получить и лучшие оценки:

$$KP'(x) \leqslant l(x) + 2\log l(x) + O(1)$$

  получится, если положить $\hat{x} = \overline{\mathrm{bin}(l(x))}01x$; итерируя эту конструкцию, получаем

$$KP'(x) \leqslant l(x) + \log l(x) + 2\log\log l(x) + O(1)$$

  и так далее.

- Свойство невозрастания сложности при алгоритмическом преобразовании остаётся верным:
$$KP'(A(x)) \leqslant KP'(x) + O(1)$$
(константа зависит от алгоритма A, но не от x). В самом деле, легко проверить, что если D — беспрефиксный способ описания, то композиция $x \mapsto A(D(x))$ также является беспрефиксным способом описания. Заменив слово «беспрефиксный» на «префиксно корректный», получаем аналогичное утверждение для KP вместо KP'. Это свойство позволяет говорить о префиксной сложности произвольных конструктивных объектов (пар слов, натуральных чисел, конечных множеств слов и т. п.), не уточняя способа их кодирования.

- Неравенство для сложности пары слов с префиксной сложностью верно без логарифмической добавки:
$$KP(x, y) \leqslant KP(x) + KP(y) + O(1)$$
(теорема 54 в разделе 4.6, с. 95).

- Свойство невозрастания сложности при алгоритмическом преобразовании можно применить, взяв в качестве A оптимальный способ описания D для обычной (не префиксной) колмогоровской сложности. Если p является кратчайшим описанием x относительно D, то $D(p) = x$ и $l(p) = KS(x)$, поэтому
$$KP(x) = KP(D(p)) \leqslant KP(p) + O(1) \leqslant l(p) + 2\log l(p) + O(1) = \\ = KS(x) + 2\log KS(x) + O(1).$$
Мы воспользовались (доказанным ранее) свойством $KP(p) \leqslant l(p) + 2\log l(p) + O(1)$; если использовать более точные оценки, то получим неравенство
$$KP(x) \leqslant KS(x) + \log KS(x) + 2\log\log KS(x) + O(1)$$
и аналогичные ему неравенства.

4.4. Отступление: машины с самоограниченным входом

[prefix-sd]

Этот раздел почти не используется в дальнейшем. Мы попытаемся проанализировать идею «самоограниченного» входа и её возможные уточнения, тем самым мотивировав определения префиксно корректных и беспрефиксных способов описания.

Обычно, подавая на вход машины двоичное слово, мы указываем начало и конец этого слова. Например, при определении вычислимости на машине Тьюринга обычно предполагают, что машина изначально видит первый символ слова, а конец слова отмечен специальным указателем (им может быть, например, символ «пробел», который идёт за последним битом слова).

Говоря о самоограниченном входе, мы имеем в виду несколько другую ситуацию: машина получает биты слова один из другим (слева направо) и в некоторый момент выдаёт ответ.

4.4.1. Префиксные функции

Вот одно из возможных уточнений. Будем считать, что у машины, помимо рабочей ленты, есть *входная лента*, на которой имеется односторонняя читающая головка. Крайняя левая клетка ленты содержит специальный маркер #, справа от которого записана бесконечная последовательность нулей и единиц (рис. 10).

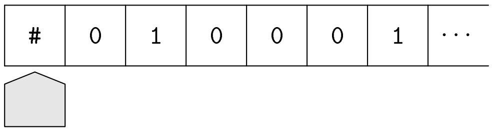

Рис. 10. Односторонняя читающая головка на входной ленте.
[read-only-tape]

Изначально читающая головка стоит у левого края ленты (и видит специальный маркер). Поведение машины Тьюринга определяется символом, который видит читающая головка, а также (как всегда) символом, который видит головка на рабочей ленте. В зависимости от этих символов и текущего состояния машина предпринимает то или иное действие. Это действие состоит в изменении внутреннего состояния, записи нового символа на рабочей ленте, а также может включать в себя сдвиг на рабочей ленте и сдвиг вправо читающей головки. Результат работы машины обычным образом записывается на рабочей ленте (изначально пустой).

Пусть M — такая машина. Будем запускать её на всевозможных входных лентах. Как только машина останавливается, мы записываем два слова: ту часть входной ленты, которую она успела прочесть (слово x), и результат работы (y). Множество полученных таким образом пар $\langle x, y\rangle$ обозначим Γ_M. Если $\langle x_1, y_1\rangle$ и $\langle x_2, y_2\rangle$ принадлежат Γ_M, то слова x_1 и x_2 несравнимы. В самом деле, если x_1 является началом x_2, то вычисление с x_2 на входе должно протекать так же, как и вычисление с x_1 (поскольку символы на входе те же) и вместе с ним должно закончиться. Поэтому часть слова x_2 останется непрочитанной, что противоречит определению множества Γ_M.

В частности, первые члены всех пар из Γ_M различны, поэтому Γ_M задаёт некоторую функцию γ_M, аргументами и значениями которой являются двоичные слова. Будем говорить, что машина M *беспрефиксно вычисляет* эту функцию. Легко понять, что функция γ_M вычислима в обычном смысле: чтобы вычислить $\gamma_M(x)$, надо запустить M на x и после остановки дополнительно проверить, что головка на входной ленте дошла до конца слова x, но не вышла за его пределы. Ясно также, что функция γ_M является беспрефиксной (любые два слова из её области определения несравнимы). Верно и обратное утверждение:

Theorem 44. *Любая беспрефиксная вычислимая функция беспрефиксно вычислима некоторой машиной.*

$\triangleleft$ Это утверждение вовсе не самоочевидно, поскольку машина F, вычисляющая (в обычном смысле) некоторую беспрефиксную функцию f, знает, где кончается вход, и может использовать эту информацию. Но если функция f беспрефиксная, мы можем построить другую машину M, для которой $\gamma_M = f$.

Это делается следующим образом. Машина M на рабочей ленте параллельно моделирует вычисления машины F на всех аргументах, время от времени читая очередные символы на входной ленте (когда именно это нужно делать, описано ниже). При появлении новой пары слов x, y, для которых $f(x) = y$, мы сравниваем x с уже прочитанной частью входа (вначале она пуста). Если прочитанная часть входа не является началом x, то мы не делаем ничего (и продолжаем порождать пары слов). Если прочитанная часть совпадает с x, то мы останавливаемся и выдаём на вход y. Если же прочитанная часть является собственным началом слова x, то мы читаем следующий бит входа и повторяем сравнение — до тех пор, пока мы не прочтём на входе либо само слово x (в этом случае M останавливается с выходом y), либо какое-то слово, не являющееся началом слова x. На этом обработка пары $\langle x, y\rangle$ заканчивается, и мы ждём появления следующей пары.

Как выглядит этот процесс вычисления? Вначале на входе не прочитано ничего. Когда появляется первая пара $\langle x, y\rangle$, мы смотрим, пусто ли x. Если да, то (так ничего и не прочтя на входе) печатаем на выходе y и останавливаемся. Если нет, то читаем вход до тех пор, пока либо не прочтём всё x, либо не отклонимся от x (то есть прочтём кусок входа, который не является началом слова x). В первом случае мы печатаем на выходе y и останавливаемся, во втором заканчиваем обработку пары $\langle x, y\rangle$ и ждём появления следующей пары.

Формально говоря, инвариант, который выполнен после обработки нескольких пар, таков (обозначим прочитанную часть входа через r): либо

(1) $f(r)$ определено, машина закончила работу и выдала на выход $f(r)$, либо

(2) r не является началом слова x ни для какой из обнаруженных пар $\langle x, y\rangle$, но всякое собственное начало r' слова r является собственным началом одного из таких слов. (Собственное начало — это начало, не совпадающее со всем словом.)

Используя этот инвариант, легко завершить формальную проверку, но мы вместо этого подчеркнём основную идею: если мы уже прочли некоторое слово r, и выяснилось, что функция f определена на некотором собственном продолжении слова r, то заведомо $f(r)$ не определено, и потому можно прочесть следующий бит входа, не опасаясь прочесть лишнее. $\triangleright$

Фактически та же модель вычислений может быть описана в более привычных для программистов терминах. Представим себе программу, которая использует оператор

$$b := \mathit{NextBit}$$

При выполнении этого оператора работа программы приостанавливается, на экране появляется надпись «Введите следующий бит». Когда клиент это делает (скажем, нажав клавишу «0» или «1»), соответствующий бит помещается в переменную b, и работа программы продолжается.

Каждой такой программе соответствует функция: её аргумент x бит за битом сообщается программе в ответ на её запросы; значение равно y, если программа напечатала y и остановилась, при этом запросив все биты слова x и только их. (Если она запросила следующий бит, прочтя всё слово x, или недоузнала биты этого слова, то x не входит в область определения функции.)

Адаптируя приведённые выше рассуждения, легко показать, что функции, соответствующие таким программам — это в точности вычислимые беспрефиксные функции. (Сдвиг головки на входной ленте в точности соответствует запросу следующего бита.)

4.4.2. Префиксно корректные функции

[prefix-nonblocking]

Есть и другая, более привычная схема работы программы, получающей свой вход бит за битом (причём конец входа никак не отмечается). Будем считать, что клиент набирает входное слово, не дожидаясь запросов, нажимая на клавиатуре клавиши «0» и «1» (и никак не обозначая конец входа). Нажатые им клавиши запоминаются и поступают в программу по требованию.

Очередной бит читается командой

$$b := \mathit{NextBit}$$

Помимо неё, есть команда

$$b := \mathit{NextExists}$$

Она позволяет выяснить, имеется ли в очереди нажатых клавиш ещё не прочитанная. Следует уточнить также, что происходит, если при выполнении команды *NextBit* очередь нажатых клавиш пуста. Можно считать, что это ведёт к аварии, а можно считать, что работа всего лишь приостанавливается до появления следующего бита (нажатия следующей клавиши). Какой из этих вариантов выбрать, не имеет значения, поскольку можно дожидаться появления входного бита, прежде чем его читать. Для этого нужно написать что-то вроде

while not *NextExists* **do** $\{$nothing$\}$;
b:=*NextBit*

Программисты бы назвали описываемый в этом разделе механизм доступа к входу «неблокирующим» чтением в отличие от ранее рассмотренного «блокирующего». Неблокирующее чтение позволяет программе продолжать внутреннюю работу, одновременно следя за входом и ожидая появления там нового символа.

При этом возникает очевидная проблема: вообще говоря, выход теперь зависит не только от входных битов, но также и от момента, в который они были поданы на вход.

Назовём программу *корректной*, если для неё результат работы не зависит от моментов нажатия клавиш: для данного входного слова x либо работа программы не заканчивается, в какие моменты его ни подавай, либо всегда заканчивается с одним и тем же результатом.

Каждой корректной программе соответствует частичная функция (зависимость выхода от входа).

Theorem 45. [prefix-correct-programs] **(а)** *Эта функция вычислима и является префиксно корректной.* **(б)** *Любая префиксно корректная вычислимая функция соответствует некоторой корректной программе.*

$\lhd$ (а) Вычислимость функции очевидна: если есть программа, вычисляющая её в описанном режиме (вход подаётся по частям), то есть и программа, вычисляющая её в обычном смысле. Проверим префиксную корректность. По определению (раздел 4.3) мы должны доказать, что если корректная программа выдаёт выход y на входе x, то она даёт тот же выход и для входа x', являющегося продолжением входа x. Это очевидно: проследим за работой программы на входе x. По предположению она напечатает y и остановится. В самый последний момент, когда машина уже всё сделала и собирается остановиться, добавим к входу x недостающий до x' кусок. Это уже не повлияет на работу программы, и она по-прежнему напечатает y при входе x'. В силу корректности программа будет печатать слово y при любом другом выборе моментов нажатия клавиш для букв из x'.

(Замечание в скобках: на самом деле спешка тут нужна лишь для наглядности, ведь мы можем нажать оставшиеся клавиши и после завершения работы программы, определение этого не запрещает.)

(б) Покажем, что для любой префиксно корректной вычислимой функции можно построить соответствующую программу. Пусть f — такая функция.

Корректная программа действует следующим образом. Она параллельно применяет f ко всем словам, а также регулярно проверяет, не поступил ли на вход новый символ (и читает его). Если выяснится, что $f(x) = y$ для некоторых x и y, причём текущее слово на входе равно x или некоторому продолжению слова x, то на выход выдаётся y и работа завершается.

Если $f(x) = y$, то эта программа выдаст y, в какие моменты не подавай биты слова x на вход. В самом деле, если подождать достаточно долго, то слово x будет подано целиком, а моделирование $f(x)$ завершится. В этот момент машина остановится и выдаст y, если она не остановилась раньше. А раньше она могла остановиться, если $f(x')$ оказалось определённым для некоторого начала x' входной последовательности. Но тогда это начало x' совместно с x (одно из двух слов является началом другого) и $f(x') = y$, поскольку функция f префиксно корректна по предположению. Поэтому на ответ это не повлияет.

С другой стороны, если $f(x)$ не определено, а функция f префиксно корректна, то $f(x')$ не определено и для всех начал слова x, поэтому работа машины не закончится. $\triangleright$

Эта теорема показывает, что корректным машинам с неблокирующим чтением соответствуют вычислимые префиксно корректные функции и только они, тем самым мотивируя определение префиксно корректной функции.

68 Докажите, что существует алгоритм, преобразующий любую программу p описанного вида (с командами *NextBit* и *NextExists*) в другую программу p' того же вида, причём p' всегда корректна и вычисляет ту же функцию, что и p, если программа p корректна. [Указание. Воспользуйтесь конструкцией из доказательства теоремы 45 в обе стороны.]

69 (Продолжение.) Докажите, что тем не менее не существует алгоритма, который по заданной программе p проверял бы, является ли она корректной. [Указание. Это делается обычным для теории вычислимых функций способом: сведением проблемы остановки. См., например, [77].]

4.4.3. Непрерывные вычислимые отображения

[prefix-sd-continuous] Существует и другая, более абстрактная мотивировка понятия префиксно корректной функции. Она исходит из общей схемы определения вычислимости для объектов высших типов, но мы изложим её применительно к конкретной ситуации (уже и так этот раздел — предназначенный дотошным читателям! — чрезмерно скучен).

Рассмотрим множество Σ конечных и бесконечных двоичных последовательностей: $\Sigma = \Xi \cup \Omega$. Для каждого (конечного) слова x рассмотрим множество Σ_x его конечных и бесконечных продолжений. Введём на Σ частичный порядок, считая, что $x \leqslant y$, если y является продолжением x.

Введём на множестве Σ топологию, считая базовыми открытыми множествами множества Σ_x (это значит, что открытыми множествами считаются всевозможные объединения множеств вида Σ_x). Легко проверить, что это действительно топология (не обладающая свойством отделимости).

Имеет место следующий (почти очевидный) факт:

Theorem 46. *Множество $A \subset \Sigma$ открыто в этой топологии тогда и только тогда, когда оно обладает двумя свойствами:*

(1) *вместе с каждым двоичным словом оно содержит все его (конечные или бесконечные) продолжения;*

(2) *если бесконечная последовательность принадлежит A, то некоторое её конечное начало также принадлежит A.*

$\triangleleft$ Объединение базовых открытых множеств очевидно обладает свойствами (1) и (2). С другой стороны, если A обладает этими свойствами, то оно является объединением множеств Σ_x по всем (конечным) словам x, принадлежащим A. $\triangleright$

Теперь введём топологию на множестве $\mathbb{N}_\perp$ натуральных чисел с добавленным к нему элементом $\perp$ (неопределённость). На этом множестве также полезно ввести частичный порядок, считая, что $\perp$ меньше всех остальных элементов, а между собой они несравнимы (рис. 11).

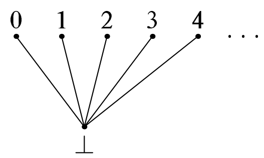

Рис. 11. Топологическое пространство $\mathbb{N}_\perp$

[n-bottom]

Будем считать открытыми любые множества, не содержащие элемента $\perp$, а также всё пространство $\mathbb{N} \cup \{\perp\}$. (Легко проверить, что аксиомы топологического пространства выполнены; пространство это также неотделимо.)

Частичные функции из Σ в $\mathbb{N}$ отождествим с отображениями вида $\Sigma \to \mathbb{N}_\perp$, определёнными на всём Σ (значение $\perp$ соответствует не определённым значениям функции). Легко описать класс непрерывных отображений (напомним, что отображение называется открытым, если прообраз любого открытого множества открыт):

Theorem 47. *Отображение $F\colon \Sigma \to \mathbb{N}_\perp$ непрерывно тогда и только тогда, когда выполнены два условия:*

(1) *F монотонно (если $x \leqslant y$, то $F(x) \leqslant F(y)$ в смысле введённых на Σ и $\mathbb{N}_\perp$ порядков);*

(2) *Если $F(x) \neq \perp$ для бесконечной последовательности x, то $F(x') \neq \perp$ для некоторого конечного начала $x' \leqslant x$.*

$\lhd$ Пусть F непрерывно. Проверим условие (1). Пусть $x \leqslant y$. Если $F(x) \nleqslant F(y)$, то $F(x)$ есть натуральное число (а не $\perp$) и $F(x) \neq F(y)$. Тогда прообраз открытого множества $\{F(x)\}$ содержит x, но не содержит y и потому не является открытым.

Проверим условие (2). Если $F(x) \neq \perp$ для бесконечной последовательности x, то открытый прообраз открытого множества $\{F(x)\}$ содержит x и потому должен содержать некоторое конечное начало последовательности x.

Осталось проверить, что если для F выполнены условия (1) и (2), то отображение F непрерывно. Достаточно проверить, что прообраз каждого натурального числа открыт (прообраз всего пространства открыт, а остальные открытые множества состоят из натуральных чисел). Для этого достаточно проверить условия (1) и (2) предыдущей теоремы, которые прямо следуют из предположения. (Заметим, что если x' есть начало x и $F(x') \neq \perp$, то $F(x') = F(x)$ в силу монотонности.) $\rhd$

С каждым непрерывным отображением $F\colon \Sigma \to \mathbb{N}_\perp$ свяжем множество Γ_F всех пар $\langle x, n\rangle \in \Xi \times \mathbb{N}$, для которых $F(x) = n$. Заметим, что множество Γ_F является лишь частью графика отображения F (мы рассматриваем лишь конечные слова x и не разрешаем элементу $\perp$ быть вторым членом пары.)

Theorem 48. *Соответствие $F \mapsto \Gamma_F$ является взаимно однозначным соответствием между непрерывными отображениями $\Sigma \to \mathbb{N}_\perp$ и множествами $A \subset \Xi \times \mathbb{N}$, обладающими двумя свойствами:*

(1) $\langle x, n\rangle \in A,\ x \leqslant y \ \Rightarrow\ \langle y, n\rangle \in A$;

(2) $\langle x, n\rangle \in A,\ \langle x, m\rangle \in A \ \Rightarrow m = n$.

$\lhd$ Пусть отображение F непрерывно. Если $F(x) = n \in \mathbb{N}$, то условие (1) предыдущей теоремы гарантирует, что $F(y) = n$ для любого слова y, продолжающего x. Тем самым условие (1) выполнено для множества Γ_F. Выполнено и условие (2), поскольку $F(x)$ не может быть равно двум разным натуральным числам одновременно. Таким образом, для любого непрерывного F множество Γ_F обладает свойствами (1) и (2).

Легко видеть, что Γ_F однозначно определяет F: чтобы найти $F(x)$ для конечного x, мы ищем пару $\langle x, n\rangle \in \Gamma_F$. Если такой пары нет, то $F(x) = \perp$. Для бесконечного x значение $F(x)$ однозначно определяется по непрерывности.

Осталось показать, что любое множество A, обладающее свойствами (1) и (2), равно Γ_F при некотором F. В самом деле, определим $F(x)$ при конечном x как то единственное n, для которого $\langle x, n\rangle \in A$ (свойство (2) гарантирует единственность) или как $\perp$, если такого n не существует. Свойство (1) гарантирует, что построенная таким образом функция (определённая на конечных словах) монотонна. Осталось доопределить её на бесконечных словах. Для бесконечного $x \in \Sigma$ положим $F(x)$ равным $F(x')$, если найдётся конечное $x' \leqslant x$, для которого $F(x') \neq \perp$; если такого x' не найдётся, то $F(x) = \perp$. Корректность определения (независимость $F(x)$ от выбора $x' \leqslant x$) гарантируется свойством (1). Очевидно, построенная таким образом функция F обладает свойствами (1) и (2) предыдущей теоремы и потому непрерывна. Столь же очевидно, что $\Gamma_F = A$. $\rhd$

Легко понять, что свойства (1) и (2) последней теоремы означают, что A представляет собой график префиксно корректной функции. Тем самым мы получаем взаимно однозначное соответствие между непрерывными отображениями $\Sigma \to \mathbb{N}_\perp$ и префиксно корректными функциями.

Назовём непрерывное отображение $F\colon \Sigma \to \mathbb{N}_{\perp}$ ***вычислимым***, если множество Γ_F перечислимо. Легко проверить, что это равносильно вычислимости соответствующей частичной функции из Ξ в $\mathbb{N}$. (Частичная функция из Ξ в $\mathbb{N}$ вычислима тогда и только тогда, когда её график перечислим.) Поэтому вычислимые отображения типа $\Sigma \to \mathbb{N}_{\perp}$ соответствуют вычислимым префиксно корректным функциям, что и является обещанной дополнительной мотивировкой понятия префиксно корректной функции.

4.5. Основная теорема о префиксной сложности

[prefix-eq]

В этом разделе мы докажем, что три меры сложности — два варианта префиксной (*KP* с префиксно корректными способами описания и *KP*$'$ с беспрефиксными), а также минус логарифм априорной вероятности совпадают с точностью до $O(1)$. Для этого мы докажем три неравенства в цепочке

$$-\log m(x) \leqslant \mathit{KP}(x) \leqslant \mathit{KP}'(x) \leqslant -\log m(x)$$

(с точностью до $O(1)$). Начнём с двух простых неравенств.

Theorem 49. [kp-kpprime]

$$\mathit{KP}(x) \leqslant KP'(x) + O(1).$$

$\triangleleft$ Неравенство было бы очевидным, если бы всякий беспрефиксный способ описания был бы префиксно корректным. Однако это не так: если беспрефиксный способ D определён на каком-то слове u, то он не определён ни на каком продолжении слова u (в то время как префиксно корректный способ описания должен быть определён на всех продолжениях и иметь то же значение).

Поэтому требуется чуть более сложная конструкция. Пусть D — беспрефиксный способ описания. Построим новый способ описания D'. Чтобы вычислить $D'(y)$, мы применяем D ко всем началам слова y. Лишь одно из значений $D(y')$ (где y' — начало слова y) может быть определённым (в силу беспрефиксности). Как только такое y' найдётся, мы полагаем $D'(y)$ равным $D(y')$. Другими словами, $D'(y) = x$ тогда и только тогда, когда $D(y') = x$ для некоторого начала y' слова y.

Из построения видно, что функция D' вычислима, префиксно корректна и является продолжением функции D. Поэтому сложность относительно D' не превосходит сложности относительно D. (На самом деле сложность относительно D' в точности равна сложности относительно D, так как кратчайшие описания остались теми же самыми.) $\triangleright$

Можно пытаться доказать обратное неравенство аналогичным способом. Действительно, для каждого префиксно корректного отображения D можно построить беспрефиксное отображение D', сузив D на «нижние точки» своей области определения (то есть положив $D'(y) = z$, если $D(y) = z$ и $D(y')$ не определено ни для какого начала y' слова y).

Это преобразование префиксно корректных отображений в беспрефиксные в точности обратно описанному выше преобразованию беспрефиксных отображений в префиксно корректные. Но тут мы сталкиваемся с проблемой, поскольку новое преобразование (в отличие от прежнего) не сохраняет вычислимость.

$\boxed{70}$ Приведите пример вычислимой префиксно корректной функции, для которой соответствующая беспрефиксная функция не является вычислимой. [Указание. Пусть A — перечислимое множество натуральных чисел, не являющееся разрешимым и потому имеющее неперечислимое дополнение. Положим $f(0^n 11x) = 0$ при всех n и всех x и $f(0^n 1x) = 0$ при всех $n \in A$ и всех x.]

Эта задача в некотором смысле говорит, что неблокирующее чтение является более гибким средством, чем блокирующее (см. раздел 4.4).

Theorem 50. `[m-kp]`

$$-\log m(x) \leqslant \mathit{KP}(x) + O(1).$$

$\lhd$ Мы должны показать, что $-\log m(x) \leqslant \mathit{KP}(x) + O(1)$ или что $2^{-\mathit{KP}(x)} \leqslant cm(x)$ для некоторой константы c. В силу максимальности полумеры m достаточно показать, что функция $x \mapsto 2^{-\mathit{KP}(x)}$ не больше некоторой перечислимой снизу полумеры. (Здесь, говоря о полумерах, мы считаем аргументами двоичные слова, то есть рассматриваем вероятностные машины, выходом которых являются двоичные слова, см. раздел 4.1.)

Укажем эту полумеру, построив соответствующую вероятностную машину. Она действует так: оптимальный префиксно корректный способ описания D (использованный в определении KP) применяется ко всем началам последовательности случайных битов $b_0, b_1, b_2, \ldots$. Как только одно из вычислений

$$D(\Lambda), D(b_0), D(b_0 b_1), D(b_0 b_1 b_2), \ldots$$

закончится, его результат и будет результатом работы машины. Заметим, что не имеет значение, какое именно из закончившихся вычислений мы возьмём (корректность гарантирует, что результат один и тот же).

Для любого слова x рассмотрим кратчайшее описание p слова x (относительно D). Тогда вероятность появления слова x как выхода вероятностной машины не меньше $2^{-l(p)}$. В самом деле, если первые $l(p)$ случайных битов совпали с p, то результат работы машины гарантированно равен x. Тем самым вероятность появления x не меньше $2^{-l(p)} = 2^{-\mathit{KP}(x)}$, что и требовалось. $\rhd$

Несколько другое доказательство того же результата обходится без вероятностных машин. В самом деле, мы знаем, что функция сложности перечислима сверху, поэтому функция $x \mapsto 2^{-\mathit{KP}(x)}$ перечислима снизу. Остаётся сослаться на следующую почти очевидную теорему:

Theorem 51. `[prefix-coding-bound]`

$$\sum_x 2^{-\mathit{KP}(x)} \leqslant 1.$$

$\lhd$ В самом деле, пусть p_x — кратчайшая программа для слова x. Тогда слова p_x и p_y несравнимы (при $x \neq y$). Остаётся воспользоваться такой простой леммой.

Лемма. Пусть $p_0, p_1, p_2, \ldots$ — попарно несравнимые слова (это значит, что ни одно из них не является началом другого). Тогда $\sum_i 2^{-l(p_i)} \leqslant 1$.

В самом деле, $2^{-l(p_i)}$ есть (равномерная бернуллиева) мера множества Ω_{p_i} всех (бесконечных) продолжений слова p_i. Поскольку слова p_i несравнимы, множества их продолжений не пересекаются и потому меры этих множеств (вероятности получить p_i в начале случайной последовательности битов) в сумме не превосходят единицы. Лемма (а с ней и теорема 51) доказана. $\triangleright$

Кстати, мы заодно доказали, что для префиксной сложности не выполняется неравенство $KP(x) \leqslant l(x) + O(1)$. В самом деле, если бы это было так, то сумма

$$\sum_x 2^{-l(x)}$$

была бы конечной. А между тем для каждого n сумма по всем x длины n равна единице (ибо состоит из 2^n членов по 2^{-n} каждый), и потому общая сумма бесконечна.

$\boxed{71}$ Докажите, что более слабая оценка $KP(x) \leqslant l(x) + \log l(x) + O(1)$ также не имеет места (то есть что разность $KP(x) - l(x) - \log l(x)$ не ограничена). [Указание. Гармонический ряд расходится.]

Осталось доказать третье, наиболее сложное неравенство:

Theorem 52. [kpprime-m]

$$KP'(x) \leqslant -\log m(x) + O(1).$$

$\triangleleft$ Попытаемся вначале объяснить идею доказательства. Имеется перечислимая снизу полумера $m(x)$. Это значит, что мы постепенно узнаём всё бо́льшие и бо́льшие оценки снизу для $m(x)$. Большое значение $m(x)$ означает для нас, что $KP'(x)$ должно быть малым, то есть означает необходимость предусмотреть для x короткое описание p. При этом описания различных объектов должны быть несравнимыми. Несравнимость описаний p_1 и p_2 означает, что отрезки I_{p_1} и I_{p_2} не пересекаются. (Напомним, что отрезок I_p — не считая концов — состоит из чисел, двоичная запись которых начинается на слово p.) При этом неравенство $l(p) \leqslant -\log_2 m(x)$ можно переписать как $2^{-l(p)} \geqslant m(x)$: длина отрезка I_p должна быть не меньше $m(x)$.

Тем самым нашу задачу можно сформулировать так: каждому объекту x нужно выделить отрезок длины не менее $m(x)$, причём отрезки, выделенные разным объектам, не должны перекрываться.

Сказанное требует уточнений. Во-первых, разрешается выделять отрезки длиной не $m(x)$, а $\varepsilon m(x)$ для некоторого фиксированного ε (что соответствует аддитивной константе в исходном неравенстве). Во-вторых, нужны не произвольные отрезки, а только «регулярные» (половинки, четвертинки и другие отрезки вида I_p). Но с учётом первого замечания это не так важно, поскольку любой отрезок (не выходящий за пределы отрезка $[0,1]$) содержит строго внутри себя регулярный подотрезок длиной не менее четверти исходного.

Таким образом, мы приходим почти к той же задаче, что была рассмотрена в разделе 4.1. К нам приходят клиенты с просьбами выделить им место на отрезке $[0,1]$. По-прежнему сумма всех запросов не превосходит единицы и клиенты сообщают свои запросы постепенно (время от времени меняя их в сторону увеличения). Теперь, однако, клиентам важна не общая длина выделенных для них отрезков, а длина непрерывного участка, что усложняет задачу. В качестве компенсации мы имеем право выделять место с некоторым коэффициентом (урезать все запросы в фиксированное число раз). Но это нам не поможет, если при каждом увеличении запроса отводить новый участок — при такой стратегии никакого коэффициента не хватит. Решение проблемы: рассматривать новый запрос (выделяя новый отрезок) только, когда запрос увеличился вдвое. При этом общая сумма рассмотренных запросов не более чем вдвое превосходит последний запрос (ведь сумма геометрической прогрессии со знаменателем 2 не больше удвоенного последнего члена). С другой стороны, требования выполнены с точностью до мультипликативной константы.

Эту схему можно довести до вполне строгого доказательства. Мы, однако, изложим немного другое доказательство. Оно основано на следующей «лемме о шторах», называемой часто леммой Крафта – Чейтина (см. [6]). (Она является вычислимым аналогом леммы Крафта из теории информации, см. с. 185).

Лемма. [kraft-chaitin] Пусть $l_0, l_1, l_2, \ldots$ — вычислимая последовательность натуральных чисел, для которой

$$\sum_i 2^{-l_i} \leqslant 1.$$

Тогда существует вычислимая последовательность попарно несравнимых двоичных слов $x_0, x_1, x_2, \ldots$, для которых $l(x_i) = l_i$.

Заметим, что указанное в лемме неравенство необходимо для существования несравнимых слов x_i с длинами l_i, поскольку отрезки I_{x_i} не перекрываются и имеют длины как раз 2^{-l_i}. Лемма утверждает, что это необходимое условие одновременно является достаточным.

Наглядно можно сформулировать лемму так. Есть окно шириной 1 (отрезок $[0,1]$). Нам приносят одну за другой шторы, которыми мы занавешиваем окно (так, чтобы они не перекрывались). При этом шторы могут иметь ширину $1, 1/2, 1/4, 1/8, \ldots$, и вешать их можно не куда угодно, а с ограничениями. Штору размера 1 можно повесить единственным способом. Штору размера $1/2$ можно вешать двумя способами (закрыв левую половину окна или закрыв правую). Для шторы размера $1/4$ имеется четыре возможных положения (четыре четверти отрезка), и так далее. Выбор положения для шторы шириной 2^{-l} соответствует выбору слова длины l. Требование, чтобы шторы не перекрывались, означает, что соответствующие им слова несравнимы. Наконец, условие леммы утверждает, что суммарная ширина штор не превосходит ширины окна.

(Более компьютерная метафора: распределение памяти между процессами; каждый процесс требует $1/2^k$ доступной памяти, и требует, чтобы выделенный участок был выровнен [aligned]; память не освобождается; мы доказываем, что если сумма всех запросов не больше общего размера памяти, то их можно удовлетворить в порядке поступления, применяя описанный далее алгоритм.)

Выбор положения для очередной шторы происходит по такому алгоритму: свободная часть окна представляется в виде объединения непересекающихся «виртуальных штор», причём размеры всех виртуальных штор различны. Виртуальная штора обозначает место, которое пока свободно, но на которое потом можно будет повесить реальную штору того же размера. (Свободная память представлена в виде списка правильно выровненных блоков разного размера.)

Вначале окно свободно, и есть одна виртуальная штора ширины 1. Пусть теперь нам приносят штору ширины w. Если среди виртуальных штор имеется штора ширины w, то всё просто: изымаем её из набора и вешаем на её место реальную. Пусть это не так. Заметим, что тогда среди виртуальных штор обязательно есть штора ширины больше w. (В самом деле, иначе в наборе были бы только виртуальные шторы $w/2, w/4, \ldots$, и общая ширина свободного пространства была бы меньше w. Напомним, что все виртуальные шторы разного размера.)

Итак, мы знаем, что есть хотя бы одна виртуальная штора ширины $w' > w$. Выберем из таких штор самую узкую (подобная стратегия при распределении памяти называется best fit). Отрежем от неё слева кусок размера w, а остальное разрежем на части $w, 2w, 4w, \ldots, w'/2$ (ведь $w + w + 2w + 4w + \ldots + (w'/2) = w'$). Эти части и будут новыми виртуальными шторами. Поскольку w' было выбрано минимальным, штор промежуточного размера (от w до $w'/2$) раньше не было, так что наше условие (требующее, чтобы виртуальные шторы были разных размеров) продолжает выполняться. Лемма доказана.

$\boxed{72}$ Докажите, что ту же стратегию можно описать другими словами: каждую следующую штору вешаем на самое левое из возможных мест. [Указание: описанная в лемме конструкция сохраняет такое свойство: длины виртуальных штор возрастают слева направо.]

Следствие. Пусть l_i — вычислимая последовательность натуральных чисел, причём $\sum_i 2^{-l_i} \leqslant 1$. Тогда $KP'(i) \leqslant l_i + O(1)$.

В самом деле, лемма позволяет указать последовательность несравнимых слов x_i длины l_i. Определим способ описания D, положив $D(x_i) = i$. Несравнимость слов x_i гарантирует, что этот способ описания будет беспрефиксным. Вычислимость последовательности x_i гарантирует вычислимость D (имея вход x, мы сравниваем его по очереди со всеми x_i до совпадения, после чего выдаём на выход i).

(Заметим, что мы, как и раньше, свободно переходим от натуральных чисел к словам, говоря об их сложности и априорной вероятности.)

Вернёмся к доказательству теоремы. Мы рассматриваем наибольшую (максимальную) перечислимую снизу полумеру m. Согласно определению перечислимости снизу,

$$m(x) = \lim_{i\to\infty} m(x,i),$$

где $m(x,i)$ — вычислимая и монотонная по i функция двух аргументов с рациональными значениями. Округлим $m(x,i)$ в сторону увеличения до ближайшей степени двойки (то есть до одного из чисел $1, 1/2, 1/4, 1/8, \ldots$), при этом нули оставляем нулями. Назовём результат округления $m'(x,i)$. По-прежнему m' — вычислимая функция, монотонная по второму аргументу. Очевидно, $m'(x,i)$ не меньше $m(x,i)$, но превосходит $m(x,i)$ не более чем вдвое.

Назовём пару $\langle x,i\rangle$ *граничной*, если $m'(x,i)$ больше $m'(x,i-1)$ (или если $i=0$ и $m'(x,0) > 0$). Граничные пары отмечают те места, где запросы $m'(x,i)$ увеличиваются с ростом i (при данном x).

Покажем, что сумма всех $m'(x,i)$ по всем граничным парам $\langle x,i\rangle$ не превосходит 4. Для этого достаточно показать, что сумма $m'(x,i)$ по всем граничным i (для данного x) не превосходит $4m(x)$. А это делается так. В указанной сумме (для фиксированного x) каждый следующий член по крайней мере вдвое больше предыдущего, поэтому вся сумма не превосходит удвоенного последнего члена. А он, в свою очередь, не более чем вдвое превосходит $m(x,i)$ для какого-то i. Вспоминая, что $m(x,i) \leqslant m(x)$, заключаем, что интересующая нас сумма не больше $4m(x)$, что и требовалось.

Множество всех граничных пар $\langle x,i\rangle$ разрешимо. В самом деле, по паре можно проверить, является ли она граничной, посмотрев на рациональные числа $m'(x,i)$ и $m'(x,i-1)$.

Перенумеровав все пары в каком-то порядке и отобрав граничные, построим вычислимую последовательность $\langle x_0,i_0\rangle,\langle x_1,i_1\rangle,\ldots$ содержащую каждую граничную пару по одному разу. Рассмотрим вычислимую последовательность чисел l_n, заданных соотношением

$$2^{-l_n} = m'(x_n,i_n)/4.$$

Тогда по доказанному

$$\sum_n 2^{-l_n} = (1/4)\sum_n m'(x_n,i_n) \leqslant 1,$$

и потому $KP'(n) \leqslant l_n + O(1)$. Поскольку x_n алгоритмически получается по n, то (как мы видели в разделе про префиксную сложность)

$$KP'(x_n) \leqslant KP'(n) + O(1) \leqslant l_n + O(1) = -\log m'(x_n,i_n) + O(1).$$

Данное слово x появляется среди x_n столько раз, сколько имеется граничных пар вида $\langle x,i\rangle$. Если $\langle x_n,i_n\rangle$ — граничная пара с данным x и наибольшим i, то $m'(x_n,i_n)$ больше или равно $m(x,i)$ при всех i (поскольку округляем мы в сторону увеличения, и больших граничных пар нет). Поэтому в этом случае $m'(x_n,i_n) \geqslant m(x)$. Отсюда следует, что

$$KP'(x) \leqslant -\log m(x) + O(1),$$

что и требовалось доказать. $\triangleright$

Итак, три доказанных неравенства позволяют заключить, что KP, KP' и $-\log m$ отличаются не более чем на $O(1)$.

Имея это в виду, мы не будем различать KP и KP' (кроме тех случаев, когда в каком-то рассуждении годится только одно из этих двух определений).

Отметим ещё (это нам пригодится при доказательстве критерия случайности в терминах монотонной сложности, раздел 5.6), что фактически мы доказали нечто большее:

Theorem 53. [prefix-explicit] *По любой перечислимой снизу последовательности $p_0,p_1,\ldots$, у которой $\sum_i p_i \leqslant 1$, можно эффективно указать беспрефиксный способ описания D, для которого $KP'_D(i) \leqslant -\log_2 p_i + 2$.*

(Имеется в виду, что дан алгоритм, перечисляющий множество всех пар $\langle r,i\rangle$, у которых $r < p_i$. Описанная в доказательстве теоремы конструкция алгоритмически преобразует его в беспрефиксный способ описания с нужным свойством.)

4.6. Свойства префиксной сложности

[prefix-pr]

Вернёмся теперь к свойствам префиксной сложности. Для начала покажем, как некоторые уже известные нам свойства могут быть получены с использованием априорной вероятности (универсальной полумеры).

Ряд $\sum 1/n^2$, как известно из курса анализа, сходится. Умножив его на подходящую константу, получаем перечислимую снизу полумеру (с суммой меньше единицы). Поэтому априорная вероятность числа n не меньше c/n^2 при некотором c, то есть

$$KP(n) \leqslant 2\log n + O(1).$$

Пусть x_n — слово номер n в последовательности $\Lambda, 0, 1, 00, 01, 10, 11, 000, \ldots$. Тогда

$$KP(x_n) \leqslant KP(n) + O(1) \leqslant 2\log n + O(1) = 2l(x_n) + O(1)$$

(легко проверить, что x_n есть двоичная запись числа $n+1$ без старшей цифры 1 и потому имеет длину $\log n + O(1)$). (Строго говоря, случай $n = 0$ следовало бы разбирать отдельно, поскольку $1/0^2$ и $\log 0$ не определены, но необходимые исправления очевидны.) Получается оценка $KP(x) \leqslant 2l(x) + O(1)$.

Чтобы получить лучшую оценку префиксной сложности (уже встречавшуюся нам на с. 81), рассмотрим более медленно сходящийся ряд

$$\sum \frac{1}{n\log^2 n}$$

(его сходимость можно установить, например, с помощью интегрального признака сходимости и интегрирования по частям). Он даёт оценку $KP(n) \leqslant \log n + 2\log\log n + O(1)$, которая после перехода к словам превращается в

$$KP(x) \leqslant l(x) + 2\log l(x) + O(1).$$

Можно и дальше улучшать оценку, рассмотрев ряды $\sum 1/(n\log n(\log\log n)^2)$, $\sum 1/(n\log n\log\log n(\log\log\log n)^2)$ и так далее.

Докажем теперь обещанное неравенство для энтропии пары.

Theorem 54. [prefix-pair]

$$KP(x, y) \leqslant KP(x) + KP(y) + O(1).$$

$\triangleleft$ Под $KP(x, y)$ мы понимаем сложность слова $[x, y]$, где $\langle x, y\rangle \mapsto [x, y]$ — какое-либо вычислимое однозначное кодирование пар. (С точностью до ограниченного слагаемого сложность пары не зависит от выбора кодирования, поскольку переход от одного кодирования к другому — вычислимое преобразование.)

Рассмотрим функцию m', определённую соотношением

$$m'([x, y]) = m(x)m(y),$$

где m — априорная вероятность. (Если не всякое слово является кодом пары, полагаем $m'(z) = 0$ для слов z, не являющихся кодами.) Очевидно, m' перечислима снизу (перемножая вычислимые последовательности, сходящиеся снизу к $m(x)$ и $m(y)$, получаем возрастающую вычислимую последовательность с нужным пределом). Кроме того,

$$\sum_z m'(z) = \sum_{x,y} m'([x, y]) = \sum_{x,y} m(x)m(y) = \sum_x m(x)\sum_y m(y) \leqslant 1\cdot 1 = 1.$$

Следовательно, m' — перечислимая снизу полумера. Сравнивая её с априорной вероятностью, находим, что $m'([x,y]) \leqslant cm([x,y])$ для некоторого c, и потому

$$KP([x,y]) \leqslant KP(x) + KP(y) + O(1),$$

что и требовалось доказать. $\triangleright$

73 [m-projection] Докажите, что $\sum_y m([x,y])$ отличается от $m(x)$ не более чем на мультипликативную константу (в обе стороны).

74 Докажите, что для любой возрастающей вычислимой функции $f\colon \mathbb{N} \to \mathbb{N}$ величина $\sum\{m(k) | f(n) \leqslant k < f(n+1)\}$ отличается от $m(n)$ не более чем на мультипликативную константу (в обе стороны): каким вычислимым способом ни разбивай сходящийся ряд $\sum m(n)$ на группы, получается тот же самый (с точностью до ограниченного множителя) ряд!

Попытаемся теперь доказать теорему 54 в терминах описаний. Здесь обнаруживается такой любопытный факт: определение с префиксно корректными способами не помогает, а с беспрефиксными всё получается. Итак, мы доказываем, что $KP'([x,y]) \leqslant KP'(x) + KP'(y) + O(1)$. Пусть D — оптимальный беспрефиксный способ описания, использованный при определении KP'.

Определим новый способ описания формулой

$$D'(pq) = [D(p), D(q)],$$

где pq означает конкатенацию слов p и q. Эта формула имеет смысл, если $D(p)$ и $D(q)$ определены. Требуется, однако, проверить, что это определение корректно, то есть что одно и то же слово x не может быть представлено двумя разными способами $x = pq = p'q'$, если $D(p), D(q), D(p'), D(q')$ все определены. В самом деле, в этом случае слова p и p' согласованы (они являются началами одного и того же слова x, и потому одно из слов p и p' есть начало другого), поэтому $D(p)$ и $D(p')$ не могут быть одновременно определены при $p \neq p'$ (функция D является беспрефиксной). А раз $p = p'$, то и $q = q'$.

Аналогичным образом легко проверить, что функция D' является беспрефиксной: если pq есть начало $p'q'$, то p и p' согласованы. Раз $D(p)$ и $D(p')$ определены, то $p = p'$. Поэтому q является началом q'; раз $D(q)$ и $D(q')$ определены, то $q = q'$.

Функция D' вычислима: чтобы найти $D'(x)$, мы разбиваем x на p и q всеми возможными способами и параллельно вычисляем $D(p)$ и $D(q)$ для всех вариантов. Как мы видели, удачным может быть не более одного разбиения, и оно даёт ответ (если существует).

Остаётся заметить, что

$$KP_{D'}([x,y]) \leqslant KP_D(x) + KP_D(y).$$

В самом деле, если p и q — кратчайшие описания для x и y относительно D, то pq — описание $[x,y]$ относительно D', имеющее длину $KP_D(x) + KP_D(y)$.

Неформально машину D' можно описать так: она читает вход, пока не прочтёт там описание для x. После этого остаток входа читается как описание для y.

75 Докажите теорему 54, используя определение беспрефиксных способов описания в терминах машин с блокирующим чтением.

$\boxed{76}$ Назовём множество слов *беспрефиксным*, если никакие два слова в нём не сравнимы. Покажите, что если A и B — беспрефиксные множества слов, то и множество

$$AB = \{ab \mid a \in A, b \in B\}$$

является беспрефиксным.

В данном случае рассуждение с априорной вероятностью было, пожалуй, проще. В следующем примере это уже не так.

Theorem 55. [prefix-addco]

$$\mathit{KP}(x, \mathit{KP}(x)) = \mathit{KP}(x) + O(1).$$

Аналогичное равенство для обычной (не префиксной) сложности составляло содержание задачи 19.

$\lhd$ Легко видеть, что $\mathit{KP}(x) \leqslant \mathit{KP}(x, \mathit{KP}(x)) + O(1)$, поскольку x получается алгоритмически по $[x, \mathit{KP}(x)]$. (Квадратные скобки обозначают вычислимое однозначное кодирование, используемое при определении префиксной сложности пары.)

Для доказательства обратного неравенства снова используем KP', то есть беспрефиксные способы описания. Пусть D — оптимальный беспрефиксный способ описания. Рассмотрим новый способ описания

$$D'(p) = [D(p), l(p)],$$

имеющий ту же область определения, что и D, и потому беспрефиксный. Если p — кратчайшее описание слова x относительно D, то $l(p) = \mathit{KP}'(x)$, и потому p является описанием пары $[x, \mathit{KP}'(x)]$ относительно D'. Поэтому $\mathit{KP}_{D'}([x, \mathit{KP}'(x)]) \leqslant l(p) = \mathit{KP}'(x)$.

Теорема доказана? На самом деле тут есть тонкость, требующая разъяснения. Мы доказали её для функции KP', соответствующей беспрефиксному способу описания. Если заменить её на KP (соответствующую префиксно корректному способу описания), то правая часть изменится не более чем на константу. Вопрос в том, почему это верно и для левой части, поскольку KP стоит внутри в качестве аргумента.

Чтобы ответить на этот вопрос, достаточно показать, что $\mathit{KP}(x, n)$ (сложность пары, в которой второй аргумент мы считаем натуральным числом) меняется не более чем на константу при изменении n на единицу. Это следует из того, что отображения $[x, n] \mapsto [x, n+1]$ и $[x, n] \mapsto [x, n-1]$ вычислимы и потому могут увеличивать сложность не более чем на константу. $\rhd$

Поучительно провести доказательство с помощью априорной вероятности. Пусть $m(x)$ — априорная вероятность слова x. Рассмотрим функцию m', определённую так:

$$m'([x, k]) = \begin{cases} 2^{-k}, & \text{если } 2^{-k} < m(x); \\ 0 & \text{в противном случае.} \end{cases}$$

Эта функция перечислима снизу (для данных x и k мы выдаём нуль, пока не выяснится, что $2^{-k} < m(x)$, а потом выдаём 2^{-k}).

Сумма $m'([x, k])$ по всем k (при данном x) представляет собой геометрическую прогрессию со знаменателем 2, и потому не больше $2m(x)$ (последний член прогрессии меньше

$m(x)$). Следовательно, сумма $m'([x,k])$ по всем x и k конечна. Сравнивая m' с априорной вероятностью и переходя к логарифмам, находим, что

$$KP(x,k) \leqslant k + O(1)$$

при $2^{-k} < m(x)$. Теперь положим $k = -\lfloor \log m(x) \rfloor + 1$, тогда $2^{-k} < m(x)$ и получается, что

$$KP(x, -\lfloor \log m(x) \rfloor + 1) \leqslant KP(x) + O(1).$$

Остаётся (как и раньше) заметить, что изменение второго члена пары на единицу меняет сложность не более чем на константу, и второе доказательство теоремы завершено.

$\boxed{77}$ Приведённое только что доказательство теоремы 55 фактически доказывает чуть больше: $KP(x,m) \leqslant m + O(1)$ при $KP(x) \leqslant m$. Как вывести это из утверждения теоремы 55, не апеллируя к доказательству?

Что можно сказать об алгоритмических свойствах функции $KP(x)$? Как и обычная колмогоровская сложность, она перечислима сверху, но не вычислима и даже не имеет никакой нетривиальной (неограниченной) вычислимой нижней оценки. (Поскольку $KP(x) \leqslant 2\,KS(x) + O(1)$, любая нетривиальная нижняя оценка для KP немедленно давала бы нетривиальную нижнюю оценку для KS.)

Ранее (теорема 8, с. 25) мы видели, что $KS(x)$ можно определить как минимальную перечислимую сверху функцию K, для которой множества $\{x \mid K(x) < n\}$ имеют размер $O(2^n)$. Аналогичное утверждение для префиксной сложности выглядит так:

Theorem 56. [kp-minimal-convergent] *KP есть минимальная (с точностью до $O(1)$) перечислимая сверху функция K, для которой сумма ряда $\sum_x 2^{-K(x)}$ конечна.*

◁ Как мы видели, функция KP перечислима сверху и ряд $\sum_x 2^{-KP(x)}$ сходится. Если K — перечислимая сверху функция с теми же свойствами, то функция $M(x) = c2^{-K(x)}$ при достаточно малом рациональном c является полумерой (она перечислима снизу и сумма меньше 1). Следовательно, $M(x) = O(m(x))$ в силу максимальности априорной вероятности, и потому $\log M(x) \leqslant \log m(x) + O(1)$, то есть $KP(x) \leqslant K(x) + O(1)$. ▷

Переформулировка: пусть функция f с натуральными значениями определена на всех словах и перечислима сверху. Тогда свойства «$KP(x) \leqslant f(x) + O(1)$» и $\sum_x 2^{-f(x)} < \infty$ равносильны.

Конечность суммы ряда в только что доказанной теореме является более сильным условием на функцию K, чем использованное в теореме 8 условие (если сумма ряда не больше C, то количество членов ряда, бóльших некоторой границы ε, не превосходит C/ε). Тем самым мы заново получаем, что $KS(x) \leqslant KP(x) + O(1)$.

Поучительно сравнить обычную и префиксную сложности по двум параметрам: средней сложности слов данной длины и числу слов сложности не больше данной. Начнём с первого.

Мы видели, что большинство слов длины n имеют (обычную) сложность $n + O(1)$ (см с. 14, а также задачу 2 на с. 23). Естественно ожидать, что префиксная сложность будет несколько больше.

Theorem 57. [average-kp] (а) *$KP(x) \leqslant l(x) + KP(l(x)) + O(1)$.*

(б) *Найдётся такая константа c, что для любого n и для любого d доля слов x длины n, у которых $KP(x) < n + KP(n) - d$, не превосходит $c2^{-d}$.*

$\lhd$ (а) Пусть $m(x)$ — априорная вероятность слова x. Мы будем также говорить об априорной вероятности $m(n)$ натурального числа n. Рассмотрим функцию $m'(x)$, равную $2^{-n}m(n)$ для слов x длины n. Значения этой функции на всех словах длины n одинаковы и в сумме дают $m(n)$, поэтому $\sum_x m'(x) \leqslant 1$. Функция m' перечислима снизу. Поскольку полумера m является наибольшей, заключаем, что $m'(x) \leqslant cm(x)$ для некоторого c и для всех x. Переходя к логарифмам, получаем неравенство

$$KP(x) \leqslant n + KP(n) + O(1)$$

для слов длины n (константа в $O(1)$ не зависит от n).

(б) Теперь будем двигаться в обратном направлении и рассмотрим функцию

$$m'(n) = \sum_{l(x)=n} m(x)$$

(сумму априорных вероятностей всех слов данной длины). Поскольку эта функция перечислима снизу и $\sum_n m'(n) \leqslant 1$, то $m'(n) = O(m(n))$ (где $m(n)$ — априорная вероятность натурального числа n, или, что тоже самое, его двоичной записи). С другой стороны, слово из n нулей имеет (с точностью до константы) ту же априорную вероятность, что и n, так что

$$c_1 m(n) \leqslant \sum_{l(x)=n} m(x) \leqslant c_2 m(n).$$

Другими словами, сумма значений функции $m(x)$ на словах длины n с точностью до константы совпадает с $m(n)$, а среднее значение $m(x)$ на этих словах — с $m(n)/2^n$. Остаётся заметить, что доля тех аргументов, для которых функция в 2^d раз превышает своё среднее, не превосходит 2^{-d} (неравенство Чебышёва). $\rhd$

$\boxed{78}$ Докажите, что среднее арифметическое префиксных сложностей всех слов длины n равно $n + KP(n) + O(1)$.

(Для обычной сложности это сделано в задаче 3, с. 23.)

Теперь оценим число слов данной сложности.

Theorem 58. [bounded-kp-cardinality] *Логарифм числа слов x, для которых $KP(x) < n$, равен $n - KP(n) + O(1)$.*

$\lhd$ Пусть c_n — число слов, для которых $KP(x) < n$. Перепишем основное свойство префиксной сложности (сходимость ряда $\sum 2^{-KP(x)}$) в терминах чисел c_n. Имеется $c_{n+1} - c_n$ слов, сложность которых равна в точности n, поэтому сумму ряда можно переписать как $\sum_n 2^{-n}(c_{n+1} - c_n)$. Перегруппировав члены, убеждаемся, что сумма $\sum_n (2^{-(n-1)} - 2^{-n})c_n = \sum_n 2^{-n}c_n$ конечна. Поскольку c_n перечислимо снизу, то отсюда следует, что $2^{-n}c_n$ не превосходит априорной вероятности $m(n)$ числа n, откуда $c_n \leqslant m(n)2^n$ (с точностью до мультипликативной константы).

С другой стороны, легко построить перечислимую сверху функцию K, аргументами которой являются двоичные слова, а значениями натуральные числа и плюс бесконечность, у которой будет примерно $m(n)2^n$ значений, равных n. Например, можно договориться, что на словах длины n эта функция равна либо n, либо $+\infty$, причём изначально все значения бесконечны, а затем по мере появления всё бóльших оценок для $m(n)$ нужное число бесконечных значений заменяется на n.

Для этой функции сумма $\sum 2^{-K(x)}$ сходится, поэтому $\mathit{KP}(x) \leqslant K(x) + O(1)$, и потому $c_{n+O(1)} \geqslant m(n)2^n$. Остаётся вспомнить, что $m(n)$ и 2^n меняются не более чем в константу раз при изменении n на единицу. $\triangleright$

Последние две теоремы оценивают разницу между простой и префиксной сложностью в терминах сложности длины этого слова (раньше мы видели аналогичные оценки, где вместо сложности длины был логарифм длины). Вот ещё несколько оценок такого типа.

Неравенство $\mathit{KP}(x) \leqslant l(x) + \mathit{KP}(l(x))$ можно итерировать, получая

$$\mathit{KP}(x) \leqslant l(x) + l(l(x)) + \mathit{KP}(l(l(x)) + O(1),$$
$$\mathit{KP}(x) \leqslant l(x) + l(l(x)) + l(l(l(x))) + \mathit{KP}(l(l(l(x))) + O(1)$$

и так далее. Кроме того, можно вспомнить, что если D — оптимальный способ описания для обычной (не префиксной) сложности, то $\mathit{KP}(D(y)) \leqslant KP(y)$; сочетая это с неравенством $\mathit{KP}(x) \leqslant l(x) + \mathit{KP}(x) + O(1)$, получаем следующее утверждение:

Theorem 59. [kp-ks-bound]

$$\mathit{KP}(x) \leqslant \mathit{KS}(x) + \mathit{KP}(\mathit{KS}(x))) + O(1),$$
$$\mathit{KP}(x) \leqslant \mathit{KS}(x) + \mathit{KS}(\mathit{KS}(x))) + \mathit{KP}(\mathit{KS}(\mathit{KS}(x)) + O(1),$$
$$\mathit{KP}(x) \leqslant \mathit{KS}(x) + \mathit{KS}(\mathit{KS}(x))) + \mathit{KS}(\mathit{KS}(\mathit{KS}(x))) + \mathit{KP}(\mathit{KS}(\mathit{KS}(\mathit{KS}(x))) + O(1)$$

и так далее.

Отметим кстати, что все эти неравенства получаются также повторным применением первого из них.

Мы привели лишь самые простые результаты, связывающие простую и префиксную сложность. Более подробно об этом можно прочесть в [52].

4.7. Условная префиксная сложность и сложность пары

[prefix-co]

4.7.1. Условная префиксная сложность

[prefix-co-def] Как определить префиксный вариант условной сложности? Каждое из приведённых нами определений можно модифицировать, добавив туда условия. Вот что при этом получится.

Функция D двух аргументов называется *префиксно корректной относительно первого аргумента*, если при любом фиксированном значении второго аргумента получается префиксно корректная функция:

$D(y, z)$ определено и $y \leqslant y' \Rightarrow D(y', z) = D(y, z)$.

Здесь предполагается, что первый аргумент является двоичным словом; $y \leqslant y'$ означает, что y' является продолжением y.

При определении условной сложности в разделе 2.2 мы рассматривали всевозможные вычислимые функции двух аргументов, называя их *способами условного описания*. (В качестве аргументов и значений рассматривались двоичные слова.) Равенство $D(y, z) = x$ читалось так: y является описанием x при известном z. Сложность x при известном z определялась как длина кратчайшего описания. Среди всех способов описания выбирался оптимальный, при котором сложность меньше всего (с точностью до константы).

`[prefix-corr-opt-cond]` Ограничимся теперь префиксно корректными по первому аргументу способами описания. Среди них тоже существует оптимальный. Доказательство этого факта аналогично доказательству теоремы 42 на с. 79 (о существовании оптимального префиксно корректного способа безусловного описания), только все приведённые там рассуждения надо параллельно провести для всех условий z. Именно, надо положить

$$D'(\hat{p}y, z) = [p](y, z),$$

где $[p](y, z)$ означает результат принудительной префиксной коррекции программы p (как функции от y при каждом z в отдельности).

Выбрав оптимальный префиксно корректный способ условного описания, сложность относительно него мы называем условной префиксной сложностью и обозначаем её $\mathit{KP}(x|z)$ (читается: «условная префиксная сложность слова x при известном слове z).

Это определение можно видоизменить, заменив требование префиксной корректности на требование беспрефиксности по первому аргументу. И для этого класса существует оптимальный способ условного описания; соответствующую сложность будем обозначать $\mathit{KP}'(x|z)$. Можно доказать, что эти определения различаются не более чем на константу:

$$\mathit{KP}'(x|z) = \mathit{KP}(x|z) + O(1),$$

причём константа эта не зависит не только от x, но и от z.

Как и для безусловного случая, прямое доказательство этого равенства затруднительно, и в качестве промежуточного звена используется условная априорная вероятность $m(x|z)$. Её (как и раньше) можно определить двумя способами. Первый способ использует вероятностные машины с входом. Пусть M — такая машина. Для каждого входного слова z возникает своё распределение вероятностей: выход x (при входе z) имеет некоторую вероятность $p_M(x|z)$. Функция $\langle x, z\rangle \mapsto p_M(x|z)$ перечислима снизу (в естественном смысле), и для каждого z сумма $\sum_x p_M(x|z)$ не превосходит единицы. Эти свойства описывают класс всех функций p_M: если имеется перечислимая снизу неотрицательная функция $\langle x, z\rangle \mapsto p(x|z)$, для которой $\sum_x p(x|z) \leqslant 1$ при всех z, то можно построить вероятностную машину M, для которой $p_M = p$.

Среди всех функций p_M существует наибольшая с точностью до умножения на константу. Мы фиксируем одну из таких наибольших функций, называем её *условной априорной вероятностью слова x при известном слове z* и обозначаем $m(x|z)$.

Далее мы доказываем, что $-\log m(x|z) \leqslant \mathit{KP}(x|z) + O(1)$ и что $\mathit{KP}'(x|z) \leqslant -\log m(x|z) + O(1)$. Неравенство $\mathit{KP}(x|z) \leqslant \mathit{KP}'(x|z) + O(1)$, которое доказывается столь же просто, как и в безусловном случае, замыкает круг: все три величины $KP(x|z)$, $KP'(x|z)$ и $-\log m(x|z)$ отличаются не более чем на константу.

Мы не приводим подробно всех доказательств, поскольку они состоят в параллельном проведении «безусловных» доказательств для всех условий z. Можно сказать, что мы имеем дело с «релятивизацией», но несколько особого вида.

Поясним сказанное. В общей теории алгоритмов «релятивизация» означает, что вместо вычислимых функций мы рассматриваем функции, вычислимые относительно некоторого оракула A. (Здесь A — произвольное множество слов. Считается, что алгоритм, отвечающий на вопросы о принадлежности слова x множеству A, доступен как внешняя процедура с входным параметром x.) Практически все известные теоремы общей теории алгоритмов без труда переносятся и на такие («A-вычислимые») функции.

Кстати, и понятие сложности можно релятивизовать таким образом, определив для любого множества A колмогоровскую сложность $KS^A(x)$ и префиксную колмогоровскую сложность $KP^A(x)$ (см. раздел 6.4). Но сейчас мы делаем не это: в качестве оракула мы рассматриваем доступ к конечному слову z. Конечно, такой оракул не может изменить понятие вычислимости (слово z конечно), но он меняет колмогоровскую сложность, поскольку информация, имеющаяся в z, «выносится за скобки» и не учитывается при подсчёте сложности. В результате вместо обычной колмогоровской сложности $KS(x)$ мы получаем условную сложность $KS(x|z)$. Другой пример: величину $I(x:y|z)$ можно считать релятивизованной относительно z версией понятия взаимной информации $I(x:y)$.

Ещё одно важное замечание. Во всех наших рассмотрениях какие бы то ни было требования, связанные с префиксами (началами), относятся только к описаниям. Напротив, и описываемые объекты, и условия для нас лишены какой бы то ни было структуры.

Такой подход не является единственно возможным. Можно учитывать отношение «быть началом» для описываемых объектов (так делают при определении монотонной сложности и сложности разрешения, см. главы 5 и 6). Можно было бы также учитывать отношение «быть началом» и для условий (см. раздел 6.3); такие варианты определений вполне осмысленны, хотя и мало изучены.

Отметим также, что при определении условной префиксной сложности требования, связанные с префиксами, накладываются отдельно при каждом значении условия (второго аргумента функции). Например, мы требуем, чтобы машина сама решала, когда она кончит читать описание — но это решение может зависеть от второго аргумента. Из-за этого утверждение, аналогичное задаче 23 (с. 40), уже не имеет места:

$\boxed{79}$ `[prefix-conditional-as-problem]` Покажите, что $KP(y|x)$ не превосходит минимальной префиксной сложности программ, переводящих x в y, но обратное неверно (все неравенства понимаются с точностью до $O(1)$). (Оба утверждения верны для любого языка программирования; аддитивная константа может зависеть от выбора языка программирования.) [Указание. Легко видеть, что $KP(y|l(y)) \leqslant l(y) + O(1)$, так как при известном $l(y)$ каждое слово можно считать своей самоограниченной программой. Если бы обратное утверждение было верно, то отсюда мы получили бы 2^n различных программ префиксной сложности не более n.]

4.7.2. Свойства условной префиксной сложности

Перечислим несколько простых свойств условной префиксной сложности.

- $KP(x|z) \leqslant KP(x) + O(1)$.

  В самом деле, любой префиксно корректный (или беспрефиксный) способ безусловного описания $y \mapsto D(y)$ можно рассматривать как префиксно корректный (соответственно беспрефиксный) способ условного описания $\langle y, z\rangle \mapsto D(y)$ (второй аргумент фиктивен).

  В терминах полумер: всякая вероятностная машина без входа может рассматриваться как вероятностная машина со входом, игнорирующая свой вход. Соответственно любая перечислимая снизу полумера $q(x)$ может рассматриваться как семейство полумер $q'(x|z) = q(x)$ с фиктивным параметром z.

- $KP(x|x) = O(1)$.

  Способ описания $D(y, z) = z$ является префиксно корректным (нас интересует префиксная корректность по y, а не по z) и $KP_D(x|x) = 0$. Беспрефиксный способ описания можно определить так: $D(\Lambda, z) = z$, где Λ — пустое слово; для непустых первых аргументов D не определён. Наконец, перечислимое снизу семейство полумер таково: $q(x|x) = 1$ и $q(x|z) = 0$ при $z \neq x$.

- $KP(f(x, z)|z) \leqslant KP(x|z) + O(1)$ для любой вычислимой функции f и для любых слов x, z, на которых f определена. (При этом константа в $O(1)$ может зависеть от f, но не от x и z.)

  Если D — оптимальный префиксно корректный [беспрефиксный] способ условного описания, то $D'\colon \langle y, z\rangle \mapsto f(D(y, z), z)$ также будет префиксно корректным [соответственно беспрефиксным] и $KP_{D'}(f(x, z)|z) \leqslant KP_D(x|z)$, откуда и следует искомое неравенство.

  В терминах полумер: если $m(x|z)$ — априорная вероятность x при известном z, рассмотрим полумеру
  $$q(x|z) = \sum\{m(x'|z) \mid f(x', z) = x\}$$
  (при каждом z она представляет собой прямой образ полумеры $x \mapsto m(x, z)$ при отображении $x \mapsto f(x, z)$); легко проверить, что функция q перечислима снизу, что $\sum_x q(x|z) \leqslant 1$ и что $q(f(x, z)|z) \geqslant m(x|z)$. Используя оптимальность m, приходим к искомому неравенству (для логарифмов априорных вероятностей).

- $KP(x|z) \leqslant KP(x|f(z)) + O(1)$ для любой вычислимой функции f и любых x, z (если $f(z)$ определено; константа может зависеть от f, но не от x и z).

  (Достаточно рассмотреть способ описания $\langle y, z\rangle \mapsto D(y, f(z))$ или семейство полумер $q(x|z) = m(x|f(z))$.)

- $KP(f(x)|x) = O(1)$ для любой вычислимой функции f и для всех x, при которых $f(x)$ определено.

  (Простое следствие предыдущих.)

- $KS(x|z) \leqslant KP(x|z) + O(1)$

В самом деле, префиксно корректные (или беспрефиксные) способы условного описания входят в число способов условного описания, рассматриваемых при определении $KS(x|z)$.

- $KP(x|z) \leqslant KS(x|z) + 2\log KS(x|z) + O(1)$

  Это утверждение можно вывести из предыдущих. Пусть D — оптимальный способ условного описания (не обязательно префиксно корректный). Тогда

  $$KP(D(y,z)|z) \leqslant KP(y|z) + O(1) \leqslant KP(y) + O(1) \leqslant l(y) + 2\log l(y) + O(1).$$

  Если в качестве y взято кратчайшее описание x при известном z, то $l(y) = KS(x|z)$.

  Аналогичным образом можно доказать и более сильное неравенство

  $$KP(x|z) \leqslant KS(x|z) + \log KS(x|z) + 2\log\log KS(x|z) + O(1)$$

  и так далее.

4.7.3. Префиксная сложность пары

Мы видели (теорема 54, с. 95), что $KP(x,y) \leqslant KP(x) + KP(y) + O(1)$. Это неравенство можно усилить:

Theorem 60. `[prefix-pair2]`

$$KP(x,y) \leqslant KP(x) + KP(y|x) + O(1).$$

$\lhd$ Есть два варианта доказательства: с беспрефиксными способами описания и с полумерами. С беспрефиксными способами описания мы рассуждаем так. Пусть D — оптимальный беспрефиксный способ безусловного описания, а D_c — оптимальный беспрефиксный способ условного описания. Рассмотрим способ описания D', определяемый формулой

$$D'(uv) = [D(u), D_c(v, D(u))]$$

(для тех u и v, для которых правая часть определена). Как и при доказательстве теоремы 54, легко проверить, что D' корректно определён и является беспрефиксным способом описания; при этом конкатенация кратчайших описаний для x и для y (при известном x) даст описание для $[x,y]$.

(Заметим, что в этом рассуждении существен порядок u и v: если бы мы взяли vu (вместо uv), то сначала надо было бы выделять слово v, и возникла бы трудность: функция D_c является беспрефиксной при фиксированном втором аргументе, а $D(u)$ пока ещё не известно.)

С полумерами: пусть $m(x)$ — априорная вероятность x, а $m(y|x)$ — априорная вероятность y при известном x. Рассмотрим функцию m', определённую соотношением

$$m'([x,y]) = m(x)m(y|x)$$

(мы считаем, что $m'(z) = 0$ для слов z, не являющихся кодами пар). Тогда m' перечислима снизу (как произведение двух перечислимых снизу функций), и

$$\sum_z m'(z) = \sum_{x,y} m(x)m(y|x) = \sum_x \bigl[m(x)\sum_y m(y|x)\bigr] \leqslant \sum_x m(x) \leqslant 1.$$

Поэтому $m([x,y]) \geqslant \varepsilon m'([x,y]) = \varepsilon m(x)m(y|x)$, что и требовалось доказать. $\triangleright$

$\boxed{80}$ Докажите, что $KS(x,y) \leqslant KP(x) + KS(y|x) + O(1)$.

[Указание. Можно использовать беспрефиксные способы описания, дописав к беспрефиксному описанию для x обычное описание для y при известном x. А можно подсчитать число пар, при которых $KP(x) + KS(y|x) \leqslant n$. Если первое слагаемое равно k, то пар не больше $2^k \cdot m(k) \cdot 2^{n-k} = 2^n m(k)$, и после суммирования по k получаем $2^n \cdot O(1)$.]

Только что доказанное неравенство можно несколько усилить. Для начала заметим, что мы можем рассматривать в качестве условий не только слова, но и пары слов (поскольку замена одного способа кодирования пар на другой меняет сложность не более чем на константу). Кроме того, можно говорить о сложности троек слов. После этих замечаний можно записать такую цепочку неравенств (для краткости члены $O(1)$ опускаем):

$$KP(x,y) \leqslant KP(x, KP(x), y) \leqslant KP(x, KP(x)) + KP(y|x, KP(x)) = KP(x) + KP(y|x, KP(x)).$$

При этом мы использовали равенство $KP(x, KP(x)) = KP(x)$ (теорема 55), а также только что доказанную теорему про энтропию пары. Полученное неравенство усиливает эту теорему, поскольку $KP(y|x, KP(x)) \leqslant KP(y|x)$ (x получается алгоритмически по $[x, f(x)]$). Как обнаружили Л. А. Левин (см. [18]) и Г. Чейтин [6], такое уточнение замечательным образом превращает неравенство в равенство:

Theorem 61. [prefix-pair3]

$$KP(x,y) = KP(x) + KP(y|x, KP(x)) + O(1).$$

$\triangleleft$ В одну сторону неравенство нами уже доказано (см. обсуждение перед формулировкой теоремы). Можно привести и прямое доказательство: чтобы получить беспрефиксный код пары $\langle x, y\rangle$, надо к беспрефиксному коду x приписать справа беспрефиксный код y при известных x и $KP(x)$ (причём в качестве $KP(x)$ надо взять как раз длину написанного беспрефиксного кода для x). Заметим, что после выделения первой части мы знаем не только x, но и $KP(x)$, так что есть всё необходимое для восстановления слова y.

В терминах полумер это доказательство можно изложить так: рассмотрим функцию m', для которой

$$m'([x,y]) = \sum_{\{k \mid 2^{-k} < 2m(x)\}} 2^{-k} m(y|x,k).$$

Эта функция перечислима снизу, сумма по всем x, y конечна (сумма $m(y|x,k)$ по y не больше 1; сумма 2^{-k} по всем k, для которых $2^{-k} < 2m(x)$, не больше $4m(x)$, и суммирование по всем x даёт не больше 4). Остаётся сравнить её с априорной вероятностью m и заметить, что при $k = -\lfloor \log_2 m(x) \rfloor$ получаем как раз нужный член суммы.

Перейдём теперь к доказательству обратного неравенства:

$$KP(x) + KP(y|x, KP(x)) \leqslant KP(x,y) + O(1).$$

Для начала приведём (неправильное, но более простое) доказательство (неправильного, но более сильного) утверждения

$$\mathit{KP}(x) + \mathit{KP}(y|x) \leqslant \mathit{KP}(x, y) + O(1).$$

Переходя к полумерам, можно переписать это неравенство так:

$$m(x)m(y|x) \geqslant \varepsilon m([x, y])$$

(для некоторого ε и для всех x, y). Здесь буквой m обозначены априорные вероятности (условная и безусловная). Перепишем это неравенство как

$$m(y|x) \geqslant \varepsilon\, \frac{m([x, y])}{m(x)}.$$

Достаточно доказать, что функция

$$m'(y|x) = \varepsilon\, \frac{m([x, y])}{m(x)}$$

при любом фиксированном x является полумерой (при некотором ε), после чего сослаться на максимальность функции $m(y|x)$. Просуммируем $m'(y|x)$ по y. Нам надо убедиться, что сумма

$$\sum_y m'(y|x) = \varepsilon\, \frac{\sum_y m([x, y])}{m(x)}$$

не превосходит единицы.

Почему это так? В самом деле, функция $x \mapsto \sum_y m([x, y])$ является полумерой (её сумма по всем x есть $\sum_{x,y} m([x, y]) \leqslant 1$) и потому эта функция не больше $m(x)/\varepsilon$ при некотором ε.

Где ошибка в этом рассуждении? Мы забыли, что полумера должна быть перечислима снизу. В одном из двух случаев это действительно так: функция $\sum_y m([x, y])$, как легко понять, будет перечислимой снизу, поскольку функция m была таковой. Но вот про $m([x, y])/m(x)$ этого сказать уже нельзя, поскольку $m(x)$ стоит в знаменателе, а при возрастании знаменателя дробь не увеличивается (как нам бы хотелось), а уменьшается.

Правильное доказательство более слабого неравенства следует той же схеме, но обходит указанную только что трудность. Мы должны доказать, что при $z = \mathit{KP}(x)$ имеет место неравенство

$$m(y|x, z) \geqslant \varepsilon\, \frac{m([x, y])}{m(x)}.$$

Наша проблема, напомним, в том, что правая часть не является перечислимой снизу. Но при $z = \mathit{KP}(x)$ можно заменить $m(x) \approx 2^{-\mathit{KP}(x)}$ на 2^{-z} и рассмотреть функцию

$$m'(y|x, z) = m([x, y])2^z.$$

Эта функция перечислима снизу. Зато она может не быть полумерой: сумма $\sum_y m'(y|x, z)$ не всегда будет ограничена единицей. Это будет так, лишь если

$$\sum_y m([x, y]) \leqslant 2^{-z}.$$

Мы видели, что $\sum_y m([x,y]) = O(m(x)) = O(2^{-KP(x)})$, поэтому для некоторой константы c выполнено такое свойство:

$$z \leqslant KP(x) - c \Rightarrow \sum_y m'(y|x,z) \leqslant 1.$$

Это хорошо, но мало: нам нужно семейство полумер, в котором такое неравенство выполняется при всех x и z, а не только при некоторых. Поэтому мы исправим функцию m', получив функцию m'' с такими свойствами:

- функция $\langle y,x,z\rangle \mapsto m''(y|x,z)$ перечислима снизу;
- неравенство
$$\sum_y m''(y|x,z) \leqslant 1$$
выполнено при всех x и z;
- для некоторой константы c
$$z \leqslant KP(x) - c \Rightarrow m''(y|x,z) = m'(y|x,z).$$

Технология такого рода исправления уже обсуждалась в разделе 4.2: строя всё бо́льшие приближения снизу, мы проверяем, не нарушают ли они границу для суммы, и если нарушают, то отбрасываем.

Сравнивая m'' с априорной условной вероятностью и переходя к логарифмам, мы видим, что

$$z \leqslant KP(x) - c \Rightarrow KP(y|x,z) \leqslant KP(x,y) - z + c'$$

для некоторых констант c и c' и для всех x, y, z.

Чтобы вывести отсюда утверждение теоремы, достаточно подставить $z = KP(x) - c$ в полученное неравенство и заметить, что при изменении z на единицу значение $KP(y|x,z)$ меняется не более чем на константу (функция прибавления единицы ко второму члену пары вычислима, аналогично для вычитания), поэтому $KP(y|x, KP(x) - c) = KP(y|x, KP(x)) + O(1)$. $\triangleright$

Заметим, что теорема 22 (с. 43), утверждающая, что $KS(x,y) = KS(x) + KS(y|x) + O(\log n)$ для слов сложности не больше n, является следствием только что доказанной.

В самом деле, замена KP на KS меняет все три члена в равенстве на более чем на $O(\log n)$. Остаётся заметить, что $KS(y|x, KP(x))$ отличается от $KS(y|x)$ не более чем на $O(\log n)$. Тем самым мы получили новое доказательство теоремы 21, в которой комбинаторные подсчёты заменены оценками вероятностей.

Вспоминая, что $m(x) \approx \sum_y m([x,y])$ с точностью до ограниченного множителя (задача 73, с. 96), можно переписать утверждение теоремы 61 так:

$$m(y|x, KP(x)) \approx \frac{m([x,y])}{\sum_y m([x,y])}$$

Правую часть этого равенства можно интерпретировать как условную вероятность того, что второй член пары равен y, при условии «первый член пары равен x».

$\boxed{81}$ Докажите, что

$$KP(x|z) \leqslant KP(x|y) + KP(y|z) + O(1)$$

для любых слов x, y, z. (Это неравенство можно ещё усилить, заменить $KP(x|y)$ на меньшую величину $KP(x|y, z)$.)

$\boxed{82}$ Докажите «релятивизованный» вариант теоремы 61:

$$KP(x, y|z) = KP(x|z) + KP(y|x, KP(x|z), z) + O(1).$$

Дважды применяя теорему 61, можно получить формулу для префиксной сложности тройки слов. Эту тройку можно рассматривать как пару, первый член которой есть пара $\langle x, y\rangle$, а второй — z. Получаем, что

$$KP(x, y, z) = KP(z|x, y, KP(x, y)) + KP(x, y) + O(1).$$

Ещё раз применяя теорему 61, получаем ответ:

Theorem 62. [prefix-triple]

$$KP(x, y, z) = KP(z|x, y, KP(x, y)) + KP(y|x, KP(x)) + KP(x) + O(1).$$

Можно изменить порядок действий (применяя на втором шаге «релятивизованный» относительно z вариант теоремы 61):

$$\begin{aligned} KP(x, y, z) = KP(y, z|x, KP(x)) + KP(x) = \\ = KP(z|y, KP(y|x, KP(x)), x, KP(x)) + KP(y|x, KP(x)) + KP(x) \end{aligned}$$

(для краткости мы опускаем слагаемые $O(1)$).

Получилась немного другая формула по сравнению с утверждением теоремы 62: два последних слагаемых в ней те же, но первое слагаемое изменилось. Как и раньше, там стоит условная сложность слова z, но вместо условия $KP(x, y)$ появилось два условия $KP(x)$ и $KP(y|x, KP(x))$ — две сложности, которые в сумме дают как раз $KP(x, y)$ по теореме 61. Поэтому пара этих сложностей содержит не меньше информации, чем $KP(x, y)$. Более удивительно, что (при известных x и y) верно и обратное. В самом деле, подставим в качестве z пару $\langle KP(x), KP(y|x, KP(x))\rangle$; во второй формуле первое слагаемое обратится в нуль (точнее, в $O(1)$). Приравняв правые части формул, получаем такое следствие:

Theorem 63. [prefix-pair-paradox]

$$\begin{gathered} KP(KP(x)|x, y, KP(x, y)) = O(1), \\ KP(KP(y|x, KP(x))|x, y, KP(x, y)) = O(1). \end{gathered}$$

(Разумеется, аналогичные утверждения верны для $KP(y)$ и $KP(x|y, KP(y))$.)

$\boxed{83}$ Дайте прямое доказательство теоремы 63. [Указание. Если мы знаем x, y и $KP(x,y)$, то можем искать верхнюю оценку d для $KP(x)$, для которой $KP(y|x,d)+d$ сравняется с $KP(x,y)$. Совпадение с точностью до $O(1)$ будет означать, что $d = KP(x)+O(1)$: если d превышает $KP(x)$ на некоторое число m, то $KP(y|x,d)$ может упасть за счёт этого не более чем на $O(\log m)$, поэтому в сумме будет проигрыш.]

С помощью теоремы 61 легко показать, что базисное неравенство (теорема 24, с. 49) для префиксной сложности выполняется с точностью $O(1)$ (для обычной сложности была лишь логарифмическая оценка):

Theorem 64. [prefix-baseineq]

$$KP(x,y,z) + KP(x) \leqslant KP(x,y) + KP(x,z) + O(1)$$

для любых трёх слов x, y, z.

$\lhd$ В самом деле, правую часть можно переписать как

$$KP(x) + KP(y|x, KP(x)) + KP(x) + KP(z|x, KP(x)),$$

а левую — как

$$KP(x) + KP(y,z|x, KP(x)) + KP(x),$$

и остаётся доказать, что

$$KP(y,z|x,KP(x)) \leqslant KP(y|x,KP(x)) + KP(z|x,KP(x)),$$

что есть релятивизованный вариант теоремы 54 (с. 95). $\rhd$

Приведём также прямое доказательство теоремы 64 с помощью полумер. Нам нужно доказать, что (с точностью до ограниченных множителей)

$$m(x,y,z)m(x) \geqslant m(x,y)m(x,z),$$

где m — наибольшая перечислимая снизу полумера. Поделив на $m(x)$, получим неравенство

$$\frac{m(x,y)m(x,z)}{m(x)} \leqslant m(x,y,z).$$

Проверим, что его левая часть имеет конечную сумму (по всем тройкам слов x, y, z). Это следует из того, что

$$\sum_{y,z} \frac{m(x,y)m(x,z)}{m(x)} \leqslant m(x)$$

(ведь $\sum_y m(x,y) \leqslant m(x)$ и $\sum_z m(x,z) \leqslant m(x)$). (Для краткости мы опускаем множители $O(1)$, как если бы они были равны единице.)

Этого, однако, недостаточно: из-за $m(x)$ в знаменателе дробь

$$\frac{m(x,y)m(x,z)}{m(x)}$$

может не быть перечислимой снизу, и потому мы не можем воспользоваться свойством максимальности. Применим такой обходной манёвр (уже использованный нами при доказательстве теоремы 61): построим перечислимую снизу верхнюю оценку для этой дроби.

Это делается так: для каждого натурального n через $m_n(x,y)$ обозначим функцию $m(x,y)$, у которой для каждого x сумма $\sum_y m(x,y)$ принудительно ограничена сверху числом 2^{-n}. Заметим, что $\sum_y m(x,y)$ равно $m(x)$ (как и раньше, ограниченные множители мы опускаем), поэтому $m_n(x,y) = m(x,y)$ при $n = \mathit{KP}(x)$. Теперь рассмотрим функцию

$$\langle x,y,z\rangle \mapsto \sum_{n\geqslant \mathit{KP}(x)} \frac{m_n(x,y)m_n(x,z)}{2^{-n}}$$

Среди всех слагаемых есть и член с $n = \mathit{KP}(x)$. С другой стороны,

$$\begin{aligned}\sum_{x,y,z}\sum_{n\geqslant \mathit{KP}(x)} \frac{m_n(x,y)m_n(x,z)}{2^{-n}} &\leqslant \sum_{x}\sum_{n\geqslant \mathit{KP}(x)} \frac{\sum_y m_n(x,y)\sum_z m_n(x,z)}{2^{-n}}\\ &\leqslant \sum_{x}\sum_{n\geqslant \mathit{KP}(x)} 2^{-n} \leqslant \sum_x 2m(x) \leqslant 2.\end{aligned}$$

(Как и раньше, мы опускаем константы — их учёт даст ограниченный множитель в окончательной оценке.)

84 [condit-triple-prefix] Покажите, что неравенство из теоремы 26 (с. 50) для префиксной сложности выполняется с точностью $O(1)$:

$$2\,\mathit{KP}(x,y,z) \leqslant \mathit{KP}(x,y) + \mathit{KP}(x,z) + \mathit{KP}(y,z) + O(1)$$

для любых слов x,y,z. [Указание: сложите базисное неравенство $\mathit{KP}(x,y,z) + \mathit{KP}(z) \leqslant \mathit{KP}(x,z) + \mathit{KP}(y,z)$ с неравенством $\mathit{KP}(x,y,z) \leqslant \mathit{KP}(x,y) + \mathit{KP}(z)$.]

85 [increasing-pair-complexity] Докажите, что при некотором c для любого слова x и любого числа n найдётся такое слово y длины n, что

$$\mathit{KP}(x,y) \geqslant \mathit{KP}(x) + n - c$$

[Указание: для любых z и n существует слово y длины n, для которого $\mathit{KP}(y|z) \geqslant n$.]

Сходное утверждение можно сформулировать не для пар, а для продолжений данного слова x на n битов (аналогичное утверждение для простой колмогоровской сложности составляло содержание задачи 34 на с. 45)

Theorem 65. [increasing-prefix-complexity]

$$\max\{\mathit{KP}(xy) | l(y) = n\} \geqslant \mathit{KP}(x|n) + n - O(1).$$

Другими словами, при некотором c для всех x и n можно так продлить слово x на n битов, чтобы его сложность увеличилась на $n-c$, правда не по сравнению с $\mathit{KP}(x)$, а лишь по сравнению с $\mathit{KP}(x|n)$.

$\lhd$ Переходя к априорным вероятностям, перепишем искомое неравенство как

$$2^n \min\{m(xy) | l(y) = n\} \leqslant m(x|n) \cdot O(1)$$

Левая часть не превосходит $\sum\{m(xy)|l(y) = n\}$ (если заменить все слагаемые на наименьшее из них, сумма лишь уменьшится). А эта величина представляет собой при каждом n (как функция от x) перечислимую снизу полумеру, так что остаётся воспользоваться максимальностью. $\rhd$

$\boxed{\mathbf{86}}$ Покажите, что чуть более слабое утверждение с $\mathit{KP}(x) - \mathit{KP}(n)$ вместо $\mathit{KP}(x|n)$ в правой части можно вывести из задачи 85.

5. Монотонная сложность

`[monot]`

5.1. Вероятностные машины и полумеры на дереве

`[monotsm]` Определяя априорную вероятность в главе 4, мы рассматривали вероятностные алгоритмы (машины), которые на выходе печатали натуральное число и останавливались. Теперь мы рассмотрим другой класс вероятностных алгоритмов, которые выдают на выход бит за битом и не обязаны останавливаться. Алгоритм такого вида задаёт случайную величину, значениями которой являются элементы Σ (конечные и бесконечные последовательности нулей и единиц).

В качестве примера рассмотрим алгоритм, выдающий на выходе случайные биты, получаемые от датчика:

while true do
  b:=*random*;
  OutputBit(*b*);
od

Соответствующая случайная величина сосредоточена на бесконечных последовательностях (любое множество конечных последовательностей имеет вероятность нуль) и имеет там равномерное распределение. Но в общем случае конечные последовательности могут появляться с положительной вероятностью (это соответствует ситуации, когда алгоритм с некоторого момента новых битов не выдаёт).

Для каждого алгоритма A описанного типа рассмотрим функцию, определённую на двоичных словах и принимающую вещественные значения:

$$a(x) = \Pr[\text{выход } A \text{ начинается на } x]$$

Формально говоря, эту функцию надо определять так. Каждому вероятностному алгоритму A соответствует некоторое отображение $\bar{A}$ множества Ω (бесконечные последовательности нулей и единиц) в множество Σ. Именно, $\bar{A}(\omega)$ есть последовательность битов, которая получается на выходе, если использовать в качестве случайных битов члены последовательности ω (очередной вызов *b* := *random* берёт следующий бит последовательности ω). Для приведённой выше программы, очевидно, $\bar{A}(\omega) = \omega$.

После этого $a(x)$ определяется как мера прообраза множества Σ_x при отображении $\bar{A}$ (где Σ_x есть множество всех конечных и бесконечных последовательностей, начинающихся на слово x).

$\boxed{87}$ Найдите $\bar{A}$ и a для алгоритма A, который печатает на выходе бесконечную последовательность нулей (и не использует случайных битов).

В этом разделе мы опишем функции, соответствующие вероятностным алгоритмам указанного вида.

Theorem 66. `[monotsm-crit1]` *Пусть A — вероятностный алгоритм описанного вида, a — соответствующая ему функция. Тогда:*

(а) $a(x) \geqslant 0$ при всех x;
(б) $a(\Lambda) = 1$ (здесь Λ обозначает пустое слово);
(в) $a(x) \geqslant a(x0) + a(x1)$ для любого слова x;
(г) функция a перечислима снизу.

Перечислимость снизу определялась в разделе 4.1 (с. 72) для последовательностей действительных чисел. Для функций на словах определение аналогично: требуется, чтобы $a(x) = \lim_i a(x, i)$, где a — вычислимая функция двух аргументов, $a(x, i)$ определено всех слов x и для всех натуральных i, является рациональным числом (или специальным символом $-\infty$) и не убывает по i.

$\triangleleft$ Первые три утверждения очевидны:

(а) Вероятность всегда неотрицательна.

(б) $a(\Lambda) = 1$, поскольку пустое слово является началом любого выхода.

(в) $a(x) \geqslant a(x0) + a(x1)$, поскольку события «на выходе появилось $x0$» и «на выходе появилось $x1$» несовместны и являются подмножествами события «на выходе появилось x».

Заметим, что неравенство пункта (в) не обязано быть равенством: разность

$$a(x) - a(x0) - a(x1)$$

есть вероятность того, что на выходе появится слово x и больше никаких битов не появится.

(г) Чтобы доказать перечислимость снизу функции a, нам надо уметь получать приближения снизу к $a(x)$ для данного слова x. Будем моделировать поведение алгоритма A для всех возможных значений случайных битов. Время от времени будут обнаруживаться значения случайных битов, которые гарантируют появление на выходе слова x, то есть интервалы в пространстве Ω, на которых функция $\bar{A}$ равна x или какому-нибудь продолжению слова x. Вероятность $a(x)$ равна суммарной мере всех таких интервалов, и в качестве искомого приближения $a(x, i)$ мы берём меру интервалов, обнаруженных к шагу i. $\triangleright$

Функции a, определённые на двоичных словах, принимающие действительные значения и удовлетворяющие условиям (а) – (г) теоремы 66, называют *перечислимыми снизу полумерами на дереве двоичных слов*. Важно не смешивать их с ранее рассматривавшимися полумерами на натуральных числах (или на словах как изолированных конструктивных объектах): прежние полумеры соответствовали алгоритмам, которые выдавали на выходе какое-либо число (или слово) и останавливались.

Мы будем называть *полумерами на дереве* функции, удовлетворяющие условиям (а) – (в); условие (г) выделяет из них перечислимые снизу.

$\boxed{88}$ Покажите, что полумеры на дереве (функции со свойствами (а) – (в)) соответствуют мерам (в смысле теории меры) на пространстве Σ конечных и бесконечных последовательностей. Найдите меру множества всех бесконечных продолжений слова x при таком соответствии. [Ответ: если a — полумера на дереве, то мера этого множества равна пределу убывающей по n последовательности

$$\alpha_n = \sum \{a(y) | y \text{ есть продолжение } x \text{ длины } n\}.$$

Здесь α_n определено при $n \geqslant l(x)$ и равно $a(x)$ при $n = l(x)$.]

$\boxed{89}$ Покажите, что для перечислимой снизу полумеры на дереве сумма $\sum_x a(x)$ может быть бесконечной. [Указание. Рассмотрите алгоритм, копирующий случайные биты на выход.]

Верно и обратное к теореме 66 утверждение:

Theorem 67. [monotsm-crit2] *Всякая перечислимая снизу полумера на дереве соответствует некоторому вероятностному алгоритму.*

$\triangleleft$ Идею доказательства легко объяснить в терминах выделения места, по аналогии с доказательством теоремы 40 (с. 75). Изменение состоит в том, что теперь запросы носят иерархический характер. Две организации (которые мы будем условно называть 0 и 1) требуют выделить им непересекающиеся подмножества в пространстве Ω (которое можно отождествить с отрезком $[0,1]$), при этом размеры запрашиваемых ими областей увеличиваются со временем, но в сумме ни в какой момент не превышают единицы.

В каждой из организаций есть по два подразделения (в организации 0 есть подразделения 00 и 01, в организации 1 есть два подразделения 10 и 11), которые требуют выделения им места внутри области, выделенной для всей организации. При этом их требования растут со временем, но в сумме не превышают требований всей организации. Аналогично для ещё более мелких подразделений 000, 001 и так далее: в пределе каждое подразделение x требует (не обязательно сплошного) участка, имеющего размер $a(x)$, где a — заданная полумера. При этом однажды выделенная любому подразделению область навсегда остаётся за ним и перераспределена быть не может.

Соответствие между этой схемой выделения места и вероятностным алгоритмом таково: если последовательность случайных битов попадает в область, выделенную подразделению x, это означает, что выход алгоритма начинается с x (для данной последовательности случайных битов).

Представив себе описанную ситуацию, легко понять, что такое распределение осуществимо. Тем не менее мы приведём более формальное рассуждение (одновременно объясняя смысл его этапов с точки зрения описанной метафоры).

Лемма 1. Пусть a — произвольная перечислимая снизу полумера на дереве. Тогда существует вычислимая всюду определённая неубывающая по i функция $\langle x,i\rangle \mapsto a(x,i)$, значения которой есть неотрицательные двоично-рациональные числа, $\lim_i a(x,i)=a(x)$ и для каждого i функция $x\mapsto a(x,i)$ является полумерой, у которой лишь конечное число значений отлично от нуля.

Другими словами, выделяющий место (кладовщик, менеджер памяти) может наложить дополнительные ограничения:

- запрашиваемые размеры должны быть двоично-рациональными (иметь вид $k/2^n$ для некоторых целых k и n);

- на каждом шаге лишь конечное число подразделений может подать ненулевые запросы;

- запросы должны быть согласованы (запрос любого подразделения должен быть не меньше суммы запросов его частей).

Доказательство леммы. Будем постепенно корректировать функцию a из определения перечислимой снизу полумеры (не меняя само́й предельной полумеры). Сначала добьёмся, чтобы все значения были двоично-рациональными. Это можно сделать, заменив $a(x,i)$ на ближайшее снизу двоично-рациональное число со знаменателем 2^i (отрицательные числа заменяем на нули).

Затем можно добиться выполнения второго условия, положив $a(x,i) = 0$ для слов x длины больше i.

Наконец, можно добиться выполнения третьего условия, выполнив замену

$$a(x,i) := \max\left(a(x,i),\, a(x0,i) + a(x1,i)\right)$$

последовательно в порядке убывания длин слов. Поскольку функция $a(x)$ является полумерой по предположению, такие замены сохранят неравенство $a(x,i) \leqslant a(x)$. Легко проверить, что значения полумеры (пределы $\lim_i a(x,i)$ при каждом x) при нашей коррекции не изменятся.

Лемма 1 доказана.

Следующую лемму удобно формулировать, введя несколько вспомогательных определений. Будем называть полумеру с конечным числом ненулевых двоично-рациональных значений «простой» полумерой.

Будем называть «простым множеством» объединение конечного числа интервалов в Ω. (Напомним, что интервал в Ω есть множество вида Ω_z, состоящее из всех бесконечных продолжений данного слова z. Таким образом, простыми являются множества, принадлежность к которым определяется конечным числом битов последовательности.)

Будем называть «простым семейством» семейство простых множеств A_x, индексированное двоичными словами x, в котором все множества, кроме конечного числа, пусты, и для каждого слова x множества A_{x0} и A_{x1} являются непересекающимися подмножествами множества A_x.

Для такого семейства функция $x \mapsto \mu(A_x)$, где μ — равномерная мера на Ω, является простой полумерой.

Лемма 2. Для всякой простой полумеры существует простое семейство, которому она соответствует.

Доказательство. Начнём строить такое семейство, начиная с пустого слова и постепенно увеличивая длину слова-индекса. На каждом шаге нам нужно будет внутри простого множества A_x найти два непересекающихся простых подмножества A_{x0} и A_{x1} заданных мер (при этом сумма этих мер не превосходит меры объемлющего множества). Ясно, что это возможно. Лемма 2 доказана.

Лемма 3. Пусть дана простая полумера $b(x)$ и соответствующее ей простое семейство B_x. Пусть дана также простая полумера c, причём $c(x) \geqslant b(x)$ для всех x. Тогда можно построить простое семейство C_x, соответствующее полумере c, для которого $C_x \supset B_x$ при всех x.

Доказательство. Повторим рассуждение из доказательства леммы 2, только теперь внутри простого множества есть два непересекающихся простых подмножества, и надо увеличить их меры до заданных значений, оставив подмножества непересекающимися и не выйдя за пределы множества. Ясно, что это возможно. Лемма 3 доказана.

Доказательства лемм 2 и 3 эффективны в том смысле, что по таблицам значений простых полумер можно алгоритмически построить соответствующие простые семейства.

Будем теперь применять лемму 3 по очереди к простым полумерам, получающимся фиксацией $i = 0, 1, 2, \ldots$ в $a(x, i)$. Получим двупараметрическое семейство простых множеств $U(x, i)$, при этом

- описание множества $U(x, i)$ (список входящих в него интервалов) строится алгоритмически по x и i;
- мера множества $U(x, i)$ равна $a(x, i)$ (и стремится к $a(x)$ при $i \to \infty$);
- при любых x и i множества $U(x0, i)$ и $U(x1, i)$ являются непересекающимися подмножествами множества $U(x, i)$;
- $U(x, i) \subset U(x, i + 1)$ при любых x и i.

Теперь искомый вероятностный алгоритм, порождающий полумеру a, можно построить так: строим множества $U(x, i)$ при всех x и i и параллельно запрашиваем случайные биты, записывая их в последовательность ω. Как только обнаруживается, что $\omega \in U(x, i)$ для каких-то x и i, выдаём на выход ещё не выданные биты слова x.

Заметим, что если ω оказалась в $U(x, i)$, то она автоматически оказывается в $U(y, i)$ для любого начала y слова x, а также что ω не может оказаться одновременно в $U(x, i)$ и $U(x', i)$ для несовместных слов x и x' (слов, не являющихся началами друг друга). Поэтому уже выданные на выход биты не потребуется «отзывать».

Слово x или его продолжение появится на выходе такого алгоритма тогда и только тогда, когда последовательность ω принадлежит объединению возрастающей последовательности множеств $U(x, i)$ при $i = 0, 1, 2, \ldots$; вероятность этого события есть предел мер множеств $U(x, i)$, а этот предел совпадает с $a(x)$, что и требовалось. $\triangleright$

Теоремы 66 и 67 показывают, что перечислимые снизу полумеры на дереве можно эквивалентно определить как функции, соответствующие распределениям на выходе вероятностных алгоритмов.

Среди всех вероятностных алгоритмов описанного вида выделяются алгоритмы, которые почти наверно выдают бесконечную последовательность (вероятность получить конечную последовательность равна нулю). Им соответствуют полумеры, являющиеся вычислимыми мерами. Точнее говоря, верна следующая теорема:

Theorem 68. [monotsm-semi-measure] (**а**) *Пусть μ — вычислимая мера на пространстве Ω. Тогда функция p, определяемая формулой $p(x) = \mu(\Omega_x)$, является перечислимой снизу полумерой, причём $p(x) = p(x0) + p(x1)$ для всех x.*

(**б**) *Если для перечислимой снизу полумеры p выполнено равенство $p(x) = p(x0)+p(x1)$ при всех x, то она определяет некоторую вычислимую меру на Ω.*

$\triangleleft$ (а) Если действительное число α вычислимо, и a_n — рациональное приближение к нему с ошибкой $1/n$, то $b_n = a_n - 1/n$ будет приближением снизу с ошибкой $2/n$. Вообще говоря, последовательность b_n может не быть монотонной (неубывающей), но последовательность

$$c_n = \max(b_0, b_1, \ldots, b_n)$$

заведомо будет вычислимой неубывающей последовательностью рациональных чисел, стремящейся к α. Таким образом, каждое вычислимое действительное число перечислимо снизу. Проводя эту конструкцию параллельно для всех x, мы убеждаемся, что любая вычислимая мера задаёт перечислимую снизу полумеру. Поскольку Ω_x есть объединение непересекающихся подмножеств Ω_{x0} и Ω_{x1}, то $p(x) = p(x0) + p(x1)$ (аддитивность меры).

(б) Пусть p — перечислимая снизу полумера, для которой $p(x) = p(x0) + p(x1)$ при всех x. Покажем индукцией по длине x, как найти приближение к $p(x)$ с любой заданной точностью. Для пустого слова значение $p(\Lambda)$ равно 1 по определению. Если мы уже умеем находить приближения к $p(x)$ сверху и снизу с любой точностью, а хотим сделать это для $p(x0)$ и $p(x1)$, то надо ждать, пока сумма постепенно растущих нижних оценок для $p(x0)$ и $p(x1)$ не приблизится к (убывающей) верхней оценке для $p(x)$. Другими словами, верхние оценки для $p(x1)$ можно получать, вычитая из верхних оценок для $p(x)$ (которые возможны по предположению индукции) нижние оценки для $p(x0)$. $\triangleright$

Эту теорему можно интерпретировать следующим образом. Пусть нам нужен датчик случайных чисел, который выдаёт последовательность нулей и единиц, распределённую по некоторой вычислимой мере p (вероятность того, что выдаваемая датчиком последовательность начинается на x, должна быть равна $p(x)$). Тогда теоремы 67 и 68 говорят, что такой датчик можно построить в виде вероятностного алгоритма, который внутри себя имеет равномерный датчик, а на выходе имеет нужное распределение.

Заметим, что в частном случае вычислимых мер можно воспользоваться более простой и явной конструкцией, чем дана в доказательстве теоремы 67. А именно, разобьём отрезок $[0,1]$ на две части длиной $p(0)$ и $p(1)$. Первую из них разобьём ещё на две части длиной $p(00)$ и $p(01)$, вторую — на части длиной $p(10)$ и $p(11)$ и так далее. Для каждого двоичного слова z получится некоторый отрезок π_z длиной $p(z)$; отрезки π_z для всех слов z данной длины образуют разбиение отрезка $[0,1]$.

Теперь рассмотрим вероятностный алгоритм, который бросает честную монету и получает последовательность α случайных битов. Эта последовательность рассматривается как двоичное разложение некоторого числа (которое мы тоже будем обозначать α). Параллельно с получением α алгоритм ищет такие z, при которых α содержится строго внутри π_z (и это можно сказать на основе уже имеющейся информации о α и о границах отрезков π_z, которые параллельно вычисляются с возрастающей точностью).

Возникающие при этом слова z являются началами друг друга (чем больше битов числа α известно, тем длиннее может быть z). Они являются началами последовательности битов, которая выдаётся на выход.

Вообще говоря, алгоритм этот может дать конечную последовательность, если число α совпадёт с концом одного из отрезков π_z, но таких концов счётное множество, и вероятность этого события равна 0. Выход алгоритма начинается на z тогда и только тогда, когда число α содержится строго внутри π_z, так что вероятности правильные.

Говоря более формально, описанное преобразование T последовательности битов α в выходную последовательность битов $\beta = T(\alpha)$ переводит равномерную меру в меру p.

(Ничего неожиданного в этом алгоритме нет. Если у вас есть датчик случайных равномерно распределённых на $[0,1]$ чисел, а нужно имитировать несимметричную монету, где 0 и 1 имеют вероятности, скажем, $2/3$ и $1/3$, то надо сравнить случайную точку с границей $2/3$. Для получения второго бита с тем же распределением надо каждый из отрезков $[0,2/3]$ и $[2/3,1]$ снова разделить на неравные части в пропорции $2:1$. Ровно это и делается в описанном алгоритме.)

Чтобы понять связи между классами случайных последовательностей (в смысле Мартин-Лёфа) по разным мерам, полезно изучить преобразование T более внимательно. Его удобно описать так: наряду с семейством отрезков π_z рассмотрим аналогичное семейство I_z для равномерной меры. Другими словами, I_z — отрезок, образованный действительными числами, двоичные разложения которых начинаются на z (плюс концы).

Отображение $T\colon \Omega \to \Sigma$ теперь можно описать так: слово y является началом $T(\alpha)$, если существует слово x, являющееся началом α, для которого отрезок I_x лежит во внутренности отрезка π_y.

Аналогично можно определить и отображение $U\colon \Omega \to \Sigma$, поменяв семейства отрезков местами: слово x является началом $U(\beta)$, если β имеет начало y, для которого π_y лежит строго внутри I_x.

Хочется сказать, что преобразования T и U взаимно обратны и осуществляют переход от двоичного разложения действительного числа на отрезке $[0, 1]$ к его «разложению по мере p», в котором числа из π_z имеют разложения, начинающиеся на z, и наоборот. Однако при буквальном понимании это неверно. Во-первых, двоично-рациональные числа создают неоднозначность при двоичном разложении. С другой стороны, и вторая половина соответствия, переход от действительного числа к его разложению по мере p, также не является однозначной, поскольку концы отрезков π_z соответствуют нескольким последовательностям. Кроме того, может случиться, что какой-то из отрезков π_z имеет нулевую длину (тогда целый конус последовательностей соответствует одному числу). Возможно также, что некоторая конкретная последовательности имеет положительную меру (тогда целый интервал действительных чисел соответствует этой последовательности).

Но в остальном, по модулю этих проблем, действительно имеет место соответствие, переводящее равномерную меру в меру p. Это неформальное заявление можно уточнять различными способами. Мы уже видели, что образ равномерно распределённой случайной последовательности при отображении T имеет распределение p. В другую сторону:

$\boxed{\mathbf{90}}$ Докажите, что если никакая последовательность не имеет положительной p-меры, а случайная величина β имеет распределение p, то последовательность $U(\beta)$ бесконечна с вероятностью 1 и имеет равномерное распределение на Ω. [Указание. Всякой последовательности соответствуют вложенные отрезки семейства π; их общая точка единственна; вероятность того, что она окажется двоично-рациональной, равна нулю. Вероятность для $U(\beta)$ попасть в отрезок π_z будет правильной, а любой другой отрезок можно приблизить отрезками вида π_z.]

Эта задача показывает, как получить равномерно распределённые случайные биты, имея датчик с известным распределением p. Модельный пример: у нас есть несимметричная монета с вероятностью орла в $2/3$, а мы хотим бросить жребий честно (чтобы оба игрока имели равные шансы). В этом конкретном случае это несложно безо всякой теории: дадим каждому бросить монету по разу, если оба раза орёл или решка, то ничья (и бросания повторяются), если орёл только у одного, то этот один и выиграл. Этот приём даже не зависит от вероятности орла (в отличие от предыдущей конструкции).

Условие отсутствия последовательностей положительной меры существенно: не из всякого распределения можно получить равномерное. Например, если $p(000\ldots 000) = 1$ для слова из любого числа нулей, а на остальных словах p равно нулю, то такой датчик случайных битов никакой случайности в себе не содержит, он просто выдаёт нули, и с его помощью ничего не сделаешь. Примерно такой же будет ситуация, если некоторая бесконечная последовательность имеет положительную меру (если есть бесконечная последовательность ω и число $\delta > 0$, для которых $p(x) \geqslant \delta$ для любого начала x последовательности ω). В этом случае с вероятностью не менее δ датчик даст последовательность ω, и потому появиться на выходе равномерное распределение не может.

Нас, однако, больше интересует связь между классами случайных (в смысле Мартин-Лёфа) последовательностей относительно различных мер.

Theorem 69. (а) *Если последовательность α случайна (в смысле Мартин-Лёфа) по равномерной мере, то последовательность $\beta = T(\alpha)$ бесконечна и случайна (в смысле Мартин-Лёфа) по мере p.*

(б) *Если последовательность β случайна по мере p и невычислима, то последовательность $\alpha = U(\beta)$ бесконечна, случайна по равномерной мере, и $T(\alpha) = \beta$.*

(Случайность в этой теореме понимается в смысле Мартин-Лёфа. Отметим также, что часть утверждений этой задачи следует из теоремы 99, с. 154.)

$\triangleleft$ (а) Случайная по равномерной мере последовательность не может быть вычислимой, поэтому она не может задавать двоично-рациональное число (то есть содержать конечное число нулей или конечное число единиц) и даже вычислимое действительное число (а концы отрезков π_z вычислимы). Значит, последовательность $\beta = T(\alpha)$ бесконечна.

Если имеется алгоритм, который по заданному $\varepsilon > 0$ покрывает последовательность β интервалами Ω_u с суммой p-мер меньше ε, то можно перейти к отрезкам π_u, заменить их на чуть большие интервалы и получить алгоритм, покрывающий действительное число, имеющее двоичную запись α, интервалами с малой суммой длин. Отсюда легко получаем покрытие последовательности α интервалами Ω_v с малой суммой равномерных мер (интервал на прямой заменяем объединением непересекающихся интервалов Ω_v, при этом сумма мер сохраняется). А это противоречит предположению о случайности последовательности α.

(б) В обратную сторону требуются некоторые дополнительные предосторожности. Прежде всего заметим, что если одноэлементное множество $\{\beta\}$ имеет положительную p-меру, то последовательность β вычислима. (Напомним, что мера p вычислима.) В самом деле, пусть мера $\{\beta\}$ больше рационального числа ε. Последовательностей β с таким свойством конечное число (не больше $1/\varepsilon$). Пусть их, скажем, k. Увеличим ε до некоторого ε' так, чтобы все k последовательностей по-прежнему имели меру больше ε'. Тогда для достаточно бóльших n (скажем, начиная с некоторого N) имеется ровно k слов z длины n, для которых $p(z) > \varepsilon'$. Зная ε' и N, мы можем эффективно найти эти слова, и потому образуемые ими бесконечные ветви вычислимы.

Таким образом, если β невычислима, то длины отрезков π_z для слов z, являющихся началами β, стремятся к нулю. Таким образом, соответствующие отрезки на прямой имеют единственную точку пересечения. Эта точка является внутренней для всех отрезков, поскольку β содержит бесконечное число нулей и единиц (невычислимость). Далее замечаем, что эта точка не является двоично-рациональной (иначе β была бы опять вычислимой), и потому последовательность $\alpha = U(\beta)$ бесконечна и $\beta = T(\alpha)$.

Осталось доказать, что последовательность α случайна. В самом деле, если нам дают её покрытие интервалами малой равномерной меры, то мы можем перенести эти интервалы на прямую и немного их расширить, а затем взять отрезки π_z, попавшие в эти интервалы.

Среди соответствующих слов z обязательно будет некоторое начало последовательности β, так как длины вложенных отрезков стремятся к нулю, а интервалы открыты и содержат α с некоторой окрестностью. $\triangleright$

Отсюда непосредственно вытекает такое утверждение: если последовательность ω случайна по некоторой вычислимой мере, то она либо вычислима, либо эквивалентна по Тьюрингу некоторой последовательности, случайной по равномерной мере.

Эквивалентность по Тьюрингу двух последовательностей $\alpha, \beta \in \Omega$ означает, что α сводится по Тьюрингу к β и наоборот. *Сводимость* α к β означает, что существует программа, которая вычисляет α, вызывая внешнюю процедуру («оракул») для получения битов β.) В нашем случае преобразования T и U обеспечивают сводимость в обе стороны.

Последовательности, случайные по некоторой вычислимой мере, были названы «правильными» в статье [85] (в английском переводе этой статьи использован термин “proper”).

В связи с этой задачей возникает такой вопрос: а существует ли вообще последовательность, не случайная ни по какой вычислимой мере? или даже не эквивалентная по Тьюрингу никакой случайной по равномерной мере? (В последнем вопросе можно заменить «равномерной» на «вычислимой»; как мы видели, это всё равно.)

По этому поводу можно сказать следующее.

1. Несложно построить последовательность, не случайную ни по какой вычислимой мере. При этом удобно использовать понятие критерий случайности в терминах дефектов (раздел 5.9, с. 152):

$\boxed{\textbf{91}}$ Покажите, что для любого конечного слова x и любой вычислимой меры P существует продолжение y слова x, имеющее сколь угодно большой дефект случайности (в смысле раздела 5.9) относительно P. [Указание. Будем продолжать x так, чтобы добавление очередного бита уменьшало меру P не менее чем (скажем) в полтора раза. Это можно сделать вычислимо, поэтому сложность будет расти медленно, а мера убывать быстро.] Рассматривая поочерёдно все вычислимые меры (каждую бесконечное число раз), покажите, что существует бесконечная последовательность, не случайная ни по какой вычислимой мере.

По существу то же самое рассуждение можно изложить с помощью так называемых «генерических» последовательностей. Напомним, что подмножество A пространства Ω называется *всюду плотным*, если оно пересекается с любым интервалом. Знаменитая *теорема Бэра* утверждает, что пересечение счётного семейства всюду плотных открытых множеств (открытые множества — объединения интервалов) непусто (и даже всюду плотно).

$\boxed{\textbf{92}}$ Докажите это утверждение, взяв произвольное конечное слово и прибавляя к нему фрагменты так, чтобы попасть внутрь очередного открытого всюду плотного множества.

Рассмотрим теперь эффективно открытые множества (объединения перечислимых семейств интервалов) и отберём из них всюду плотные. Получится счётное семейство всюду плотных открытых множеств, которое по теореме Бэра имеет непустое (и даже всюду плотное) пересечение. Назовём эти последовательности *генерическими*.

Генерические последовательности можно неформально описать как «не подчиняющиеся никаким законам», если под законом понимать алгоритмически проверяемое утверждение, запрещающее хотя бы одно конечное продолжение у любой конечной последовательности.

$\boxed{\textbf{93}}$ Докажите, что никакая генерическая последовательность не удовлетворяет усиленному закону больших чисел. [Указание. Множество двоичных слов длины более N, в которых доля единиц больше 99%, является эффективно открытым и всюду плотным. Аналогично для слов с долей единиц менее 1%.]

$\boxed{94}$ Докажите, что генерическая последовательность не может быть вычислимой. [Указание. Множество всех последовательностей, отличных от данной вычислимой, эффективно открыто и всюду плотно.]

В отличие от случайности, определение генерической последовательности не предполагает никакой меры.

$\boxed{95}$ Докажите, что генерическая последовательность не может быть случайной ни по какой вычислимой мере. [Указание: достаточно построить эффективно открытое плотное множество сколь угодно малой меры. Для этого можно из двух половинок каждого интервала выбирать меньшую (или чуто-чуть большую) по мере.]

Как написано в статье [85] (замечание после определения 4.4), несложно показать, что характеристическая последовательность универсального перечислимого множества не является случайной ни по какой вычислимой мере, но там не указано, в каком смысле понимается универсальность. Скорее всего утверждение, которое авторы этой статьи имели в виду, вытекает из следующего результата:

$\boxed{96}$ Покажите, что существует перечислимое множество, характеристическая последовательность которого не случайна ни по какой вычислимой мере.

[Указание. Поскольку сложности начальных отрезков будут логарифмическими для любого перечислимого множества, достаточно гарантировать, что их меры быстро убывают. Разделим натуральный ряд на счётное число арифметических прогрессий; i-я из них будет отвечать за i-ю вычислимую меру. Поскольку мы на самом деле не знаем, имеем ли мы дело с мерой или нет, получается перечислимое (а не разрешимое) множество.]

2. Сложнее — но тоже возможно — построить последовательность, которая не эквивалентна по Тьюрингу никакой последовательности, случайной по равномерной мере. При этом заранее известно, в какую сторону будет проблема: к этой последовательности не будет сводиться никакая случайная. Более того, можно построить вероятностную машину, которая выдаёт такие последовательности с положительной вероятностью. (Эта конструкция приведена в статье Вьюгина [80].)

3. Как мы увидим дальше, в обратную сторону сводимость есть всегда (всякая последовательность сводится по Тьюрингу к `[gacs-reducibility-remark]`случайной по равномерной мере, теорема 101, с. 159).

4. Можно также построить последовательность, эквивалентную случайной по равномерной мере, но не случайную ни по какой вычислимой мере. Для этого на чётных местах поместим генерическую последовательность, а на нечётных — ту случайную по равномерной мере, к которой она сводится.

(Здесь надо использовать результат предыдущего пункта, а также тот факт, что если последовательность случайна в смысле Мартин-Лёфа по мере P, то подпоследовательность её членов с чётными номерами случайна в смысле Мартин-Лёфа относительно проекции P на эти координаты.)

Последовательности, не случайные ни по какой вычислимой мере, в некотором смысле аналогичны нестохастическим по Колмогорову конечным объектам (см. раздел 16.2). Более того, несложно показать, что если последовательность случайна по некоторой вычислимой мере, то её начальные отрезки являются стохастическими (задача 268, с. 388).

В заключение раздела укажем следующее простое следствие доказанных нами теорем:

Theorem 70. [monotsm-anybits] *Определение перечислимой полумеры не изменится, если разрешить использовать вероятностные алгоритмы с произвольным вычислимым распределением вероятности на случайных битах (вместо равномерного): всё равно полученная полумера будет перечислима снизу.*

$\lhd$ В самом деле, любое вычислимое распределение вероятностей может быть получено на выходе вероятностного алгоритма, и тем самым можно имитировать любой датчик с вычислимым распределением, использовав комбинацию двух вероятностных алгоритмов.

Другой способ доказательства состоит в повторении доказательства теоремы 66, где обнаруживающиеся интервалы будут иметь вычислимые меры и потому вероятность можно приближать снизу. $\rhd$

5.2. Наибольшая перечислимая полумера на дереве

[monotmax]

Theorem 71. *Среди всех перечислимых снизу полумер на дереве существует наибольшая (с точностью до константы) полумера a: для любой перечислимой снизу полумеры a' на дереве выполнено неравенство $a'(x) \leqslant ca(x)$ для некоторой константы c и для всех x.*

$\lhd$ Как и для полумер на $\mathbb{N}$ (теорема 41, с. 76), рассмотрим вероятностный алгоритм A, который сначала случайно выбирает некоторый вероятностный алгоритм, а затем моделирует выбранный алгоритм. Если полумера a' соответствует вероятностному алгоритму A', то $a'(x) \leqslant (1/\varepsilon)a(x)$, где ε — вероятность того, что будет выбран именно алгоритм A'. $\rhd$

Другой способ доказательства состоит в том, что мы располагаем все перечислимые снизу полумеры в вычислимую последовательность $a_0, a_1, \ldots$, а затем рассматриваем полумеру $a = \sum_i \lambda_i a_i$, где λ_i — ряд с вычислимыми членами и суммой 1 (например, $\lambda_i = 2^{-i-1}$).

Тонкость здесь в том, что надо построить вычислимую последовательность всех полумер, то есть перечислимую снизу функцию $u(i,x)$, для которой: (1) при каждом фиксированном i функция $u_i\colon x \mapsto u(i,x)$ является полумерой; (2) среди u_i содержатся все перечислимые снизу полумеры.

Это можно сделать, либо перечисляя все вероятностные алгоритмы (что соответствует ранее приведённому доказательству), либо принудительно корректируя программы так, чтобы они перечисляли полумеры. Это рассуждение аналогично случаю полумер на $\mathbb{N}$ (раздел 4.2, страница 76), и мы не будем приводить детали. (Скажем лишь, что в случае нарушения условия $p(x) \geqslant p(x0) + p(x1)$ надо увеличивать $p(x)$, если только это не приведёт к тому, что $p(\Lambda)$ станет больше единицы.)

$\boxed{97}$ Проведите это рассуждение подробно.

(Замечание в скобках: доказательство даёт даже немного больше, чем утверждается. А именно, не только вероятность появления слова *с началом* x (то есть $p(x)$), но и вероятность появления на выходе *в точности слова* x, то есть разность $p(x) - p(x0) - p(x1)$, для построенной «универсальной» машины больше, чем для любой другой.)

$\boxed{98}$ Покажите, что всё сказанное выше естественно переносится на вероятностные алгоритмы, которые выдают на выход (по одному) не биты, а натуральные числа (каждое число выдаётся как единое целое). Такие алгоритмы соответствуют перечислимым снизу мерам на пространстве конечных и бесконечных последовательностей натуральных чисел.

$\boxed{99}$ (Продолжение.) Пусть m — наибольшая перечислимая снизу полумера на пространстве конечных и бесконечных последовательностей натуральных чисел. Покажите, что её ограничение на последовательности длины 1 даёт априорную вероятность на натуральных числах в смысле главы 4, а её ограничение на последовательности нулей и единиц даёт только что рассмотренную нами наибольшую меру на двоичном дереве.

Фиксируем некоторую наибольшую перечислимую снизу полумеру на дереве и обозначим её $a(x)$. Её также называют *универсальной полумерой* на дереве. Величину $a(x)$ можно было бы назвать *априорной вероятностью слова* x (как элемента дерева), но важно не путать её с априорной вероятностью в смысле главы 4. Величину

$$KA\,(x) = -\log a(x)$$

будем называть *априорной сложностью* слова x. (Здесь можно не опасаться путаницы, так как в главе 4 аналогичное определение приводило к префиксной сложности и специальное наименование излишне.) Различные варианты выбора наибольшей полумеры приводят к разным функциям априорной сложности, но они отличаются друг от друга не более чем на аддитивную константу (поскольку различные наибольшие полумеры на дереве отличаются не более чем на мультипликативную константу).

В следующем разделе мы изучим свойства априорной сложности. Заметим сразу же, что формально по нашему определению априорная сложность может не быть целым (и даже рациональным) числом. Но поскольку большинство интересующих нас утверждений всё равно справедливы «с точностью до $O(1)$», то можно было бы заменить $-\log a(x)$ на минимальное число n, при котором $a(x) > 2^{-n}$. Тут есть небольшая тонкость: мы используем строгое неравенство, чтобы получаемая функция была перечислима сверху. В дальнейшем мы специально оговариваем те редкие случаи, когда это округление (или его отсутствие) существенно.

5.3. Свойства априорной сложности

[monotapr]

Theorem 72. [ka-properties]

(а) *$KA\,(x) \leqslant l(x) + O(1)$ для любого x;*

(б) *$KA\,(x) \leqslant KP(x) + O(1)$ для любого x;*

(в) *если $x_0, x_1, \ldots$ — вычислимая последовательность несравнимых слов (ни одно из них не является началом другого), то $KA\,(x_i) = KP(x_i) + O(1) = KP(i) + O(1)$;*

(г) *$KP(x) \leqslant KA\,(x) + 2\log l(x) + O(1)$;*

(д) *более того, $KP(x) \leqslant KA\,(x) + KP(l(x)) + O(1)$;*

(е) *и более того, $KP(x|l(x)) \leqslant KA\,(x) + O(1)$;*

(ё) *бесконечная последовательность нулей и единиц вычислима тогда и только тогда, когда априорная сложность её начальных отрезков ограничена.*

(ж) *если* $f\colon \Sigma \to \mathbb{N}_\perp$ — *вычислимое непрерывное отображение, то* $\mathit{KP}(f(x)) \leqslant \mathit{KA}(x) + O(1)$ *для любого слова* x, *на котором* $f(x)$ *определено* (*не равно* $\perp$).

$\lhd$ (а) Функция $p(x) = 2^{-l(x)}$ является перечислимой снизу полумерой, и потому $p(x) \leqslant ca(x)$ для некоторого c, откуда и следует требуемое.

(б) Машины, которые выдают двоичное слово (как единое целое) и останавливаются, можно рассматривать как частный случай машин, выдающих последовательности битов. Поэтому $m(x) \leqslant ca(x)$, где m — априорная вероятность в смысле главы 4, откуда и следует требуемое.

Поучительно провести это рассуждение в терминах полумер. Положим $m'(x)$ равным сумме $m(y)$ по всем продолжениям y слова x, где m — априорная вероятность в смысле предыдущей главы, на словах как на изолированных объектах. (Ещё положим в порядке исключения $m'(\Lambda) = 1$.) Тогда m' будет полумерой на дереве и потому $m(x) \leqslant m'(x) = O(a(x))$.

(в) Если x_i — вычислимая последовательность несравнимых двоичных слов, а a — априорная вероятность на дереве (универсальная полумера на дереве), то функция $i \mapsto a(x_i)$ является полумерой на $\mathbb{N}$. В самом деле, она перечислима снизу, а события «x_i появляется на выходе вероятностного алгоритма» несовместны, и потому сумма их вероятностей не превосходит единицы. Поэтому $\mathit{KP}(i) \leqslant \mathit{KA}(x_i) + O(1)$.

С другой стороны, $\mathit{KP}(x_i) = \mathit{KP}(i) + O(1)$, поскольку из i можно алгоритмически получить x_i и наоборот; наконец, $\mathit{KA}(x_i) \leqslant \mathit{KP}(x_i) + O(1)$ согласно (б).

(г) Пусть a — априорная вероятность на дереве. Рассмотрим перечислимую снизу функцию $u(x) = a(x)/l(x)^2$. Сумма значений $a(x)$ по всем словам длины n не превосходит 1, поскольку эти слова попарно несравнимы, поэтому

$$\sum_x u(x) = \sum_n \sum_{l(x)=n} \frac{a(x)}{n^2} \leqslant \sum_n \frac{1}{n^2} = O(1),$$

откуда и следует требуемое.

(д) доказывается аналогично, только надо положить $u(x) = a(x)m(l(x))$, где m — априорная вероятность на $\mathbb{N}$.

(е) Рассмотрим функцию

$$u(x,n) = \begin{cases} a(x), & \text{если } l(x) = n,\\ 0, & \text{если } l(x) \neq n. \end{cases}$$

Тогда при каждом n функция $x \mapsto u(x,n)$ является полумерой в смысле предыдущей главы (сумма не превосходит 1), откуда и следует требуемое.

(ё) Для данной бесконечной вычислимой последовательности ω нулей и единиц рассмотрим вероятностный (по форме) алгоритм, который не использует датчика случайных чисел и печатает один за другим биты последовательности ω. Соответствующая полумера равна 1 на любом начале последовательности, и потому априорные вероятности (мы рассматриваем априорную вероятность на дереве) всех её начал отделены от нуля.

Обратное рассуждение несколько сложнее. Пусть априорные вероятности всех начал последовательности ω больше некоторого рационального $\varepsilon > 0$. Рассмотрим множество B

всех двоичных слов, для которых $a(x) > \varepsilon$. Множество B содержит все начала последовательности ω. Оно является деревом (вместе с любым словом содержит все его начала). Кроме того, любое его подмножество, состоящее из попарно несравнимых слов, содержит не более $1/\varepsilon$ элементов (поскольку соответствующие события не пересекаются и их суммарная вероятность не больше 1). Наконец, B перечислимо (строя приближения снизу к $a(x)$, мы рано или поздно обнаружим любое x, для которого $a(x) > \varepsilon$).

Этих свойств уже достаточно, чтобы заключить, что последовательность ω вычислима. В самом деле, рассмотрим максимально возможное число попарно несравнимых слов $x_1, \ldots, x_N$ из B. Для каждого из слов x_i рассмотрим все его продолжения, лежащие в B. Все они сравнимы (иначе можно было бы заменить x_i на два несравнимых продолжения и увеличить N). Поэтому из каждого x_i выходит конечная или бесконечная ветвь без ветвлений, и эта ветвь вычислима (поскольку можно перечислять множество B). Последовательность ω является одной из таких ветвей (если бы ω не проходила ни через одно из слов x_i, то её достаточно длинное начало можно было бы добавить к списку несравнимых слов и увеличить N).

(ж) Построим вероятностную машину в смысле главы 4, применяя f к выходу вероятностной машины, соответствующей наибольшей перечислимой полумере на дереве, и сравним результат с наибольшей перечислимой полумерой на $\mathbb{N}$ (логарифм которой равен $\mathit{KP} + O(1)$). $\triangleright$

Отметим, что априорная сложность отличается по своим свойствам от уже знакомых нам (обычной и префиксной) сложностей. Прежде всего, её определение использует структуру начал на множестве двоичных слов, и потому алгоритмические преобразования, не сохраняющие этой структуры, могут увеличивать сложность более чем на константу.

100 Покажите, что априорная сложность слова x может быть ограниченной, а сложность слова x^R (те же биты в обратном порядке) — сколь угодно большой. (Формально: существует такое c, что для любого n найдётся слово x с $\mathit{KA}(x) < c$ и $\mathit{KA}(x^R) > n$.) [Указание: слово x начинается с единицы, дальше идут одни нули.]

Тем самым нет смысла (в отличие от префиксной и обычной сложностей) говорить об априорной сложности, скажем, натурального числа или графа.

Априорная сложность слова длины n отличается от уже известных нам вариантов сложностей не более чем на $O(\log n)$, но здесь важно, что под логарифмом стоит именно длина слова, а не его сложность, поскольку, скажем, для слов из одних нулей априорная сложность ограничена, а обычная и префиксная — нет.

101 Покажите, что разности $\mathit{KS}(x) - \mathit{KA}(x)$ и $\mathit{KA}(x) - \mathit{KS}(x)$ могут быть порядка $\log n$ для некоторых слов длины n (и для сколь угодно больших n). [Указание. $\mathit{KS}(x)$ будет больше $\mathit{KA}(x)$, если взять слово из одних нулей. Обратное соотношение имеет место для начал случайных по равномерной мере последовательностей, у которых $\mathit{KA}(x) = l(x) + O(1)$ (см. раздел 5.6), а сложность бывает меньше длины примерно на логарифм длины, см. задачу 38.]

102 Докажите, что

$$\mathit{KA}(xy) \leqslant \mathit{KP}(x) + \mathit{KA}(y) + O(1),$$

где xy — соединение слов x и y. Здесь существенно, что речь идёт о соединении именно в таком порядке: покажите, что для другого порядка это неверно. [Указание. Пусть U —

вероятностный алгоритм в смысле главы 4, порождающий наибольшую перечислимую снизу полумеру (на словах как на изолированных объектах), а V — вероятностный алгоритм в смысле этой главы, порождающий наибольшую полумеру на дереве. Рассмотрите алгоритм, который вначале действует как U, а после остановки продолжает работу как V (читая остаток входной последовательности, не прочитанный алгоритмом U, и добавляя биты к уже выданным). Чтобы показать, что для $\mathit{KA}\,(yx)$ аналогичная оценка места не имеет, положите $y = 0^n$ и $x = 1$.]

$\boxed{\mathbf{103}}$ [increasing-apriory-complexity] Докажите, что для любого слова x хотя бы одно из чисел $\mathit{KA}\,(x0)$ и $\mathit{KA}\,(x1)$ не меньше $\mathit{KA}\,(x) + 1$. (Здесь существенно, что $\mathit{KA}\,(x)$ определено как $-\log a(x)$ без округления.) Выведите отсюда, что для любого слова x и числа $n \in N$ можно найти слово y длины n, для которого $\mathit{KA}\,(xy) \geqslant \mathit{KA}\,(x) + n$.

(Ср. теорему 65 на с. 110 и задачу 34 на с. 45; заметим, что здесь нет условия n и даже константы $O(1)$.)

Ещё одно свойство априорной сложности непосредственно следует из определения. Пусть дана произвольная вычислимая мера μ на пространстве Ω. Тогда для некоторой константы c и для всех слов x выполняется неравенство

$$\mathit{KA}\,(x) \leqslant -\log \mu(\Omega_x) + c.$$

В самом деле, априорная вероятность на дереве не меньше меры μ (и даже любой другой перечислимой снизу полумеры) с точностью до ограниченного множителя, что после логарифмирования и даёт требуемое неравенство.

Мы обращаем внимание на это (простое) свойство, поскольку на нём основан критерий случайности по Мартин-Лёфу: последовательсть ω случайна по вычислимой мере μ тогда и только тогда, когда для её начальных отрезков x это неравенство превращается в равенство, то есть когда разница $-\log \mu(\Omega_x) - \mathit{KA}\,(x)$ ограничена сверху (снизу она всегда ограничена в силу только что рассмотренного свойства).

Этот критерий является следствием теоремы Левина – Шнорра (критерий случайности в терминах монотонной сложности) и будет доказан вместе с ней в разделе 5.6. Но прежде нам необходимо дать определение монотонной сложности (раздел 5.5), для чего подробнее рассмотреть понятие вычислимого отображения пространства Σ в себя (раздел 5.4).

Можно охарактеризовать априорную сложность и как минимальную перечислимую сверху функцию, удовлетворяющую некоторым условиям — подобно тому, как это сделано в теореме 8 (с. 25) для простой сложности и в теореме 56 (с. 98) для префиксной сложности. А именно, справедлива следующая теорема:

Theorem 73. [ka-criterion] *Функция KA является минимальной с точностью до константы перечислимой сверху функцией K с таким свойством:*

$$\sum_{x \in M} 2^{-K(x)} \leqslant 1$$

для любого множества M попарно несравнимых двоичных слов.

$\lhd$ Поскольку слова $x \in M$ несравнимы, соответствующие множества Σ_x не пересекаются и вероятности попадания в них в сумме не превосходят единицы.

С другой стороны, пусть дана произвольная перечислимая сверху функция K, обладающая указанным в теореме свойством. Нам нужно определить перечислимую снизу полумеру, которая была бы не меньше (перечислимой снизу) функции 2^{-K}. Заметим, что функция 2^{-K} не обязана быть полумерой, её значения на словах x, $x0$ и $x1$, вообще говоря, никак не связаны. Поэтому её надо увеличить — в той мере, в которой это неизбежно. А именно, положим $a(x)$ равным точной верхней грани всех сумм вида

$$\sum_{y \in M} 2^{-K(y)}$$

по всем множествам M попарно несравнимых продолжений слова x. Легко проверить, что это действительно будет перечислимая снизу полумера, которая не меньше 2^{-K}, что и требовалось. $\triangleright$

5.4. Вычислимые отображения $\Sigma \to \Sigma$

[tree-mappings]

Алгоритмы (машины), рассматриваемые нами при определении априорной вероятности (универсальной полумеры) на дереве, состоят из двух частей: датчика случайных чисел, порождающего последовательность случайных битов, и алгоритма, который порождает выходные биты, используя эту последовательность. Сейчас мы изучим подробнее эту вторую составляющую, введя понятие вычислимого отображения пространства Σ (конечных и бесконечных последовательностей нулей и единиц) в себя. Отметим, что мы рассматриваем отображения, определённые на всём Σ (зато среди значений может быть пустое слово, что в некотором смысле заменяет неопределённость).

5.4.1. Непрерывные отображения $\Sigma \to \Sigma$

[tree-continuous] Пусть дано отображение $f\colon \Sigma \to \Sigma$, определённое на всём Σ. Мы будем называть его *непрерывным*, если выполнены два свойства:

(1) оно монотонно: если $x \in \Sigma$ является началом $y \in \Sigma$, то $f(x)$ является началом $f(y)$;

(2) значение отображения f на бесконечной последовательности ω является объединением (наименьшим общим продолжением) значений $f(x)$ на всех конечных началах x последовательности Ω.

Мы будем использовать обозначение $x \preccurlyeq y$ для отношения «x является началом y»; предполагается, что $x, y \in \Sigma$ могут быть и конечными, и бесконечными. Если $x \preccurlyeq y$ для бесконечной последовательности x, то $x = y$. Требование (1) представляет собой монотонность f с точки зрения частичного порядка $\preccurlyeq$ на Σ. Оно гарантирует, что значения $f(x)$ на конечных началах x последовательности ω будут продолжать друг друга; их объединение (точная верхняя грань в смысле $\preccurlyeq$-порядка) согласно требованию (2) должно совпадать с $f(\omega)$.

104 Покажите, что это определение соответствует стандартному понятию непрерывности для отображений топологических пространств, если Σ снабдить топологией раздела 4.4.3 (с. 86). [Указание: ср. аналогичный результат для непрерывных отображений пространства Σ в $\mathbb{N}_\perp$ в том же разделе.]

С каждым непрерывным отображением $f\colon \Sigma \to \Sigma$ свяжем множество Γ_f, которое естественно назвать *подграфиком* отображения f; его элементами являются пары слов $\langle x, y\rangle$, где x, y конечные слова, для которых $y \preccurlyeq f(x)$. Имеет место следующее простое наблюдение: для любого непрерывного f множество Γ_f обладает следующими тремя свойствами:

(1) $\langle x, \Lambda\rangle \in \Gamma_f$ для любого слова x;

(2) если $\langle x, y\rangle \in \Gamma_f$, то $\langle x', y'\rangle \in \Gamma_f$ при любых $x' \succcurlyeq x$, $y' \preccurlyeq y$.

(3) если $\langle x, y_1\rangle$ и $\langle x, y_2\rangle$ принадлежат Γ_f, то слова y_1 и y_2 сравнимы (одно из них является началом другого).

Первые два свойства очевидны, третье следует из того, что два слова, являющиеся началами конечной или бесконечной последовательности, сравнимы. Следующая теорема показывает, что непрерывные отображения однозначно задаются своими подграфиками.

Theorem 74. [continuous-graphs] *Соответствие $f \mapsto \Gamma_f$ является взаимно однозначным соответствием между непрерывными отображениями $\Sigma \to \Sigma$ и множествами пар слов, обладающими свойствами* (1) – (3).

$\lhd$ Пусть дано множество F пар слов, обладающее свойствами (1) – (3). Требования (1) – (3) гарантируют, что для любого слова x множество тех y, при которых $\langle x, y\rangle \in F$, непусто и состоит из сравнимых слов. Их объединение (точную верхнюю грань) и будем считать значением $f(x)$. Свойство (2) гарантирует, что $x \preccurlyeq x'$ влечёт $f(x) \preccurlyeq f(x')$ (с ростом x рассмотренное множество слов y также растёт). Тем самым можно корректно определить $f(\omega)$ как объединение $f(x)$ для всех слов $x \preccurlyeq \omega$. Построенное отображение f будет непрерывно в смысле данного выше определения. Легко проверить, что построенное соответствие $F \mapsto f$ обратно к $f \mapsto \Gamma_f$. $\rhd$

5.4.2. Монотонные машины с неблокирующим чтением

Непрерывное отображение $f\colon \Sigma \to \Sigma$ будем называть *вычислимым*, если соответствующее множество Γ_f перечислимо. (Вычислимые отображения по определению всегда непрерывны.)

Формально нам не требуется никакого «оправдания» для этого определения, и всё сказанное дальше в этом разделе о «машинно-зависимой» интерпретации вычислимых отображений нигде не потребуется. Тем не менее поучительно понять, какому виду машин (программ) соответствует такое определение.

Будем рассматривать программы, которые используют неблокирующее чтение со входа (можно прочесть очередной бит из очереди входных битов и проверить, пуста ли эта очередь). Такие программы подробно обсуждались в разделе 4.4.2, с. 85. Но теперь будем считать, что машина строит выход по битам, имея команду *OutputBit*(b) с битовым аргументом.

Выходная последовательность такой программы может быть конечной или бесконечной, и зависит, вообще говоря, не только от входной последовательности битов (нажатых клавиш «0» и «1»), но и от моментов нажатия клавиш. Будем называть программу *корректной*, если такой зависимости нет (моменты нажатия клавиш могут влиять на моменты появления битов на выходе, но не на выходную последовательность). Корректная программа задаёт некоторое отображение множества Σ в себя.

Theorem 75. *Корректные программы задают вычислимые отображения (в описанном выше смысле); всякое вычислимое отображение может быть задано некоторой корректной программой.*

$\lhd$ Пусть имеется некоторая корректная программа M и две входные последовательности x и x', причём $x \preccurlyeq x'$. Покажем, что $M(x) \preccurlyeq M(x')$, где через $M(z)$ обозначается выход программы M на входе z (не зависящий, по предположению, от моментов времени подачи на вход битов слова z). При бесконечном x это заведомо так (поскольку $x = x'$). Пусть x конечно. Подадим x на вход и будем ждать, пока на выходе не появится $M(x)$. Это рано или поздно должно случиться, если $M(x)$ конечно, после чего мы подадим на вход недостающие символы из x'. Полученная после этого на выходе последовательность $M(x')$ неизбежно будет продолжением $M(x)$. Таким образом, случай конечного $M(x)$ рассмотрен. Если же $M(x)$ бесконечно, то любой бит $M(x)$ в некоторый момент должен появиться после подачи x на вход. Поскольку мы можем подать остаток x' после этого момента, тот же бит должен присутствовать и в $M(x')$, так что в этом случае $M(x) = M(x')$.

Столь же ясно, что для бесконечного ω значение $M(\omega)$ есть объединение значений $M(x)$ для конечных $x \preccurlyeq \omega$, поскольку в каждый момент вычисления на вход подано лишь конечное число битов последовательности ω.

Множество пар слов x, y, для которых $y \preccurlyeq M(x)$, перечислимо, так как его можно перечислять, моделируя работу машины M на всевозможных входах. Таким образом, каждой корректной машине соответствует некоторое вычислимое (в абстрактном смысле) отображение.

Напротив, пусть имеется произвольное вычислимое отображение f. Построим машину M, которая корректно вычисляет его. А именно, M перечисляет подграфик Γ_f и параллельно с этим читает биты со входа. Как только в Γ_f обнаружится некоторая пара $\langle x, y\rangle$, для которой x является началом входа, мы выдаём недостающие биты слова y (требования (1) – (3) гарантируют, что все обнаруживающиеся слова y будут сравнимы, так что отзывать биты с выхода не придётся). $\rhd$

5.4.3. Перечислимость множества вычислимых отображений

Определение вычислимости с корректными программами кажется более естественным, но оно (как и для случая префиксно корректных программ) имеет важный недостаток: нет алгоритма, позволяющего по данной программе определять, будет ли она корректной. Тем не менее можно алгоритмически преобразовать каждую программу в корректную: надо перейти от неё к множеству пар, «скорректировать» это множество и затем обратно перейти к программе. Мы не будем это описывать подробно, поскольку корректные программы для нас — всего лишь мотивировка понятия вычислимого отображения, а ограничимся утверждением о перечислимости множества вычислимых отображений, которое нам понадобится в дальнейшем.

Theorem 76. [monot-enumerable] *Существует перечислимое множество U троек $\langle n, x, y\rangle$ (где n — натуральные числа, а x и y — слова) с такими свойствами:*

(1) *при любом n множество $U_n = \{\langle x, y\rangle \mid \langle n, x, y\rangle \in U\}$ является подграфиком некоторого вычислимого отображения $u_n\colon \Sigma \to \Sigma$ (то есть удовлетворяет свойствам* (1) – (3) *теоремы 74*).

(**2**) *среди отображений u_n встречаются все вычислимые отображения Σ в себя.*

$\triangleleft$ Рассмотрим универсальное множество W троек того же вида, среди сечений W_n которого встречаются все перечислимые множества пар слов. Далее изменим W, «скорректировав» все сечения W_n. Мы хотим, чтобы после коррекции сечение заведомо удовлетворяло свойствам (1) – (3) (и тем самым было подграфиком вычислимого отображения по теореме 74), но чтобы корректные сечения при этом не менялись.

Коррекция состоит из двух этапов: сначала устраняются «противоречия», а затем восполняются «пробелы». Противоречие образуют две пары $\langle x_1, y_1\rangle$ и $\langle x_2, y_2\rangle$, в которых x_1 сравнимо с x_2, а y_1 не сравнимо с y_2. (Легко видеть, что в подграфике таких пар быть не может.) Противоречия устраняются естественным способом: выбрасываются пары, противоречащие ранее появившимся (и не выброшенным). Полученное множество перечислимо. После этого восполняются пробелы: добавляются все пары вида $\langle x, \Lambda\rangle$; если в множестве есть пара $\langle x, y\rangle$, то добавляются также и все пары $\langle x', y'\rangle$ с $x' \succcurlyeq x$ и $y' \preccurlyeq y$. Восполнение пробелов также сохраняет перечислимость; легко проверить, что полученное после этого множество удовлетворяет нашим требованиям. $\triangleright$

Это утверждение, которое можно назвать перечислимостью множества всех вычислимых отображений Σ в себя, будет использовано при построении монотонной сложности в следующем разделе.

5.5. Монотонная сложность

[monotone-complexity] При определении монотонной сложности в качестве способов описания (декомпрессоров) рассматриваются вычислимые отображения $D\colon \Sigma \to \Sigma$. *Монотонной сложностью* слова x при данном способе описания D называется наименьшая длина слова y, для которого $x \preccurlyeq D(y)$. Она обозначается $\mathit{KM}_D(x)$.

(Это определение безо всяких изменений переносится и на бесконечные последовательности x, но мы следуем традиции и рассматриваем лишь двоичные слова, если это не оговорено особо.)

$\boxed{\mathbf{105}}$ Докажите, что естественным образом определённая сложность бесконечной последовательности равна пределу неубывающей последовательности сложностей её начальных отрезков.

Theorem 77. *Существует оптимальный способ описания, то есть вычислимое отображение $D\colon \Sigma \to \Sigma$, для которого KM_D минимальна с точностью до константы: для всякого вычислимого $D'\colon \Sigma \to \Sigma$ найдётся такая константа c, что*

$$\mathit{KM}_D(x) \leqslant \mathit{KM}_{D'}(x) + c$$

для любого слова x.

$\triangleleft$ Пусть U — множество троек, сечениями которого являются все подграфики вычислимых отображений и только они (теорема 76, с. 129). Пусть D_n — вычислимое отображение, соответствующее сечению U_n. Определим отображение D формулой

$$D(\hat{n}z) = D_n(z),$$

где $\hat{n}$ — беспрефиксный код числа n (скажем, его двоичная запись с удвоенными цифрами, за которой следует 01), а z — произвольный элемент Σ. В терминах подграфика: рассмотрим множество всех пар $\langle \hat{n}u, v\rangle$, для которых $\langle n, u, v\rangle \in U$. Легко проверить, что действительно получается вычислимое отображение и что если способ описания D' имеет номер n (подграфик D' совпадает с U_n), то $\mathit{KM}_D(x) \leqslant \mathit{KM}_{D'}(x) + l(\hat{n})$ при всех x. $\triangleright$

Как обычно, мы фиксируем некоторый оптимальный «монотонный способ описания» (вычислимое отображение D, для которого выполнено утверждение этой теоремы) и *монотонной сложностью* слова x называем $\mathit{KM}_D(x)$. Обозначение: $\mathit{KM}(x)$.

Theorem 78. (а) *Монотонная сложность монотонна: $\mathit{KM}(x) \leqslant \mathit{KM}(y)$ при $x \preccurlyeq y$;*

(б) *функция KM перечислима сверху;*

(в) $\mathit{KM}(x) \leqslant l(x) + O(1)$;

(г) $\mathit{KM}(x) \leqslant \mathit{KP}(x) + O(1)$;

(д) $\mathit{KA}(x) \leqslant \mathit{KM}(x) + O(1)$;

(е) *бесконечная последовательность нулей и единиц вычислима тогда и только тогда, когда монотонная сложность её начальных отрезков ограничена;*

(ё) *если $f\colon \Sigma \to \Sigma$ — вычислимое отображение, то $\mathit{KM}(f(x)) \leqslant KM(x) + O(1)$ (скрытая в $O(1)$ константа может зависеть от f, но не от x);*

(ж) *если $f\colon \Sigma \to \mathbb{N}_\perp$ — вычислимое отображение, то $\mathit{KP}(f(x)) \leqslant KM(x) + O(1)$ (скрытая в $O(1)$ константа может зависеть от f, но не от x).*

Поучительно сравнить эти утверждения со свойствами априорной сложности (теорема 72, с. 123). Поскольку монотонная сложность не меньше априорной (утверждение (д)), некоторые свойства априорной сложности сразу же переносятся на монотонную. В частности, мы получаем, что $\mathit{KP}(x|l(x)) \leqslant \mathit{KM}(x) + O(1)$ и $\mathit{KP}(x) \leqslant \mathit{KM}(x) + \mathit{KP}(l(x)) + O(1)$. Кроме того, отметим, что для вычислимых последовательностей несравнимых слов префиксная и априорная сложности совпадают с точностью до $O(1)$ — и потому совпадают с находящейся между ними монотонной: если $x_0, x_1, \ldots$ — такая последовательность, то $\mathit{KM}(x_i) = \mathit{KA}(x_i) + O(1) = \mathit{KP}(x_i) + O(1)$.

$\triangleleft$ Утверждение (а) непосредственно следует из определения: если $D(u) \succcurlyeq y$, то и $D(u) \succcurlyeq x$ для любого начала x слова y. Можно сказать, что при определении монотонной сложности требуется описать не само слово, а любое из его продолжений, и при удлинении слова эта задача усложняется (множество продолжений становится меньше).

Утверждение (б) сразу следует из определения: подграфик вычислимого отображения перечислим.

Чтобы доказать (в), достаточно заметить, что тождественное отображение $\Sigma \to \Sigma$, для которого $D(x) = x$ при всех $x \in \Sigma$, вычислимо.

Чтобы сравнить KM и KP (пункт (г) теоремы), достаточно заметить, что любое вычислимое отображение $\Sigma \to \mathbb{N}_\perp$ можно переделать в отображение $\Sigma \to \Sigma$ (отобразив $\mathbb{N}_\perp$ в Σ: элемент $\perp$ переходит в пустое слово). Более формально, если D — оптимальный префиксно корректный способ описания, используемый при определении KP, то его же можно считать вычислимым отображением $\Sigma \to \Sigma$, доопределив пустым словом на тех аргументах, где D не был определён, и продолжив на бесконечные последовательности по непрерывности.

Для сравнения *KM* и *KA* (пункт (д) теоремы) надо вспомнить, с чего мы начинали обсуждение вычислимых отображений: мы говорили, что вероятностный алгоритм представляет собой датчик случайных битов, к выходу которого применяется вычислимое отображение пространства Σ в себя. Пусть D — оптимальный способ описания, используемый при определении монотонной сложности. Рассмотрим вероятностную машину, которая состоит в применении D к последовательности случайных битов. Нас интересует вероятность того, что на её выходе появится слово x или какое-то его продолжение. Ясно, что эта вероятность не меньше $2^{-l(y)}$ для любого слова y, при котором $D(y) \succcurlyeq x$, поскольку датчик случайных битов с вероятностью $2^{-l(y)}$ выдаст последовательность, начинающуюся на y, а применение D после этого даст последовательность, начинающуюся на x. (Подробнее о сравнении *KM* и *KA* см. формулировку теоремы 80 и её обсуждение.)

Утверждение (е) в одну сторону прямо следует из соответствующего утверждения теоремы 72, а в другую очевидно: начальные отрезки вычислимой последовательности ω имеют ограниченную монотонную сложность, поскольку можно рассмотреть вычислимое отображение $\Sigma \to \Sigma$, равное ω на всех аргументах.

В пункте (ё) допустим и случай, когда $f(x)$ бесконечно (если стремиться этого избежать, то надо говорить о сложности всех конечных начал $f(x)$). Для доказательства достаточно рассмотреть способ описания, являющийся композицией оптимального (использованного при определении монотонной сложности) способа и отображения f.

Наконец, в пункте (ж) надо заметить, что композиция оптимального для монотонной сложности способа описания и f является префиксно корректным способом описания. Кроме того, это можно вывести из аналогичного утверждения про априорную сложность. $\triangleright$

$\boxed{106}$ `[prefix-monotone]` Докажите, что $KM(xy) \leqslant KP(x) + KM(y) + O(1)$ (здесь xy — конкатенация слов x и y). В частности, $KM(xy) \leqslant KP(x) + l(y) + O(1)$. [Указание. Рассмотрим оптимальный беспрефиксный способ описания D_p и оптимальный монотонный способ описания D_m. Теперь положим $D'(uv) = D_p(u)D_m(v)$ (когда D_p прочитает всё, что захочет, непрочитанная часть рассматривается как вход для D_m).]

$\boxed{107}$ `[kp-km]` Покажите, что в предыдущей задаче можно заменить $KM(y)$ в правой части на «условную» монотонную сложность $KM(y|x)$, определив её естественным образом (монотонности по аргументу x не требуется; подробнее см. в главе 6).

$\boxed{108}$ Докажите, что утверждение (ё) остаётся верным, если заменить *KM* на *KA* (в обеих частях). [Указание: отображение f можно применять к выходу вероятностной машины; получится новая вероятностная машина, которая не хуже оптимальной.]

Определение монотонной сложности можно дать и формально более простым (хотя, на наш взгляд, менее естественным) образом. Вот как это делается. Напомним, что через Ξ мы обозначаем множество всех двоичных слов. На этом множестве рассмотрим отношение «быть сравнимым»: x сравнимо с y, если $x \preccurlyeq y$ или $y \preccurlyeq x$. Перечислимое отношение $D \subset \Xi \times \Xi$ будем называть *корректным*, если оно обладает таким свойством: для всех x_1, x_2, y_1, y_2

$$\langle x_1, y_1\rangle \in D, \quad \langle x_2, y_2\rangle \in D \quad \text{и} \quad (x_1 \text{ сравнимо с } x_2) \ \Rightarrow\ (y_1 \text{ сравнимо с } y_2)$$

Далее монотонную сложность слова y относительно D определяем как минимальную длину x, при котором $\langle x, y\rangle \in D$, и доказываем утверждение о существовании оптимального корректного перечислимого отношения.

$\boxed{109}$ Покажите, что такой подход приводит к величине монотонной сложности, отличающейся от определённой нами не более чем на ограниченное слагаемое. [Указание. Подграфик всякого вычислимого отображения $\Sigma \to \Sigma$ является корректным множеством в только что описанным смысле. Напротив, если множество D корректно, то после «восполнения пробелов» (см. доказательство теоремы 76) оно превращается в подграфик вычислимого отображения.]

Сравнивая это определение с определением обычной колмогоровской сложности (где роль D играют графики всех вычислимых функций, то есть произвольные однозначные по второму аргументу множества), можно увидеть отличия между *KS* и *KM*. В данном случае мы не требуем однозначности: при одном и том же x могут быть пары $\langle x, y\rangle$ с разными y; нужно лишь, чтобы все эти y были согласованы, то есть были различными началами какой-то одной последовательности нулей и единиц. За счёт этого сложность уменьшается. Например, это позволяет всем началам данной вычислимой последовательности иметь ограниченную сложность (скажем, 0^n имеет ограниченную монотонную сложность, в то время как $KS(0^n) = KS(n)$ может достигать логарифма n).

С другой стороны, мы накладываем и дополнительные ограничения: если x является описанием какого-то y, то сравнимые с x слова могут описывать лишь слова, сравнимые с y. За счёт этого сложность увеличивается. Это особенно хорошо видно, когда мы говорим о несравнимых словах (элементах вычислимой последовательности): монотонная сложность при этом обращается в префиксную, и разница между ней и обычной сложностью также может иметь порядок $O(\log n)$ для слов длины n.

Суммируя эти наблюдения (и вспоминая, что и априорная сложность, и обычная сложность отличаются от префиксной не более чем на $O(\log n)$ для слов длины n), приходим к такому утверждению:

Theorem 79. [km-ks-difference] *Разница между KS(x) и KM(x) не превышает $O(\log n)$ для слов длины n и может достигать $\log n - O(1)$ в обе стороны для слов длины n при бесконечно многих n.*

Мы вернёмся к вопросу о связи различных вариантов определения сложности в главе 6. Сейчас мы приведём (без доказательства) лишь одно утверждение такого рода:

Theorem 80. [gacs-difference] *Величина KM(x) − KA(x) не ограничена сверху.*

Вспомним, что и при определении *KM*(x), и при определении *KA*(x) выбирается некоторое непрерывное отображение $f\colon \Sigma \to \Sigma$. Затем рассматривается прообраз множества всех продолжений слова x при этом отображении. Разница между *KA* и *KM* состоит в том, что в первом случае нас интересует мера этого прообраза, а во втором случае — мера наибольшего интервала вида Σ_y, содержащегося в этом прообразе. Из этого описания видно, что $KA_f \leqslant KM_f$, и разница может быть большой, если прообраз «разреженный». Вопрос в том, насколько это возможно для оптимального способа описания.

Однако этого мало. Вспомним метафору с выделением места на отрезке $[0, 1]$ счётному числу клиентов в доказательствах теорем 40 (с. 75) и 52 (с. 91). Разница между префиксной сложностью и логарифмом априорной вероятности на $\mathbb{N}$ также возникала из-за того, что в одном случае нас интересовал общий размер прообраза, а в другом случае — длина наибольшего непрерывного участка). Однако тогда за счёт перехода к другому способу описания удавалось обойтись лишь постоянным множителем.

Теперь задача усложняется за счёт того, что клиенты объединены в иерархическую структуру типа дерева. Из-за этого дополнительное ужесточение требований (учёт лишь максимального непрерывного участка памяти, а не общего её количества) приводит к бо́льшим потерям памяти.

К сожалению, авторам книги не удалось настолько хорошо разобраться в (технически довольно сложном) доказательстве этого утверждения из статьи Питера Гача [17], чтобы изложить его просто и понятно, так что приходится отсылать читателей к указанной статье.

[*А всё-таки жаль, что...* может быть, можно разобраться?!]

110 [km-ka-relation] Докажите, что $\mathit{KM}(x) \leqslant \mathit{KA}(x) + O(\log \mathit{KA}(x))$. [Указание. Верно более сильное утверждение: $\mathit{KM}(x|\,\mathit{KA}(x)) \leqslant \mathit{KA}(x) + O(1)$. В самом деле, если $\mathit{KA}(x) = k$, то x рано или поздно попадёт в растущее дерево слов априорной сложности менее $k+1$. Это дерево в каждый момент имеет ширину (максимальное число попарно несравнимых элементов) не более 2^{k+1}, поэтому его можно покрыть не более чем 2^{k+1} растущими ветвями, следя за максимальными элементами дерева. Подробно см. теорему 102, с. 164.]

5.6. Теорема Левина – Шнорра

[monotlev]

Определение априорной сложности гарантирует, что для любой перечислимой снизу полумеры p выполняется неравенство $\mathit{KA}(x) \leqslant -\log p(x) + O(1)$. Оказывается, что если p является мерой, то это неравенство выполнено не только для KA, но и для (вообще, говоря большей) величины KM.

Theorem 81. [monotlev-upper-bound] *Пусть μ — вычислимое распределение вероятностей на Ω и $p(x) = \mu(\Omega_x)$. Тогда существует такая константа c, что*

$$\mathit{KM}(x) \leqslant -\log p(x) + c$$

для любого слова x.

$\triangleleft$ Идею доказательства можно объяснить так: различие между KM и KA возникает из-за невозможности выделить сплошные участки отрезка $[0,1]$ по требованию. А это невозможно из-за того, что мы не знаем, какие из уже предъявленных требований возрастут, а какие нет, и не можем зарезервировать место. Но когда имеется не полумера, а вычислимая мера, то такой проблемы не возникает.

Если эти объяснения непонятны, их можно пропустить — сейчас мы изложим формальное доказательство.

Для каждого двоичного слова x рассмотрим отрезок π_x — часть отрезка $[0,1]$, определяемую по следующим правилам:

- длина отрезка π_x равна $p(x)$;
- $\pi_\Lambda = [0,1]$ (где Λ — пустое слово);

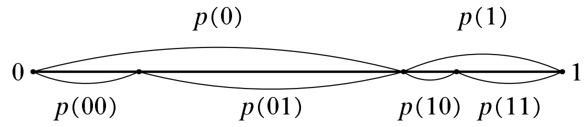

Рис. 12. Построение отрезков π_x.

[monotlev-pict1]

- для любого слова x отрезок π_x делится некоторой своей точкой на π_{x0} (левая часть) и π_{x1} (правая часть).

(См. рисунок 12.)

Мы будем сравнивать отрезки π_x с аналогичными отрезками для равномерной меры: через I_x мы обозначим отрезок, в котором двоичные записи чисел начинаются на x. Отрезки вида I_x будем называть *двоичными.*

Теперь рассмотрим множество G всех пар слов $\langle x, y\rangle$, для которых (двоичный) отрезок I_x лежит строго внутри отрезка π_y. Множество G перечислимо (поскольку функция p вычислима, мы можем находить концы отрезков π_y с любой точностью, и если они строго больше (или меньше) некоторого рационального числа, то рано или поздно это обнаружится). Кроме того, свойство $\langle x, y\rangle \in G$ сохранится, если заменить x на любое его продолжение (отрезок I_x станет меньше) или заменить y на любое начало (отрезок π_y станет больше). Если $\langle x, y_1\rangle \in G$ и $\langle x, y_2\rangle \in G$, то отрезки π_{y_1} и π_{y_2} имеют общую внутреннюю точку (они оба содержат I_x), и потому слова y_1 и y_2 сравнимы. Поэтому по теореме 74 (с. 128) существует вычислимое отображение Σ в себя, подграфик которого равен G. Используем его в качестве декомпрессора D в определении монотонной сложности. При этом $KM_D(y)$ есть минус двоичный логарифм длины самого большого двоичного отрезка, лежащего строго внутри π_y. Остаётся заметить, что (строго) внутри любого отрезка длины h есть двоичный отрезок длины не менее $h/4$, и воспользоваться оптимальностью. $\triangleright$

$\boxed{\mathbf{111}}$ Проверьте, что сформулированное утверждение о двоичных отрезках действительно верно. [Указание: пусть u — степень двойки, для которой $h/4 \leqslant u < h/2$. Тогда отрезок длины h должен пересечься по крайней мере с тремя последовательными отрезками длины u, и содержит внутри себя средний из них.]

Теорема 81 лежит в основе следующего способа оценки сложности (он применялся А. Н. Колмогоровым и его учениками для русских литературных текстов). Читая текст буква за буквой, экспериментатор пытается угадать следующую букву, формулируя свою догадку в виде распределения вероятностей на всех буквах алфавита. Затем открывается следующая буква и к оценке сложности добавляется величина $-\log p$, где p — приписанная этой букве вероятность. Полученная величина (если считать деятельность экспериментатора вычислимой) является верхней оценкой сложности: в самом деле, экспериментатор задаёт вычислимую меру на последовательностях букв (в форме условных вероятностей следующей буквы при известном начальном отрезке), и сложность текста не больше логарифма этой меры.

Конечно, при реальном проведении таких опытов непрактично требовать указания вероятности для каждой из букв алфавита; разумно ограничиться каким-то классом предсказаний типа «буква А с вероятностью $0{,}5$, остальные гласные равновероятны и в сумме $0{,}3$, остальные буквы равновероятны». Отметим ещё, что происходит при этом скорее оценка сложности данного текста при условии жизненного опыта экспериментатора; ничего удивительного, что получится почти что нуль, если экспериментатор знает текст наизусть (и вообще оценка может уменьшиться, если он хорошо знаком с текстами того же автора).

Теперь всё готово для формулировки критерия случайности (в смысле Мартин-Лёфа) в терминах монотонной сложности; последовательность случайна, если доказанное только что неравенство обращается для её начальных отрезков в равенство. Вот точная формулировка: пусть μ — вычислимое распределение вероятностей на пространстве Ω бесконечных двоичных последовательностей и $p(x) = \mu(\Omega_x)$.

Theorem 82 (Левина – Шнорра). [levin-schnorr1] *Последовательность $\omega \in \Omega$ случайна по Мартин-Лёфу относительно μ тогда и только тогда, когда*

$$-\log p(x) - KM(x) \leqslant c$$

для некоторого c и для всех начальных отрезков x последовательности ω.

$\lhd$ Доказательство состоит из двух частей. Для начала покажем, что если для данной последовательности ω разность $-\log p(x) - KM(x)$ принимает сколь угодно большие значения, то эта последовательность не случайна (множество $\{\omega\}$ является эффективно нулевым).

Пусть дано некоторое число c. Рассмотрим те слова x, у которых разность $-\log p(x) - KM(x)$ больше c. (Эту разность иногда называют *дефектом случайности*, но нужно быть осторожным — это слово употребляется во многих хотя и близких, но всё же различных смыслах; один из них описан в главе 16.) Множество таких слов обозначим D_c.

Множество D_c перечислимо (поскольку p вычислима, а KM перечислима сверху, разность перечислима снизу).

Лемма 1. Множество всех последовательностей, у которых некоторое начало принадлежит D_c, имеет меру не больше 2^{-c} (относительно рассматриваемой нами меры μ).

Неформально говоря, эта лемма верна потому, что на этом множестве мера μ в 2^c раз меньше априорной вероятности на Σ (а последняя в сумме не превосходит единицы). Более формально следует рассуждать так.

Рассматриваемое множество последовательностей есть объединение всех Ω_x при $x \in D_c$. В этом объединении можно оставить лишь минимальные $x \in D_c$, то есть те x, у которых никакие начала не принадлежат D_c. Пусть это будут $x_0, x_1, \ldots$. (Мы не утверждаем, что множество минимальных элементов будет перечислимо, так что эта последовательность может и не быть вычислимой.)

Для каждого x_i рассмотрим кратчайшее описание p_i (согласно определению монотонной сложности: $x_i \preccurlyeq D(p_i)$, где $D\colon \Sigma \to \Sigma$ — оптимальный способ описания). Тогда $l(p_i) = KM(x_i) < -\log p(x_i) - c$. Кроме того, ни одно из p_i не является началом другого (иначе соответствующие x_i были бы сравнимы). Поэтому $\sum_i 2^{-l(p_i)} \leqslant 1$ (как сумма равномерных мер непересекающихся множеств Ω_{p_i}). Соответствующие $p(x_i)$ в 2^c раз меньше, откуда и получаем требуемое. Лемма доказана.

По предположению рассматриваемая нами последовательность ω покрыта множеством из леммы при любом c. Поэтому, чтобы (в соответствии с определением эффективно нулевого множества) указать покрытие множества $\{\omega\}$ перечислимым семейством интервалов с суммарной μ-мерой не больше 2^{-c}, достаточно перечислить интервалы из D_c.

Тут, правда, возникает техническая трудность. Для интервалов из D_c мы знаем, что *мера объединения* не больше 2^{-c} (согласно лемме), в то время как определение требует, чтобы *сумма мер* интервалов не превосходила 2^{-c}. Эту проблему нельзя решить, оставив в D_c минимальные точки, поскольку их множество может уже не быть перечислимым. Вместо этого можно использовать такое утверждение:

Лемма 2. Всякое перечислимое множество слов $x_0, x_1, \ldots$ можно преобразовать в перечислимое множество несравнимых слов, сохранив неизменным объединение $\cup_i \Omega_{x_i}$. Это преобразование эффективно (алгоритм, перечисляющий первое множество, можно вычислимо преобразовать в алгоритм перечисления второго).

В самом деле, если при перечислении появляется продолжение ранее рассмотренного слова, то его можно просто выбросить (соответствующий интервал в объединении и так покрыт). Если же появляется слово y, являющееся (собственным) началом ранее рассмотренного слова x, то нужно разбить добавляемое множество $\Omega_y \setminus \Omega_x$ в объединение интервалов и заменить y на слова, задающие эти интервалы. Лемма 2 доказана.

После применения леммы 2 мы получаем перечислимое множество несравнимых слов; заметим, что эти слова уже не обязательно принадлежат D_c, но это и не важно. Достаточно, что они задают непересекающиеся интервалы и что объединение этих интервалов (согласно лемме 1) имеет μ-меру не более 2^{-c}.

Для завершения доказательства теоремы Левина – Шнорра осталось показать, что если последовательность принадлежит эффективно нулевому множеству, то разности (между минус логарифмом меры и монотонной сложностью) для её начальных отрезков не ограничены. Идею этого рассуждения можно описать так: для данного малого по мере множества мы строим монотонный способ описания, благоприятствующий последовательностям из этого множества (дающий малую сложность некоторым их начальным отрезкам).

Более подробно. Пусть последовательность ω принадлежит эффективно нулевому (по мере μ) множеству U. Для каждого c можно эффективно указать покрытие множества U интервалами $\Omega_{x_0}, \Omega_{x_1}, \ldots$, имеющими суммарную меру меньше 2^{-c}. Если увеличить меры всех этих интервалов в 2^c раз, то их сумма останется меньше единицы. Применив к последовательности $p_i = 2^c \mu(\Omega_{x_i})$ теорему 53 (с. 94), мы получаем беспрефиксный способ описания, для которого сложность i не превосходит $-\log \mu(\Omega_{x_i}) - c + 2$. Композиция этого способа описания с вычислимым отображением $i \mapsto x_i$ даёт беспрефиксный способ описания D_c, у которого

$$KP'_{D_c}(x_i) \leqslant -\log \mu(\Omega_{x_i}) - c + 2.$$

(Индекс c у D_c подчёркивает, что всё построение зависит от c.) Монотонная сложность не больше префиксной, поэтому если разница между минус логарифмом меры и префиксной сложностью велика (как показывает только что написанное неравенство при больших c), то тем более будет велика разность с монотонной сложностью. Надо только аккуратно соединить построенные способы описания в один.

По аналогии с построением оптимального способа описания будем считать, что $\hat{c}u$ является описанием слова v, если u является описанием слова v относительно D_c. Здесь $\hat{c}$ — самоограниченная запись натурального числа c длины $O(\log c)$. Для построенного способа описания D уже можно написать неравенство, верное при всех c:

$$KP'_D(x_i) \leqslant -\log \mu(\Omega_{x_i}) - c + O(\log c)$$

Поскольку монотонная сложность не превосходит префиксной, можно заменить $KP'_D(x_i)$ на $KM(x_i)$ и заключить, что у всех слов x_i (для данного c) разность между минус логарифмом меры и монотонной сложностью не меньше $c - O(\log c)$. Для последовательности из U можно найти начало такого вида при любых c, поэтому разности между минус логарифмом меры и монотонной сложностью для начальных отрезков этой последовательности не ограничены.

Теорема Левина – Шнорра доказана. $\triangleright$

[Достаточно перечислимости сверху в одном случае и снизу в другом - В Хорошо бы вставить сюда соответствующую задачу.]

Приведённое доказательство теоремы Шнорра по существу доказывает немного больше. Вот что ещё из него можно заключить:

Theorem 83. [levin-schnorr2] *В формулировке теоремы можно заменить монотонную сложность $KM(x)$ на априорную сложность $KA(x)$.*

$\triangleleft$ Априорная сложность меньше, поэтому разность при переходе к априорной сложности только увеличится; остаётся доказать первую часть теоремы для априорной сложности. Для этого достаточно заметить при доказательстве леммы 1, что $\sum_i 2^{-KA(x_i)} \leqslant 1$, поскольку эта сумма является суммой априорных вероятностей непересекающихся интервалов Ω_{x_i}. $\triangleright$

Theorem 84. [levin-schnorr3] *В формулировке теоремы можно заменить монотонную сложность $KM(x)$ на префиксную сложность $KP(x)$.*

$\triangleleft$ Здесь, напротив, мы увеличиваем сложность, а не уменьшаем, так что сомнения вызываем лишь вторая часть теоремы. Но в доказательстве мы как раз и получали оценку для префиксной сложности. $\triangleright$

Именно такой критерий случайности сейчас наиболее распространён (см., например, [38]). Всё же авторам кажется более естественным использовать монотонную (или априорную) сложность.

Заметим, например, что при использовании префиксной сложности разность, о которой идёт речь в критерии случайности, может быть и отрицательной: скажем, для равномерной меры величина $-\log \mu(\Omega_x)$ есть длина слова x, а префиксная сложность может превышать длину (на величину порядка логарифма длины), см. теорему 57, см. 99.

Кроме того, использование монотонной сложности позволяет усилить теорему следующим образом:

Theorem 85. [levin-schnorr4] *Если последовательность ω не случайна (в смысле Мартин-Лёфа) по мере μ, то величина $-\log p(x) - KM(x)$ не только не ограничена для её начальных отрезков, но и стремится к бесконечности.*

$\triangleleft$ Вспомним доказательство: в нём для последовательности слов x_i строился беспрефиксный способ описания, при котором слово x_i имело описание p_i, причём длина p_i (и тем самым префиксная сложность x_i при этом способе описания) не превосходила $-\log \mu(\Omega_{x_i}) - c$. Чтобы получить нужную оценку для монотонной сложности, мы можем использовать (при каждом i) продолжения слова p_i в качестве описаний продолжений слова x_i, при этом длина первых соответствует логарифму меры вторых, как это делалось при доказательстве теоремы 81 (с. 134).

Более формально можно воспользоваться неравенством $KM(xy) \leqslant KP(x) + KM(y|x)$ (задача 107) и релятивизованным вариантом теоремы 81, утверждающим, что $KM(y|x) \leqslant -\log \mu_x(\Omega_y)$ для любого вычислимого семейства мер, (вычислимо) зависящего от параметра x. При этом в качестве μ_x нужно взять меру, сосредоточенную на продолжениях слова x, для которой $\mu_x(\Omega_y) = \mu(\Omega_{xy})/\mu(\Omega_x)$.

Для случая равномерной меры (когда $-\log \mu(\Omega_x) = l(x)$) всё упрощается и можно просто сказать, что слово $p_i z$ является описанием слова $x_i z$ для любого слова z. $\triangleright$

Мы привели доводы в пользу использования монотонной сложности для характеризации случайности по Мартин-Лёфу. С другой стороны, характеризация случайности в терминах сложности начальных отрезков имеет некоторый дефект: в то время как случайность последовательности нулей и единиц инвариантна относительно вычислимых перестановок индексов (и соответствующего изменения меры), понятие начального отрезка и тем самым критерий случайности в терминах начальных отрезков таковым не является. Используя префиксную сложность, можно модифицировать критерий случайности, сделав его инвариантным.

Пусть F — конечное множество натуральных чисел, а ω — последовательность нулей и единиц. Через $\omega(F)$ будем обозначать сужение ω на F, то есть двоичное слово, образованное битами ω_i при $i \in F$ (в порядке возрастания индексов).

Пусть фиксирована вычислимая мера μ на Ω. Для данного множества F и слова Z, длина которого равна числу элементов в F, можно рассмотреть событие $\omega(F) = Z$. Его вероятность будем обозначать $\mu_{F,Z}$.

112 Докажите, что если последовательность ω случайна (в смысле Мартин-Лёфа) относительно меры μ, то найдётся такое c, что

$$KP(F, \omega(F)) \geqslant -\log \mu_{F,\omega(F)} - c$$

для любого конечного множества F.

[Указание. Мера множества последовательностей, где это условие нарушается для фиксированного c, не превосходит 2^{-c}, умноженного на сумму априорных вероятностей всех пар F, Z, то есть не больше 2^{-c}.]

(Заметим, что если F — начальный отрезок, то F восстанавливается по $\omega(F)$, так что мы приходим к прежней формулировке.)

Сформулированное условие является не только необходимым, но и достаточным. Более того, достаточно потребовать его для произвольной вычислимой возрастающей последовательности множеств, покрывающей все индексы.

113 `[levin-schnorr-subset]` Пусть $F_0 \subset F_1 \subset F_2 \subset \ldots$ — вычислимая последовательность конечных множеств, в объединении дающих всё $\mathbb{N}$. Если для некоторой последовательности ω и для некоторого c неравенство

$$KP(F_i, \omega(F_i)) \geqslant -log \mu_{F_i,\omega(F_i)} - c$$

выполнено при всех i, то последовательность ω случайна в смысле Мартин-Лёфа по мере μ.

[Указание: удобно переставить индексы и считать, что F_i — начальные отрезки. Далее нужно повторить доказательство теоремы Левина – Шнорра, но использовать только интервалы нужных длин (измельчая неформатные).]

Из этого утверждения следует, например, что двумерная последовательность битов (отображение $\mathbb{Z}^2 \to \{0,1\}$) случайна в смысле Мартин-Лёфа по равномерной мере (все биты независимы, нули и единицы равновероятны) тогда и только тогда, когда квадрат $N \times N$ с центром в начале координат (при всех нечётных N) имеет сложность не меньше $N^2 - O(1)$.

Случай равномерной меры заслуживает того, чтобы выписать формулировки всех доказанных нами теорем специально для этого случая (через $(\omega)_n$ мы обозначаем начальный отрезок последовательности ω, имеющий длину n):

Theorem 86. `[levin-schnorr-uniform]`

(а) *Оценка сверху: для любого слова x выполнены неравенства*

$$\mathit{KA}\,(x) \leqslant \mathit{KM}\,(x) + O(1) \leqslant l(x) + O(1);$$

(б) *Критерий случайности: последовательность ω случайна (в смысле Мартин-Лёфа) по равномерной мере тогда и только тогда, когда для её начальных отрезков это неравенство обращается в равенство:*

$$\mathit{KA}\,((\omega)_n) = \mathit{KM}\,((\omega)_n) + O(1) = n + O(1);$$

(в) *Для неслучайных (в смысле Мартин-Лёфа) по равномерной мере последовательностей разность $n - \mathit{KM}\,((\omega)_n)$ и тем самым $n - \mathit{KA}\,((\omega)_n)$ стремятся к бесконечности.*

(г) *Последовательность ω случайна (в смысле Мартин-Лёфа) по равномерной мере тогда и только тогда, когда $\mathit{KP}((\omega)_n) \geqslant n - c$ для некоторого c и для всех n.*

(д) *Последовательность ω случайна (в смысле Мартин-Лёфа) по равномерной мере тогда и только тогда, когда $\mathit{KP}(F, \omega(F)) \geqslant |F| - c$ для некоторого c и для всех конечных множеств F.*

Имеется ещё один критерий случайности по Мартин-Лёфу для случая равномерной меры, который интересен тем, что в нём можно ограничиться использованием обычной колмогоровской сложности (и обойтись без монотонной или префиксной). Даже удивительно, что этот критерий был найден только недавно (см. [50]), поскольку этим занимались ещё в конце 1960-х годов и подошли к нему очень близко (см. [85,47]), а его доказательство достаточно просто и естественно и не выходит из круга идей, хорошо известных в то время.

Theorem 87. `[miller-yu-martin-lof]` *Пусть $f\colon \mathbb{N} \to \mathbb{N}$ — вычислимая функция и ряд $\sum 2^{-f(n)}$ сходится. Пусть ω — случайная (в смысле Мартин-Лёфа) по равномерной мере последовательность. Тогда*

$$\mathit{KS}\,((\omega)_n | n) \geqslant n - f(n) - O(1)$$

(то есть найдётся c, при котором для всех n выполнено неравенство $\mathit{KS}\,((\omega)_n | n) \geqslant n - f(n) - c$).

$\lhd$ Пусть утверждение теоремы не выполнено. Это значит, что для любого c найдётся такое n, что

$$KS((\omega)_n|n) < n - f(n) - c.$$

Другими словами, каково бы ни было c, последовательность ω покрыта некоторым интервалом Ω_x, для которого

$$KS(x|n) < n - f(n) - c,$$

где n — длина x. Для каждого n таких интервалов не более $2^{n-f(n)-c}$, поэтому их суммарная мера не более $2^{-f(n)}2^{-c}$ при данном n и не более

$$2^{-c}\left(\sum_n 2^{-f(n)}\right)$$

для интервалов всех длин. Поэтому последовательность ω образует эффективно нулевое множество: выбирая нужное c, получаем покрытие сколь угодно малой меры. Значит, она не случайна. (Заметим, что сумма ряда $\sum 2^{-f(n)}$ может быть невычислимой; это не важно, так как мы можем использовать любую верхнюю оценку.) $\rhd$

Из этой теоремы следует, например, что для случайной последовательности (по равномерной мере) обычная сложность начальных отрезков не меньше $n - 2\log n - O(1)$ или даже $n - \log n - 2\log\log n - O(1)$, поскольку соответствующие ряды сходятся.

Чем меньше функция f, тем ближе ряд к расходящемуся и тем более ограничительным является утверждение теоремы. Оказывается, что при некоторых f оно становится критерием случайности.

Theorem 88. [miller-yu-proper] *Существует всюду определённая вычислимая функция $f\colon \mathbb{N} \to \mathbb{N}$, для которой $\sum_n 2^{-f(n)} < \infty$, с таким свойством: если для некоторой последовательности ω и для некоторого c при всех n выполнено неравенство*

$$KS((\omega)_n|n) \geqslant n - f(n) - c,$$

то последовательность ω случайна (в смысле Мартин-Лёфа) по равномерной мере.

$\lhd$ Нам нужно доказать, что любая неслучайная последовательность (то есть любая последовательность, содержащаяся в наибольшем эффективно нулевом множестве) имеет «простые» начальные отрезки. При этом критерий «простоты» (точнее, функцию f) нам предстоит выбирать самим.

Как это делается? Пусть, скажем, у нас есть некоторое семейство интервалов с суммой длин не более ε. Пусть F — множество задающих их двоичных слов (семейство состоит из интервалов Ω_u при $u \in F$). Рассортируем слова множества F (и тем самым интервалы семейства) по длинам и подсчитаем для каждого n суммарную меру интервалов, соответствующих словам длины n. Выберем функцию f таким образом, чтобы эта мера была примерно равна $2^{-f(n)}$. Тогда по построению сумма ряда $\sum_n 2^{-f(n)}$ не превосходит ε. С другой стороны, в F имеется примерно $2^{n-f(n)}$ слов длины n и каждое из них может быть описано (при известном n и других параметрах конструкции) своим номером, требующим $n - f(n)$ битов, что даёт верхнюю оценку сложности для слов из F. При этом всякая покрытая нашими интервалами последовательность имеет начальный отрезок в F.

Как это соображение можно применить в нашей ситуации? Рассмотрим наибольшее эффективно нулевое множество. Для любого $\varepsilon > 0$ имеется его покрытие интервалами с суммой длин не больше ε, и описанным способом из этого покрытия получается функция f с $\sum_n 2^{-f(n)} \leqslant \varepsilon$. Далее нам нужно как-то соединить такие функции для разных ε в одну. Будем делать это следующим образом.

Для каждого $c = 0, 1, 2, \ldots$ рассмотрим покрытие с суммой мер меньше 2^{-3c} и построим по нему функцию f описанным способом, а затем уменьшим её на $2c$. Таким образом для каждого c мы получим функцию f_c, для которой

$$\sum_n 2^{-f_c(n)} < 2^{-c}$$

(вместо 2^{-3c} мы получаем 2^{-c}, поскольку уменьшили функцию на $2c$), и множество слов F_c, содержащее $2^{n-f_c(n)-2c}$ слов длины n, причём любая неслучайная последовательность имеет начальный отрезок в F_c.

Далее мы определим $f(n)$ равенством

$$2^{-f(n)} = \sum_c 2^{-f_c(n)},$$

тогда

$$\sum_n 2^{-f(n)} = \sum_n \sum_c 2^{-f_c(n)} = \sum_c \sum_n 2^{-f_c(n)} \leqslant \sum_c 2^{-c} \leqslant 1,$$

и при этом $f(n) \leqslant f_c(n)$ при любых n и c. С другой стороны, множество F_c перечислимо (это гарантируется определением эффективно нулевого множества), и потому любое слово x длины n в F_c может быть задано (при известных n и c) своим порядковым номером в перечислении, который содержит $n - f_c(n) - 2c$ битов:

$$KS(x|n, c) \leqslant n - f_c(n) - 2c + O(1),$$

откуда

$$KS(x|n) \leqslant n - f_c(n) - 2c + O(\log c) < n - f(n) - c$$

для любого слова $x \in F_c$ длины n (при достаточно бóльших c).

Пусть теперь ω — неслучайная последовательность. Из сказанного следует, что она (для каждого c) имеет начало в F_c; если n — длина этого начала (и c достаточно велико), то

$$KS((\omega)_n|n) < n - f(n) - c,$$

что противоречит предположению теоремы.

Это рассуждение, однако, ещё не доказывает теорему — в ней требуется, чтобы функция f была вычислима, в то время как F_c мы можем лишь перечислять (и ни в какой момент не можем быть уверены, что больше слов данной длины не будет). Вспомним, однако, что нас интересуют покрытия интервалами, и в этих покрытиях можно заменить один большой интервал Ω_z на много маленьких Ω_{zt}, если в качестве t взять все слова некоторой фиксированной длины. За счёт этого можно добиться вычислимости функции f_c — надо договориться, что минимально разрешённая длина выдаваемого при перечислении слова растёт со временем. Тогда количество выдаваемых слов длины n является вычислимой функцией от n, так как они могут появляться только до некоторого заранее известного момента. Это рассуждение можно провести параллельно для всех c, и тогда сумма $\sum_c 2^{-f_c(n)}$ также будет вычислима по n.

Наконец, имеется ещё совсем техническая трудность: мы строим функцию f с натуральными значениями, поэтому надо её ещё округлить. $\triangleright$

Две предыдущие теоремы вместе дают критерий случайности, в котором используется обычная (а не монотонная или префиксная) сложность. Этот критерий имеет некоторый «запас». А именно, можно заменить условную сложность $KS((\omega)_n|n)$ на безусловную сложность $KS((\omega)_n)$ или на условную префиксную сложность $KP((\omega)_n|n)$.

В самом деле, при такой замене сложность только увеличится, поэтому адаптации требует лишь теорема 88. Вариант с префиксной сложностью: надо воспользоваться тем, что для любого конечного множества A и любого элемента $x \in A$ справедливо неравенство $KP(x|A) \leqslant \log_2 |A| + O(1)$ (можно рассмотреть беспрефиксное кодирование словами длины $\log_2 |A|$). Немного сложнее удалить условие n (для обычной, не префиксной сложности). В этом случае слово $x \in F_c$, имеющее длину n, надо задавать не его порядковым номером в множестве $F_{c,n}$ слов длины n из множества F_c, а номером во всём F_c (расположенном в порядке возрастания длин). Соответственно число битов, необходимое для этого, есть

$$\log(|F_{c,0}| + |F_{c,1}| + \ldots + |F_{c,n}|),$$

и всё будет в порядке, если последний член $|F_{c,n}|$ больше всех предыдущих вместе взятых (тогда порядковый номер увеличится не более чем в два раза от добавления предыдущих членов). Ясно, что этого можно достичь тем же приёмом (заменяя короткое слово на все его продолжения некоторой длины). Заметим, что это делается при данном c, то есть условие c остаётся, но это не страшно (в итоге оно даёт вклад $O(\log c)$).

Отсюда получаем такой результат:

Theorem 89. [miller-yu-criterion] *Случайность последовательности ω (в смысле Мартин-Лёфа) по равномерной мере равносильна тому, что для любой вычислимой функции f с конечной суммой ряда $\sum_n 2^{-f(n)}$ выполнено свойство*

$$KS((\omega)_n) \geqslant n - f(n) - O(1).$$

Этот критерий использует простую колмогоровскую сложность и именно его обычно называют «теоремой Миллера – Ю».

Недостатком этого критерия можно считать то, что в нём есть квантор по f. И хотя этот квантор можно переставить (найдётся f, которая позволяет отвергнуть все неслучайные последовательности; именно такова формулировка теоремы 88), всё равно хотелось бы как-нибудь обойтись без f. Это можно сделать, но тогда снова приходится упоминать префиксную сложность:

Theorem 90. [miller-yu-prefix] *Последовательность ω является случайной (в смысле Мартин-Лёфа) относительно равномерной меры тогда и только тогда, когда*

$$KS((\omega)_n) \geqslant n - KP(n) - O(1).$$

$\triangleleft$ Если ряд $\sum_n 2^{-f(n)}$ сходится (для вычислимой функции f), то $KP(n) \leqslant f(n)+O(1)$. Поэтому приведённое в теореме условие на ω сильнее, чем в теореме 89.

Таким образом, остаётся доказать утверждение в другую сторону: если для всякого c найдётся n, при котором

$$KS((\omega)_n) < n - KP(n) - c,$$

то последовательность ω не случайна. Это делается так же, как в теореме 87, надо только заметить, что множество всех слов x, для которых

$$KS(x) < l(x) - KP(l(x)) - c$$

(здесь $l(x)$ — длина x), перечислимо. В остальном функция KP ничем не хуже функции f, которая была в том доказательстве. $\triangleright$

Заметим, что и в этой теореме можно заменить $KS((\omega)_n)$ на $KS((\omega)_n|n)$ (годится то же самое рассуждение).

114 Убедитесь в этом.

115 Покажите, что в теореме 88 нельзя положить $f(n) = 2\log n$. [Указание. Из теоремы 87 следует, что для случайной последовательности ω выполняется более сильное неравенство $KS((\omega)_n) \geqslant n - \log n - 2\log\log n - O(1)$. Поэтому, если разбавить её очень редкими нулями на вычислимых местах, то получится последовательность, для которой $KS((\omega)_n) \geqslant n - 2\log n - O(1)$, но не случайная. Аналогичное рассуждение показывает, что не годится никакая функция с вычислимо сходящимся рядом $\sum 2^{-f(n)}$.]

Всё сказанное до сих пор, однако, не отвечает на естественный вопрос: а нельзя ли вообще обойтись без функции f и потребовать, чтобы $KS((\omega)_n)$ было больше $n - O(1)$ (как для монотонной сложности)?

Такое требование было бы самым естественным, но, как сразу же заметил Мартин-Лёф, ничего хорошего не получается: в случайной последовательности должны встречаться все группы цифр, но если слово длины n кончается на k нулей, то его сложность не больше $n-k+2\log k+O(1)$ (для беспрефиксного кодирования числа k достаточно $2\log k$ битов), и разница с длиной будет $k - 2\log k - O(1)$.

Можно оценить неизбежную разницу между длиной и сложностью более точно (см. [85,47]):

Theorem 91. `[martin-lof-oscillations-bound]` *Найдётся такая константа c, что для любой последовательности $\omega \in \Omega$ при бесконечно многих n выполнено неравенство*

$$KS((\omega)_n) \leqslant n - \log n + c.$$

$\triangleleft$ Для каждого n выделим среди слов длины n некоторую часть, составляющую $1/n$ от всех слов длины n. При этом сделаем это так, чтобы у любой бесконечной последовательности было бесконечно много выделенных начал и чтобы множество выделенных слов было разрешимо.

Почему это возможно? Ряд $\sum 1/n$ расходится, поэтому можно его разделить на бесконечное число групп, в каждой из которых сумма больше единицы. С помощью слов каждой группы покроем все последовательности в один слой (каждая последовательность имеет начало из этой группы) — это можно сделать в порядке возрастания длин, покрывая ещё не покрытые слова. (Строго говоря, есть проблема с округлением, поскольку $2^n/n$ — не целое число, но и после округления ряд будет расходиться.)

Любое выделенное слово длины n задаётся (при известном n) своим порядковым номером, который имеет $n - \log n$ битов. Поэтому его условная сложность не больше $n - \log n + O(1)$. Более того, если нумеровать все выделенные слова подряд, то порядковый номер не сильно увеличивается (число выделенных слов растёт почти как геометрическая прогрессия с показателем 2, и добавление всех предыдущих слов увеличивает их число не более чем в $O(1)$ раз).

Отсюда и следует утверждение теоремы. $\triangleright$

116 Покажите, что утверждение теоремы верно не только для некоторого c, но и для всех c (в том числе отрицательных).

[Указание. Если ряд $\sum 2^{-f(n)}$ расходится, то можно немного увеличить функцию f, сохранив это свойство: найдётся функция g, для которой $g(n) - f(n) \to \infty$ и ряд $\sum 2^{-g(n)}$ расходится.]

117 Покажите, что в теореме 91 (в варианте с условной сложностью) можно заменить логарифм на любую вычислимую функцию f, для которой ряд $\sum 2^{-f(n)}$ расходится.

В статье Мартин-Лёфа [47] без доказательства (со ссылкой на его неопубликованную работу) приводится утверждение с безусловной сложностью и произвольной вычислимой функцией f с расходящимся рядом. Этот результат (также без доказательства и со ссылкой на Мартин-Лёфа) приведён и в обзоре [85]. [Как его доказать, неясно.]

Отметим также, что утверждение теоремы 87 приведено в статье [47] в немного другом варианте:

118 Докажите, что если последовательность ω случайна (в смысле Мартин-Лёфа по равномерной мере), а функция f вычислима и *ряд* $\sum 2^{-f(n)}$ *вычислимо сходится*, то $KS((\omega)_n|n) \geqslant n - f(n)$ для всех n, кроме конечного числа. [Указание. Если ряд вычислимо сходится, а неравенство нарушается бесконечно много раз, то хвосты ряда можно использовать для получения покрытий сколь угодно малой меры.]

А что будет, если требовать большой сложности начального отрезка не для всех (достаточно длинных) отрезков, а для бесконечного их числа? В той же статье Мартин-Лёфа приводятся следующие результаты:

119 Докажите, что для почти всех (по равномерной мере) последовательностей ω найдётся c, при котором $KS((\omega)_n|n) \geqslant n - c$ для бесконечно многих n.

[Указание: если это не так, то для всякого c найдётся N, начиная с которого начальные отрезки длины n имеют сложность меньше $n - c$. Для данных c и N множество таких последовательностей имеет меру меньше 2^{-c}; с ростом N множество растёт и объединение по N имеет меру не больше 2^{-c} по непрерывности.]

120 Если для последовательности ω найдётся такая константа c, что $KS((\omega)_n|n) \geqslant n - c$ при бесконечно многих n, то последовательность ω случайна в смысле Мартин-Лёфа (по равномерной мере). [Указание. Если ω покрыта одним из интервалов с общей мерой меньше 2^{-c}, то достаточно длинные её начальные отрезки (при известной длине) можно описать порядковым номером среди слов данной длины, попадающих в один из этих интервалов, для чего достаточно $2 \log c + n - c$ битов.]

$\boxed{121}$ Покажите, что утверждение предыдущей задачи можно усилить, заменив условную сложность на безусловную $KS((\omega)_n)$

[Указание: можно воспользоваться задачей 6 или, что ещё проще, задачей 39.]

Таким образом, получается некоторое множество полной меры, содержащееся в множестве всех случайных по Мартин-Лёфу последовательностей. (Если бы дополнение этого множества было не просто нулевым, но и эффективно нулевым, то мы получили бы критерий случайности по Мартин-Лёфу — но это не так.)

Сравнительно недавно выяснилось, что это за множество — оказывается, это множество всех последовательностей, случайных в смысле Мартин-Лёфа с оракулом $\mathbf{0}'$. Такие последовательности иногда называют «2-случайными» (а случайные по Мартин-Лёфу в обычном смысле, без оракула, называют «1-случайными»). См. [51,62].

5.7. Случайное число Ω

[chaitin-omega]

Интересное применение критерия случайности составляет следующая теорема. Пусть m — максимальная перечислимая снизу полумера на множестве натуральных чисел (например, $m(x) = 2^{-KP(x)}$ или распределение на выходе универсального вероятностного алгоритма, см. главу 4). Г. Чейтин предложил рассмотреть число

$$\Omega = \sum_n m(n)$$

(которое можно назвать общей вероятностью остановки универсального вероятностного алгоритма или суммой наименее сходящегося перечислимого снизу ряда) и сделал такое интересное наблюдение:

Theorem 92. [omega-is-random] *Двоичная запись числа Ω является случайной по Мартин-Лёфу последовательностью относительно равномерной меры.*

Заметим, что значение Ω зависит от выбора конкретной максимальной перечислимой снизу полумеры, но утверждение теоремы остаётся верным при любом таком выборе.

$\lhd$ Пусть нам известны первые n двоичных знаков числа Ω. Они образуют число Ω_n, которое есть приближение снизу к Ω с погрешностью не более 2^{-n} Будем перечислять снизу приближения к $m(0), m(1), \ldots$ параллельно, постепенно добавляя всё новые $m(i)$, пока найденные числа в сумме не дадут больше $\Omega_n - 2^{-n}$. (Такое рано или поздно случится, так как сумма ряда равна Ω и строго больше выбранной нами границы.) Составим список всех чисел i, которые встречаются в этой сумме (с ненулевой нижней оценкой для $m(i)$).

Заметим, что в этом списке встречаются все числа i, у которых $m(i) \geqslant 2 \cdot 2^{-n}$ (поскольку если бы такое число было пропущено, то погрешность была бы больше 2^{-n}), и потому все i с $KP(i) < n - c$ (для некоторого c, зависящего от выбора функции m, но не от n). Поэтому минимальное число, не входящее в этот список, имеет сложность не меньше $n - c$. Значит, и сам этот список, и число Ω_n, по которому этот список построен, имеют сложность не меньше $n - c'$ для некоторого другого c' и для всех n. Остаётся воспользоваться критерием случайности с префиксной сложностью (см. теоремы 84 и 86). $\rhd$

Можно было бы сразу определять случайность действительных чисел в смысле Мартин-Лёфа (относительно обычной меры на прямой), если в определении эффективно нулевого множества требовать наличия алгоритма, который по каждому рациональному $\varepsilon > 0$ выдаёт покрытие интервалами с рациональными концами и суммарной мерой не более ε.

$\boxed{122}$ Покажите, что определённая таким образом случайность равносильна случайности последовательности знаков в двоичном разложении.

$\boxed{123}$ Покажите, что квадрат (синус, экспонента) случайного действительного числа случаен. [Указание. Прообраз множества меры нуль имеет меру нуль, и это можно эффективизировать.]

$\boxed{124}$ Всегда ли сумма двух случайных действительных чисел случайна? [Указание: они могут быть «зависимы».]

Случайное число Ω (точнее, любое число такого вида, для любой максимальной перечислимой снизу полумеры) обладает разными особыми свойствами, выделяющими его среди других случайных чисел. Прежде всего, оно перечислимо снизу (заметим, что множество перечислимых снизу чисел, очевидно, счётно и тем самым имеет меру нуль). Из этого вытекают и другие любопытные свойства, например, такое:

$\boxed{125}$ Покажите, что если α — любое перечислимое снизу действительное число, то $\alpha + \Omega$ — случайно.

$\boxed{126}$ Покажите, что любое действительное число есть сумма двух случайных. [Указание: достаточно знать, что случайные числа образуют множество полной меры.]

[Здесь можно было бы написать про сводимость действительных чисел по Соловею, её максимальный элемент и т.п. — если разобраться.]

[Ω как оракул позволяет разрешать любое перечислимое множество — а что ещё можно про него сказать? Как связаны числа Ω для двух разных оптимальных способов описания? сводятся друг к другу?]

Число Ω можно рассматривать как бесконечный аналог объектов сложности n, рассмотренных в теореме 15 (с. 32). Между ними есть и более формальная связь.

Theorem 93. [chaitin-finite-infinite] *Пусть Ω_n — двоичное слово, составленное из первых n битов двоичной записи числа Ω. Тогда Ω_n обладает свойствами, указанными в теореме 15: любой из указанных в этой теореме объектов может быть алгоритмически получен из Ω_n при известном n и наоборот (в обоих случаях — с поправкой $O(\log n)$ в значении n).*

$\triangleleft$ Покажем, что по Ω_n можно построить число T, большее $B(n - O(\log n))$. В самом деле, рассмотрим процесс пошагового получения приближений снизу к Ω и оборвём его, когда приближения достигнут Ω_n (то есть первые n битов приближения получат окончательные значения). Пусть это случится на шаге T; число T можно алгоритмически найти по Ω_n. Пусть $t > T$; покажем, что число t имеет сложность более $n - O(\log n)$. В самом деле, зная t и n, можно сделать t шагов приближения и найти Ω_n, которое имеет префиксную сложность не меньше $n - O(1)$ (теорема 92), а значит, обычную сложность не меньше $n - O(\log n)$.

Теперь покажем, что верно и обратное: зная $B(n)$ и n, можно найти $\Omega_{n-O(\log n)}$. В самом деле, сделаем $B(n)$ шагов приближения снизу к Ω. Надо проверить, что в полученном приближении снизу по крайней мере $n - O(\log n)$ знаков будут «окончательными» (то есть будут совпадать с соответствующими знаками числа Ω). Если это не так, то существует конечная двоичная дробь β из $2^{n-O(\log n)}$ знаков, разделяющая достигнутое приближение и окончательное значение Ω. Эта дробь имеет сложность $n - O(\log n)$; с другой стороны, зная β, мы можем найти число, большее $B(n)$ — достаточно подсчитать число шагов, после которых приближение к Ω впервые пересекает границу β. При достаточно большой константе в $O(\log n)$ эти два обстоятельства противоречат друг другу. $\triangleright$

Отсюда видно, что знание $n + O(\log n)$ битов в числе ω позволяет ответить на любой вопрос о завершении работы программы длины не более n. Поскольку любой перечислимый вопрос, например, вопрос о выводимости формул длины n, может быть представлен в таком виде, то — при достаточном художественном воображении — можно, следуя Чейтину, назвать число Ω «числом мудрости», в котором содержится информация обо всём на свете.

5.8. Эффективная размерность Хаусдорфа

[monothaus]

В классической теории меры хорошо известно понятие *размерности Хаусдорфа* (часто упоминаемое в связи с «фракталами»). Его можно определить так. Пусть $\alpha > 0$ — некоторое действительное число. Будем говорить, что множество A является α-нулевым, если для всякого $\varepsilon > 0$ его можно покрыть счётным семейством интервалов I_k с

$$\sum_k \mu(I_k)^\alpha < \varepsilon.$$

Это определение предполагает, что A является подмножеством некоторого пространства, в котором выделен класс «интервалов» с определённой на них мерой μ. Мы ограничимся случаем $A \subset \Omega$, считая, что интервалами являются множества Ω_x, состоящие из бесконечных продолжений слова x. Мера интервала Ω_x равна $2^{-l(x)}$.

Несколько простых замечаний:

(1) Подмножество α-нулевого множества является α-нулевым множеством.

(2) При $\alpha = 1$ мы получаем обычное определение множества меры нуль.

(3) При $\alpha > 1$ любое подмножество $A \subset \Omega$ является α-нулевым (его можно покрыть 2^n интервалами, соответствующими 2^n словам длины n, и сумма α-степеней их мер стремится к нулю при $n \to \infty$).

(4) Если $0 < \alpha < \alpha'$, то любое α-нулевое множество является и α'-нулевым (мера $\mu(I)$ интервала I не превосходит 1, и потому $\mu(I)^\alpha > \mu(I)^{\alpha'}$).

127 Предложите естественное определение α-нулевого множества для подмножеств прямой и проверьте, что множество $A \subset [0,1]$ является α-нулевым тогда и только тогда, когда множество двоичных записей всех чисел из A является α-нулевым согласно определению. [Указание: по существу надо проверить, что разрешение рассматривать произвольные промежутки на прямой, а не только половины, четверти и т.п., не меняет класса нулевых множеств.]

Из наших замечаний следует, что для любого множества $A \subset \Omega$ есть некоторая граница $d \in [0,1]$ с таким свойством: при $\alpha > d$ множество A является α-нулевым, а при $\alpha <$

d — нет. (При $\alpha = d$ множество может быть нулевым, а может и не быть.) Эта граница называется *хаусдорфовой размерностью* множества A.

$\boxed{\textbf{128}}$ *Канторово множество* на отрезке получается, если выбросить из него среднюю треть $(1/3, 2/3)$, у каждого из двух полученных отрезков выбросить среднюю треть и так далее. Докажите, что получится компактное множество, гомеоморфное Ω и имеющее хаусдорфову размерность $\log_3 2$.

[Указание. Верхняя оценка для размерности получается, если в качестве покрытия использовать все отрезки, оставшиеся к некоторому шагу — точнее, немного бо́льшие их интервалы. Чтобы получить нижнюю оценку, заметим, что (1) в силу компактности можно рассматривать конечные покрытия интервалами; (2) от интервалов можно перейти к отрезкам; (3) если отрезок пересекается с выброшенным на некотором шаге интервалом (средней третью), но целиком его не содержит, то отрезок можно уменьшить, отрезав край; 4) если отрезок целиком содержит некоторый выброшенный интервал, то этот отрезок можно заменить на тот, средней третью которого интервал был; (5) если остались только «стандартные» отрезки, то они образуют покрытие двоичного дерева, откуда получается оценка на сумму степеней их мер.]

$\boxed{\textbf{129}}$ Предложите естественное определение размерности множеств в пространстве ($\mathbb{R}^3$) по Хаусдорфу. Объясните, почему мера тел равна 3, поверхностей — 2, линий — 1, а точек — 0. Покажите, что в пространстве существует множество любой размерности от 0 до 3.

Теперь понятно, каким будет эффективный аналог понятия хаусдорфовой размерности (см. [74,63]). Будем говорить, что множество $A \subset \Omega$ является *эффективно α-нулевым* (для данного $\alpha > 0$), если существует алгоритм, который по рациональному $\varepsilon > 0$ указывает покрытие множества A интервалами, у которых сумма α-х степеней мер не превосходит ε.

Как и в классическом случае, при увеличении α (и при уменьшении A) это свойство сохраняется. Принципиальная разница с классическим определением (как и для нулевых множеств) возникает дальше:

Theorem 94. [effective-hausdorf] *При любом рациональном $\alpha > 0$ существует наибольшее по включению эффективно α-нулевое множество.*

◁ Доказательство этой теоремы повторяет аналогичное рассуждение для эффективно нулевых множеств в главе 3. Счётное объединение α-нулевых множеств (классических) является α-нулевым; аналогичным образом объединение перечислимого семейства эффективно α-нулевых множеств является эффективно α-нулевым. С другой стороны, при рациональном (и даже при любом вычислимом) α можно эффективно перечислить все эффективно α-нулевые множества (точнее, алгоритмы, порождающие необходимые покрытия) — надо перечислять все, при этом корректируя излишние меры при необходимости. ▷

Следующий результат (А. Ходырев) не используется в дальнейшем (для определения эффективной хаусдорфовой размерности достаточно ограничиться рациональными α), но любопытен сам по себе. Пусть α — произвольное действительное число.

Theorem 95. *Наибольшее эффективно α-нулевое множество существует тогда и только тогда, когда α перечислимо снизу.*

◁ Пусть α перечислимо снизу. Это значит, что постепенно мы получаем всё более и более точные приближения к α с недостатком. Тем самым требования к алгоритму (сумма α-степеней мер любого конечного числа порождённых им интервалов меньше ε) постепенно ослабевают, и мы можем «снять с полки» (как говорили про фильмы во время перестройки) и пропустить на выход забракованные ранее, а ныне признанные безопасными интервалы. Легко понять, что если алгоритм удовлетворяет требованиям для предельного α, то рано или поздно любой порождённый им интервал будет пропущен.

Напротив, пусть существует наибольшее эффективно α-нулевое множество. Рассмотрим соответствующий ему алгоритм. Его можно использовать для получения нижних оценок на α. В самом деле, если этот алгоритм при некотором ε выдал интервалы, для которых сумма β-степеней мер больше ε, то это означает только одно: выбранное нами β меньше α. Пробуя разные β, разные ε и разные конечные наборы интервалов, мы получим перечислимое множество оценок снизу для α.

Остаётся лишь показать, что эти оценки могут сколь угодно близко подходить к α. Предположим, что это не так и что они меньше некоторого $\alpha' < \alpha$. В этом случае всякое эффективно α-нулевое множество будет и эффективно α'-нулевым, чего не бывает, как мы увидим дальше (существуют множества любой эффективной размерности, см. задачу 130, p. 152). ▷

Теперь естественно дать такое определение. *Эффективной хаусдорфовой размерностью* множества $A \subset \Omega$ называется точная нижняя грань всех α, при которых A является эффективно α-нулевым. Это число находится в промежутке $[0, 1]$ и не меньше (классической) хаудорфовой размерности. (Первоначально определение эффективной хаусдорфовой размерности было предложено в терминах вычислимых мартингалов, см. [43,44], где установлены приводимые ниже свойства размерности; об этом варианте определения см. далее в разделе 9.10.)

В своё время мы говорили, что свойство множества быть эффективно нулевым парадоксальным образом зависит не от того, много ли в нём элементов, а от того, какие это элементы — случайные или нет. Подобным образом обстоит дело и с эффективной размерностью.

Theorem 96. [dimension-sup] *Эффективная хаусдорфова размерность множества равна точной верхней грани эффективных хаусдорфовых размерностей его элементов.*

(Говоря об эффективной хаусдорфовой размерности точки $\omega \in \Omega$, мы имеем в виду размерность синглетона $\{\omega\}$.)

◁ Очевидно, размерность множества не меньше размерности любого его подмножества. Остаётся проверить, что если размерности всех синглетонов, образованных элементами некоторого множества A, меньше некоторого рационального r, а r' больше r (и тоже рационально), то размерность множества в целом не превосходит r'. А это сразу следует из теоремы 94: все синглетоны лежат в наибольшем эффективно r'-нулевом множестве, поэтому A является его подмножеством и имеет размерность не больше r'. ▷

Как же найти эффективную размерность синглетона? Оказывается, она имеет следующее простое описание в терминах колмогоровской сложности.

Theorem 97. `[hdim-formula]` *Эффективная хаусдорфова размерность множества* $\{\omega\}$*, где* $\omega = \omega_0\omega_1\omega_2\ldots$*, равна*

$$\liminf_{n\to\infty} \frac{KS(\omega_0\omega_1\ldots\omega_{n-1})}{n}.$$

(В формулировке теоремы говорится о простой колмогоровской сложности начальных отрезков. Однако это не принципиально: поскольку разница между различными вариантами сложности имеет порядок $O(\log n)$ для слов длины n, при делении на n и переходе к пределу она исчезает.)

$\lhd$ Надо доказать неравенство в две стороны.

Предположим, что нижний предел меньше некоторого рационального r. Надо проверить, что множество $\{\omega\}$ является эффективно r'-нулевым для любого рационального $r' > r$.

Рассмотрим для каждого n все слова длины n, имеющие сложность меньше rn. Таких слов не больше $O(2^{rn})$. (Условие про нижний предел гарантирует, что при бесконечно многих n начальный отрезок последовательности ω оказывается в числе таких слов.) Рассмотрим все интервалы Ω_z, порождённые этими словами z. Мера каждого из Ω_z в степени r' есть $2^{-r'n}$, и суммарная r'-мера равна $2^{(r-r')n}$, что с ростом n убывает как геометрическая прогрессия. Поэтому ряд

$$\sum_n 2^{(r-r')n}$$

сходится, и если отбросить некоторый его начальный кусок (рассматривать слова длины N и более), то суммарная r'-мера этих слов может быть сделана сколь угодно малой (выбором достаточно большого N). Между тем по предположению о нижнем пределе остаток будет по-прежнему покрывать ω. В одну сторону неравенство доказано.

Напротив, пусть $\{\omega\}$ имеет эффективную размерность меньше r для некоторого рационального r. Покажем, что нижний предел, о котором идёт речь в теореме, не превосходит r.

По определению для каждого рационального $\varepsilon > 0$ можно эффективно перечислять последовательность интервалов, про которую известно, что один из них содержит ω и сумма ряда из r-степеней длин не больше ε. Сделаем это для $\varepsilon = 1, 1/2, 1/4, \ldots$ и получим последовательность интервалов, у которых сумма r-степеней длин ограничена и которые покрывают ω бесконечно много раз. Другими словами, мы нашли вычислимую последовательность слов $x_0, x_1, \ldots$ с такими свойствами:

- $\sum 2^{-rl(x_i)} < \infty$;
- x_i является началом ω при бесконечно многих i.

Из первого условия следует, что $m(i) \geqslant c2^{-rl(x_i)}$ при некотором c и всех i (здесь m — наибольшая полумера на натуральных числах, рассмотренная в главе 4). Переходя к логарифмам, получаем оценку для префиксной сложности

$$KP(x_i) \leqslant rl(i) + O(1).$$

Остаётся заметить, что длины x_i стремятся к бесконечности (поскольку ряд сходится), что среди x_i имеется бесконечно много начал последовательности ω и что простая сложность не превосходит префиксной. (А также вспомнить определение нижнего предела и заметить, что если последовательность бесконечно много раз не превосходит r, то её нижний предел не превосходит r.) $\rhd$

$\boxed{130}$ [any-hausdorff-dimension] Выведите из этой теоремы, что для любого действительного $\alpha \in [0,1]$ существует множество (и даже одноэлементное множество) с эффективной размерностью α. [Указание: сложность начальных отрезков можно увеличивать, добавляя случайные биты, и уменьшать, добавляя нулевые биты.]

$\boxed{131}$ Покажите, что для всякого $\alpha \in [0,1]$ существует множество с (классической) хаусдорфовой размерностью α. [Указание. Можно взять множество последовательностей, у которых на некоторых местах стоят нули.]

$\boxed{132}$ Докажите, что определение эффективной хаусдорфовой размерности не изменилось бы, если вместо наличия для каждого ε покрытия с суммой степеней меньше ε мы бы требовали наличия одного покрытия с конечной суммой степеней, но покрывающего каждый элемент множества бесконечно много раз.

Мы вернёмся к понятию эффективной размерности, когда будем говорить о связи случайности с эффективными мартингалами (раздел 9.5 и далее). Там же будет проведено доказательство теоремы 97 (по существу то же самое) на языке мартингалов.

5.9. Дефект случайности для априорной сложности

[monotdef]

Критерий случайности (в смысле Мартин-Лёфа) по вычислимой мере P можно переформулировать следующим образом. Для каждого слова x рассмотрим величину

$$d_P(x) = -\log_2 P(\Omega_x) - \mathit{KA}\,(x).$$

Назовём её *дефектом случайности* двоичного слова x относительно вычислимой меры P. Объяснение этого названия: последовательность ω случайна (в смысле Мартин-Лёфа) тогда и только тогда, когда дефекты случайности её начальных отрезков ограничены.

Слова «дефект случайности» могут употребляться в самых разных смыслах; например, в главе 16 мы будем говорить о дефекте случайности элемента внутри конечного множества. Но в этом разделе, говоря о дефекте случайности, мы имеем в виду именно функцию d_P, определённую выше.

[Хорошо бы добавить сравнение разных вариантов определений, включая определения Левина – Гача для бесконечных последовательностей, относительно классов мер и так далее. Про это можно было бы написать специальный раздел в этой главе.]

Приведённое определение предполагает, что $P(\Omega_x) > 0$; если $P(\Omega_x) = 0$, то дефект естественно считать бесконечным.

Дефект случайности всегда неотрицателен (с точностью до константы, см. теорему 81). Критерий случайности (теоремы 83 и 85) говорит, что для случайных в смысле Мартин-Лёфа последовательностей дефект их начальных отрезков ограничен, а для неслучайных — стремится к бесконечности. В частности, не бывает последовательностей, у которых определённый таким образом дефект неограничен, но не стремится к бесконечности. Почему так происходит? На этот вопрос даёт ответ следующая теорема.

Theorem 98. [monotdef-conservation] *Пусть фиксирована вычислимая мера P. Тогда существует такая константа c, что для любого слова x и для любого его продолжения y*

выполняется неравенство:

$$d_P(y) \geqslant d_P(x) - 2\log d_P(x) - c.$$

Смысл этой теоремы: любое продолжение конечной последовательности с большим дефектом имеет (почти столь же) большой дефект. Или: любое начало (конечной) последовательности с малым дефектом имеет (почти столь же) малый дефект.

$\lhd$ Для любого k рассмотрим перечислимое множество всех конечных последовательностей, у которых дефект больше k. Все их бесконечные продолжения образуют открытое множество S_k, у которого P-мера не больше 2^{-k}. Рассмотрим теперь меру P_k на Ω, которая равна нулю вне S_k и в 2^k раз превосходит P внутри S_k. (Формально говоря, мера множества U равна $2^k P(U \cap S_k)$). Отметим, что P_k не является вычислимой мерой в нашем смысле слова, поскольку мера всего пространства не равна 1 (и даже не обязана быть вычислимой, а всего лишь перечислима снизу). Но P_k вполне можно считать перечислимой снизу полумерой (оставшуюся до 1 вероятность считаем мерой пустого слова).

Рассмотрим теперь сумму

$$\sum_k \frac{1}{2k^2} P_k,$$

которая задаёт некоторую перечислимую снизу полумеру S (двойка в знаменателе добавлена, чтобы сумма ряда $\sum 1/2k^2$ была меньше единицы; меру пустого слова надо вновь увеличить). При этом

$$-\log S(x) \leqslant -\log P(x) - k + 2\log k + O(1)$$

для всех слов x, являющихся продолжениями какого-либо слова с дефектом больше k. Поскольку S не больше априорной вероятности на дереве (с точностью до $O(1)$-множителя), получаем требуемое утверждение.

В этом рассуждении мы предполагали, что дефект x конечен, то есть что $P(\Omega_x) \neq 0$. Но если это не так, то для любого продолжения y мера $P(\Omega_y)$ также нулевая, и дефект бесконечен. $\rhd$

Отметим ещё одно свойство дефекта, непосредственно следующее из его определения:

$\boxed{133}$ Докажите, что для любого слова x хотя бы одно из слов $x0$ и $x1$ имеет не больший дефект случайности, чем само x. (Вычислимая мера, относительно которой вычисляется дефект, фиксирована.)

Согласно этой задаче, мы можем начать с любого слова и добавлять к нему бит за битом, не увеличивая дефекта. В результате, согласно критерию случайности, получится случайная в смысле Мартин-Лёфа последовательность (по выбранной заранее мере, относительно которой измеряется дефект)

Покажем теперь, как понятие дефекта может быть использовано для сравнения классов случайных (в смысле Мартин-Лёфа) последовательностей относительно разных мер.

Пусть P — вычислимое распределение вероятностей на Ω, а $f\colon \Sigma \to \Sigma$ — непрерывное вычислимое отображение. Рассмотрим образ меры P при этом отображении, то есть меру Q на пространстве Σ, для которой

$$Q(U) = P(f^{-1}(U))$$

при $U \subset \Sigma$. Другими словами, Q — это распределение случайной величины $f(\omega)$, если ω — случайная величина с распределением P. Вообще говоря, Q не обязана быть сосредоточена на бесконечных последовательностях, то есть в нашей терминологии может быть не мерой, а полумерой. Предположим, однако, что это не так и что Q является мерой. (В этом случае, как легко видеть, Q будет вычислимой мерой на Ω.)

Theorem 99. `[random-image]` (а) *Если ω — случайная по мере P последовательность, то её образ $f(w)$ будет бесконечной последовательностью, случайной по мере Q.*

(б) *Всякая случайная по мере Q последовательность может быть получена таким образом (равна $f(\omega)$ для некоторой случайной по мере P последовательности ω).*

(Говоря о случайности в этой теореме, мы имеем в виду случайность в смысле Мартин-Лёфа.)

Смысл этой теоремы можно объяснить так. Пусть у нас имеется вероятностная машина, состоящая из датчика случайных битов и алгоритма, эти биты использующего для получения выходной последовательности. При этом биты, выдаваемые датчиком, имеют распределение P, а алгоритм преобразования задаёт отображение $f\colon \Sigma \to \Sigma$.

Какие последовательности могут получиться на выходе такого устройства? Точнее, про какие последовательности мы готовы поверить, что они случайно получились на выходе? Это будут f-образы последовательностей, которые могут получиться на выходе датчика, то есть (если следовать схеме Мартин-Лёфа) образы случайных по мере P последовательностей. С другой стороны, всю эту машину (датчик плюс алгоритм) можно рассматривать как единый источник битов, распределённых по мере Q; при таком подходе на её выходе «могут получиться» случайные по мере Q последовательности. Так вот, наша теорема говорит, что это одно и то же.

$\lhd$ Прежде всего докажем, что образ P-случайной последовательности не может быть никаким конечным словом z. В самом деле, рассмотрим бесконечные последовательности, входящие в прообраз этого конечного слова, то есть в прообраз $\Sigma_z \setminus (\Sigma_{z0} \cup \Sigma_{z1})$. Прообраз Σ_z есть эффективно открытое множество (объединение перечислимого семейства интервалов), прообраз $\Sigma_{z0} \cup \Sigma_{z1}$ — также эффективно открытое множество, являющееся подмножеством первого. Нам надо доказать, что прообраз разности (то есть разность прообразов) не содержит случайных последовательностей, то есть является эффективно нулевым относительно меры P множеством. Это следует из такой общей леммы:

Лемма. Пусть P — вычислимая мера на Ω, а $U \subset V$ — два эффективно открытых множества, причём $P(V \setminus U) = 0$. Тогда $V \setminus U$ является эффективно нулевым множеством (не содержит случайных последовательностей).

Доказательство леммы. Очевидно, достаточно рассмотреть один интервал I множества V (ограничив U этим интервалом). По мере перечисления интервалов множества U обнаруживается всё бо́льшая часть интервала I, входящая в U. По непрерывности меры мера обнаруженной части стремится к мере интервала I, и рано или поздно мы дождёмся момента, когда разность по мере будет меньше ε, какое бы $\varepsilon > 0$ нам бы ни задали. Лемма доказана.

Чтобы завершить доказательство пункта (а), осталось показать, что образ P-случайной последовательности ω не может быть Q-неслучайным, то есть содержаться в эффективно Q-нулевом множестве. Но если бы $f(\omega)$ можно было покрыть интервалами сколь угодно малой меры, то беря f-прообразы этих интервалов, мы бы получили покрытие последовательности ω сколь угодно малой меры (прообраз эффективно открытого множества открыт и по определению имеет ту же меру). Утверждение (а) доказано.

Докажем теперь утверждение (б), используя понятие дефекта. Пусть последовательность τ случайна по мере Q. Это значит, что дефекты её начальных отрезков ограничены. Теперь применим следующую лемму, которую можно рассматривать как аналог утверждения (б) для конечных последовательностей:

Лемма. Пусть u — произвольное двоичное слово, для которого $Q(\Omega_u) > 0$. Тогда существует двоичное слово v, для которого $u \preccurlyeq f(v)$ (u является началом $f(v)$) и $d_P(v) \leqslant d_Q(u) + O(1)$.

(Имеется в виду, что константа в $O(1)$ не зависит от слова u, но может зависеть от выбора f, P и Q.)

Доказательство леммы. Рассмотрим прообраз множества Σ_u при отображении f. Это — эффективно открытое множество в Σ, которое мы обозначим через F_u. По построению P-мера множества F_u (сосредоточенная на бесконечных последовательностях) равна $Q(u)$. Если дефект $d_Q(u)$ мал, то эта мера не может быть сильно меньше априорной вероятности слова u.

Рассмотрим теперь априорную вероятность множества F_u, то есть вероятность того, что универсальная вероятностная машина M на выходе даст элемент из F_u, или, другими словами, вероятность того, что машина $f \circ M$ (к выходу M дополнительно применяем f) даст некоторое продолжение слова u. Сравнивая машину $f \circ M$ с универсальной, заключаем, что априорная вероятность множества F_u лишь в константу раз превышает априорную вероятность слова u, которая не более чем в $2^{d_Q(u)}$ раз превышает $Q(u)$, которое равно P-мере F_u. Таким образом мы получаем неравенство для двух мер множества F_u (априорной вероятности, которую мы обозначим через A, и меры P):

$$\frac{A(F_u)}{P(F_u)} \leqslant O(2^{d_Q(u)}).$$

Множество F_u представимо в виде объединения (пусть неперечислимого) непересекающихся интервалов, и ясно, что для одного из интервалов Σ_v выполнено аналогичное неравенство:

$$\frac{A(\Sigma_v)}{P(\Sigma_v)} \leqslant O(2^{d_Q(u)}).$$

Поскольку $\Sigma_v \subset F_u$, то $f(v) \succcurlyeq u$, а неравенство даёт оценку $d_P(v) \leqslant d_Q(u) + O(1)$, что и требовалось. Лемма доказана.

Продолжая доказательство утверждения (б) теоремы, применим доказанную лемму к начальным отрезкам $t_0, t_1, t_2, \ldots$ случайной по мере Q последовательности τ (имеющим длины $0, 1, 2, \ldots$). Их Q-дефекты ограничены, и по лемме можно найти последовательность $v_0, v_1, \ldots$ слов с ограниченными P-дефектами, для которых $f(v_i)$ является продолжением t_i. Применяя обычное рассуждение с компактностью, можно из последовательности v_i выделить либо бесконечную подпоследовательность, состоящую из равных слов, либо бесконечную подпоследовательность, сходящуюся к некоторой бесконечной последовательности ω (последнее означает, что любой начальный отрезок ω является начальным отрезком всех членов подпоследовательности, начиная с некоторого).

В первом случае последовательность τ вычислима, будучи образом конечного слова v, встречающегося среди v_i бесконечно много раз. Это не противоречит случайности, если эта вычислимая последовательность имеет (как одноэлементное множество) положительную Q-меру, но в этом случае в качестве ω можно взять любое бесконечное P-случайное продолжение слова v (оно существует, так как $P(\Omega_x) > 0$).

Осталось рассмотреть второй случай, когда бесконечная подпоследовательность последовательности v_i сходится к некоторой последовательности ω. Заметим, что в этом случае:

(1) любой начальный отрезок ω является начальным отрезком одного из слов v_i, а их P-дефекты ограничены, поэтому по теореме 98 и P-дефекты начальных отрезков ω ограничены, и потому ω является P-случайной;

(2) по ранее доказанному в пункте (а) последовательность $f(\omega)$ бесконечна;

(3) последовательность $f(\omega)$ не может иметь начальных отрезков, не являющихся начальными отрезками τ, поскольку в этом случае у ω был бы начальный отрезок, образ которого несовместим с τ, и этот отрезок был бы (в силу сходимости) начальным отрезком сколь угодно далёких v_i, образы которых имеют сколь угодно длинное общее начало с τ.

Полученное противоречие завершает доказательство пункта (б) теоремы. $\triangleright$

134 Дайте прямое доказательство утверждения (б), используя лишь определение эффективно нулевого множества.

[Указание. Для данного ε рассмотрим покрытие Z_ε наибольшего эффективно P-нулевого множества интервалами суммарной P-меры меньше ε; пусть F — замкнутое множество бесконечных последовательностей, оказавшихся непокрытыми. Все последовательности из F случайны, поэтому нам достаточно (для произвольного ε) эффективно покрывать дополнение к образу F интервалами с малой суммой Q-мер.

Будем считать слово x "особым", если покрытие Z_ε вместе с f-прообразами всех слов, несравнимых с x, покрывает всё Ω. Множество особых слов перечислимо (всякое покрытие имеет конечное подпокрытие в силу компактности Ω, что позволяет обнаруживать особые слова). Если последовательность $f(\omega)$ имеет особое начало x, то ω попадает в Z_ε, поэтому особые слова образуют покрытие суммарной Q-меры менее ε. С другой стороны, всякая последовательность τ, все начала которой неособые, является образом последовательности из F. В самом деле, раз для каждого k начало τ длины k не является особым, то есть некоторая последовательность ω_k, которая лежит в F и имеет образ (бесконечный, раз $\omega_k \in F$), совпадающий с τ в первых k битах. Выберем из $\omega_1, \omega_2, \ldots$ сходящуюся подпоследовательность; её предел ω также лежит в F (поскольку F замкнуто), поэтому $f(\omega)$ бесконечно и по непрерывности совпадает с τ.]

135 Докажите утверждение, которое можно рассматривать как конечный аналог пункта (а) теоремы: если u и v — два двоичных слова, причём $u \preccurlyeq f(v)$, то

$$d_Q(u) \leqslant d_P(v) + 2\log d_P(v) + O(1).$$

[Указание. Последовательности с большим Q-дефектом покрываются множеством малой Q-меры, поэтому их прообразы покрываются множеством малой P-меры и потому имеют большой P-дефект. Отметим, что это утверждение является обобщением теоремы 98.]

Доказанная теорема имеет довольно неожиданные следствия. Приведём несколько примеров.

136 Пусть последовательность ω — случайна в смысле Мартин-Лёфа относительно бернуллиевой меры (все испытания независимы) с вероятностью единицы $1/3$. Докажите, что существует последовательность ω', случайная относительно равномерной меры, которая получается из ω заменой некоторых нулей на единицы. [Указание. Рассмотрим случайную последовательность независимых случайных точек, равномерно распределённых на $[0,1]$, точнее, последовательность случайных битов, записанных в двумерную таблицу, строки которой — бесконечные двоичные дроби. Заменим точки, большие $2/3$, на единицы, а остальные на нули. Доказанная теорема утверждает, что получится последовательность, случайная по бернуллиевой мере с вероятностью единицы $1/3$, и что всякая случайная по этой мере последовательность может быть получена таким образом. Аналогично для границы $1/2$, откуда и следует утверждение задачи.]

137 Рассмотрим вычислимое распределение на парах последовательностей (то есть на $\Omega \times \Omega$) и соответствующее понятие случайной (в смысле Мартин-Лёфа) пары последовательностей относительно этого распределения. Докажите, что первый член случайной пары последовательностей будет случаен в смысле Мартин-Лёфа относительно распределения, являющегося проекцией исходного на первую координату, и что таким образом получатся все случайные последовательности (в смысле Мартин-Лёфа относительно указанного распределения).

В случае, когда компоненты пары независимы (распределение на парах является произведением двух распределений на компонентах), можно доказать более сильное утверждение, известное как теорема Ламбальгена [28]:

[А где именно в диссертации Ламбальгена это написано? там много всего, но есть ли это в явном виде???]

Пусть P и Q — два вычислимых распределения на пространстве Ω. Рассмотрим произведение этих распределений; получится вычислимое распределение на $\Omega \times \Omega$ (которое изоморфно Ω, и определения случайности на него очевидно переносятся).

Theorem 100. *Пара последовательностей $\langle \xi, \eta \rangle$ случайна по Мартин-Лёфу относительно распределения $P \times Q$ тогда и только тогда, когда выполнены два условия:*

(1) *последовательность ξ случайна в смысле Мартин-Лёфа по мере P;*

(2) *последовательность η случайна в смысле Мартин-Лёфа с оракулом ξ по мере Q.*

Говоря о случайности с оракулом, мы имеем в виду, что алгоритм, перечисляющий покрытие, теперь имеет ξ в качестве оракула (от этого перечислимых множеств и неслучайных последовательностей может стать больше, а случайных — меньше). Условие про оракул существенно: пара $\langle \xi, \xi \rangle$, как правило, не случайна по мере $P \times P$, даже если ξ случайна по мере P.

Можно ещё отметить, что ξ и η входят симметрично, поэтому условие (1) можно было бы усилить, разрешив η в качестве оракула. Но несимметричный вариант кажется более естественным, его можно прочесть так: «выбрать пару случайно означает случайно выбрать первый член пары, а затем, зная первый член, случайно выбрать второй».

$\triangleleft$ Докажем сначала, что для случайной пары выполнены условия (1) и (2).

(1) Если последовательность ξ не случайна и покрывается интервалами малой P-меры, то эти интервалы (умноженные на всё Ω по второй координате дают покрытие пары $\langle\xi,\eta\rangle$ той же меры (относительно $P\times Q$).

(2) Пусть для всякого ε можно перечислять (с оракулом ξ) семейство интервалов малой Q-меры, покрывающее η. Этот же процесс можно запустить и с любым другим оракулом, и он будет порождать какие-то интервалы, используя конечную информацию относительно оракула.

Таким образом мы получим (по данному ε) некоторое перечислимое (без оракула) семейство прямоугольников с таким свойством: Q-мера сечения, в котором первая координата равна ξ, не больше ε. Такое семейство легко переделать в семейство прямоугольников общей меры не больше ε, если принудительно урезать прямоугольники так, чтобы в любом сечении мера была не больше ε и при этом сечения, где это неравенство и так выполнялось, не менялись. Получаем противоречие со случайностью пары $\langle\xi,\eta\rangle$.

Теперь докажем, что если пара $\langle\xi,\eta\rangle$ не случайна, то не выполнено одно из условий (1) и (2). Пусть имеется семейство прямоугольников с общей мерой не больше ε, покрывающее $\langle\xi,\eta\rangle$. Пусть U — объединение этих прямоугольников. Посмотрим, при каких значениях первой координаты x) сечение $U_x=\{y|\langle x,y\rangle\in U\}$ имеет меру больше $\sqrt{\varepsilon}$. Соответствующее подмножество (одномерное, по первой координате) имеет меру не больше $\sqrt{\varepsilon}$ и его можно представить в виде перечислимого объединения интервалов. При этом для точки $\langle\xi,\eta\rangle$, покрытой исходным семейством прямоугольников, верно одно из двух: либо ξ покрыта построенным семейством интервалов общей меры не более $\sqrt{\varepsilon}$, либо $\langle\xi,\eta\rangle$ покрыта семейством прямоугольников, которые в ξ-сечении имеют Q-меру не больше $\sqrt{\varepsilon}$. Во втором случае η покрыта ξ-перечислимым семейством интервалов суммарной Q-меры не более $\sqrt{\varepsilon}$.

Хотелось бы сделать так для каждого ε и заключить, что либо ξ не случайна, либо η не случайна с оракулом ξ. Первое гарантировано, если при всех ε имеет место первый вариант (ξ покрыта интервалами); второе — если при всех ε случается второе. Но что делать, если при некоторых ε случается одно, а при других — другое?

Тут помогает такой трюк: выполним указанное построение для $\varepsilon=2^{-2k}$ при всех $k=1,2,3,\ldots$. Тогда для каждого k получится семейство $V(k)$ интервалов по первой координате, имеющих общую P-меру не более $2^{-k}=\sqrt{2^{-2k}}$. Теперь есть две возможности: либо при бесконечно многих k это семейство покрывает ξ, либо для всех достаточно больших k это семейство не покрывает ξ.

В первом случае для каждого K объединим все семества $V(k)$ при $k\geqslant K$; получится семейство интервалов с мерой не более $2\cdot 2^{-K}$, и оно покрывает ξ уже при любом K, так что ξ не случайна.

Во втором случае есть способ (для всех k, начиная с некоторого, пусть нам и неизвестного) получить ξ-перечислимое покрытие для η, имеющее малую меру, и потому η не случайна с оракулом ξ. $\triangleright$

[Хотелось бы обобщить эту теорему на случай зависимых величин, но предварительно надо построить конструктивный вариант теории условных вероятностей. Может быть, это уже кем-то сделано?]

Вернёмся к образу множества случайных последовательностей при вычислимых отображениях. Вопрос о том, какие последовательности могут (не вызывая отторжения гипотезы о случайности) появиться на выходе вероятностной машины, имеет смысл и в общем случае, без предположения о том, что вероятность появления конечной последовательности равна нулю. Пусть, как и прежде, имеется вычислимое распределение вероятностей P на множестве Ω, а также вычислимое непрерывное отображение $f\colon \Sigma \to \Sigma$. Тогда возникает распределение вероятностей на Σ, являющееся образом P при отображении f. Мы, однако, не предполагаем, что оно сосредоточено на бесконечных последовательностях, так что получается перечислимая снизу полумера Q, вообще говоря, не являющаяся мерой.

С другой стороны, мы можем рассмотреть образы P-случайных последовательностей при отображении f. Как связано это множество с полумерой Q? Можно ли утверждать, что оно определяется полумерой Q (как это было для случая мер, когда оно состояло из Q-случайных последовательностей)? Та же аналогия подсказывает и другую возможную гипотезу: образы P-случайных последовательностей — это те и только те последовательности, у начальных отрезков которых отношение априорной вероятности и Q ограничено.

Мы не знаем, верны ли эти предположения. Можно заметить лишь, что из последнего утверждения следовала бы следующая теорема (см. [19]):

Theorem 101. [gacs-reducibility-theorem] *Пусть α — произвольная последовательность нулей и единиц. Тогда существует случайная (в смысле Мартин-Лёфа) по равномерной мере последовательность ω, а также вычислимое отображение $f\colon \Sigma \to \Sigma$, для которого $f(\omega) = \alpha$.*

Для знакомых с теорией вычислимых функций можно было бы сказать короче: всякая последовательность нулей и единиц сводится по Тьюрингу к некоторой случайной по равномерной мере (мы уже упоминали этот результат на с. 121).

Это утверждение следует из сформулированной ранее гипотезы: если в качестве f взять универсальную вероятностную машину, то полумера Q будет (с точностью до константы) априорной вероятностью, и отношение априорной вероятности к Q будет ограничено для всех последовательностей α. К сожалению, не удаётся перенести на этот случай использованное ранее рассуждение (мы можем найти прообразы у t_i с малым дефектом и даже взять предельную точку, но образ этой предельной точки теперь может оказаться конечным). Поэтому придётся воспользоваться другой конструкцией.

$\lhd$ Будем доказывать несколько более сильное утверждение и построим вычислимое непрерывное отображение f (одно для всех α), для которого образ $f(R)$ множества R случайных относительно равномерной меры последовательностей покрывает всё Ω.

Более того, мы для любого эффективно открытого множества U (объединения перечислимого семейства интервалов) с достаточно малой мерой построим вычислимое отображение f, для которого образ $f(\Omega \setminus U)$ покрывает всё Ω. Поскольку в качестве U можно взять покрытие максимального эффективно нулевого множества, отсюда следует требуемое утверждение.

Будем действовать следующим образом. Отображение f мы строим постепенно, следя за тем, чтобы образ дополнения к U (точнее, дополнения к текущему состоянию множества U) содержал все последовательности. Вначале, пока текущее U ещё пусто, задача тривиальна — если не думать о дальнейшем, то можно было бы даже в качестве f взять тождественное отображение. Но появляющиеся интервалы множества U «выбивают» часть последовательностей, и потому нужен резерв для замены. Такой резерв несложно организовать, поскольку по предположению мера U мала и большая часть последовательностей окажется незатронутой, но тут есть деликатный момент. Нам мало, чтобы у данной последовательности был прообраз на каждом шаге построения (ещё не вычеркнутый) — надо, чтобы прообраз был в пределе, и об этом надо специально заботиться (добиваясь постепенной стабилизации прообразов во всех позициях).

Для реализации этого (пока довольно расплывчатого) плана выберем последовательность целых чисел $k_0, k_1, k_2, \ldots$, которые будем рассматривать как длины блоков, на которые разбивается последовательность битов в прообразе. Если ограничиваться словами длин $k_0, k_0 + k_1, k_0 + k_1 + k_2$ и так далее, то от двоичного дерева остаётся дерево с ветвлением 2^{k_0} в корне, 2^{k_1} в сыновьях корня и так далее. (Формально говоря, алгоритм, реализующий преобразование f, будет читать биты целыми блоками.) Удобно также предполагать, что и выбраковка попавших в U последовательностей будет происходить на уровне блоков. Другими словами, мы представляем U как объединение интервалов Ω_x, где слово x имеет длину $k_0 + \ldots + k_i$ при некотором i. (Ясно, что этого можно достичь измельчением интервалов, не меняющим суммарную меру.)

В это дерево большого ветвления мы вложим двоичное дерево, которое и отобразим на стандартное двоичное дерево с помощью f. Для этого выберем среди 2^{k_0} сыновей корня какие-либо два, скажем, два первых в алфавитном порядке слова X_0 и X_1 и объявим, что они отображаются в 0 и 1. (Это значит, что для любой последовательности ω, начинающейся на X_0 или X_1, первый бит $f(\omega)$ равен нулю и соответственно единице. Для остальных последовательностей ω, у которых первый блок отличается от X_0 и X_1, значение $f(\omega)$ пока не определено. Далее к каждому из слов X_0 и X_1 мы приписываем два блока длины k_1 (скажем, первые в алфавитном порядке). Получаются четыре слова длины $k_0 + k_1$, которые мы обозначаем $X_{00}, X_{01}, X_{10}, X_{11}$; они будут отображаться в слова 00, 01, 10 и 11 и так далее. Формально говоря, для каждого двоичного слова u мы определяем состоящее из $l(u)$ блоков слово X_u, которое отображается в u. При этом X_u является началом X_v, если u есть начало v, и X_u несовместно с X_v, если u несовместно с v.

Так мы получаем вычислимое отображение, определённое на двоичном поддереве нашего дерева (большого ветвления) и отображающее его на всё Ω. Если множество U пусто, то на этом можно и закончить. Однако появление интервалов в множестве U (выбраковка некоторых вершин дерева) заставляет перестраивать поддерево, заменяя выбракованные вершины на другие — в каждый момент будет двоичное поддеререво, не задевающее выбракованные вершины и отображающееся на Ω.

Чтобы описать эту конструкцию подробнее, введём понятие «плохой» вершины дерева (большого ветвления). А именно, будем считать вершину x плохой в двух случаях:

• если x попадает в выброшенный интервал, то есть для одного из выброшенных (попавших в U) интервалов Ω_z слово z является началом x;

• все её сыновья кроме одного (или просто все) являются плохими.

Как обычно, такое индуктивное определение задаёт минимальное множество вершин, обладающее этими двумя свойствами. Оно гарантирует, что у всякой «хорошей» (не плохой) вершины хотя бы два сына не являются плохими, что и позволит нам вкладывать двоичное дерево в подмножество хороших вершин. Заметим, что согласно определению сыновья плохой вершины тоже плохи.

Множество плохих вершин растёт по мере обнаружения новых выброшенных интервалов. Оценим, какую часть Ω надо выбросить, чтобы корень дерева стал плохим. Чтобы корень стал плохим, все его сыновья, кроме одного, должны быть плохими, что составляет

долю
$$\left(1-\frac{1}{2^{k_0}}\right)$$
от общего числа вершин первого уровня. Для этого в свою очередь на втором уровне должно быть не менее
$$\left(1-\frac{1}{2^{k_0}}\right)\cdot\left(1-\frac{1}{2^{k_1}}\right)$$
плохих вершин (доля от общего числа вершин второго уровня). В каждый момент имеется конечное число выброшенных интервалов, и потому на достаточно высоком уровне все плохие вершины попадут в выброшенные интервалы. Мы получили такое следствие: *если суммарная мера выброшенных интервалов меньше бесконечного произведения*
$$\left(1-\frac{1}{2^{k_0}}\right)\cdot\left(1-\frac{1}{2^{k_1}}\right)\cdot\left(1-\frac{1}{2^{k_2}}\right)\cdot\ldots,$$
то в любой момент корень будет оставаться хорошей вершиной.

Таким образом становится ясным требование к числам k_i: произведение должно быть положительным, что (как известно из курса математического анализа) равносильно сходимости ряда $\sum 2^{-k_i}$. Например, можно положить $k_i = 2\log i$ (для $i \geqslant 2$).

Множество плохих вершин растёт со временем. Если при этом оно не задевает нашего двоичного поддерева, то это поддерево остаётся неизменным. Если же какая-то ветвь дерева попадает в плохую вершину, то в этом месте мы заменяем плохую вершину на её хорошего брата, и оттуда продолжаем рост поддерева. Формально говоря, перестройка дерева проводится снизу вверх с заменой плохих вершин на хорошие и с выбором сыновей у новых вершин.

Одновременно с перестройкой дерева происходит доопределение преобразования f (именно дооопределение, на устаревших вершинах значения f остаются прежними). Процесс алгоритмический, поэтому легко проверить, что f будет вычислимым.

Остаётся доказать, что при таком построении f для любой последовательности $\alpha \in \Omega$ найдётся f-прообраз, не входящий в выброшенное открытое множество U. По построению в каждый момент t найдётся прообраз ω_t, не входящий в уже построенную часть U. Кроме того, с ростом t последовательности ω_t сходятся к некоторому пределу ω (индукцией по уровню доказываем, что на этом уровне будет стабилизация, так как число возможных изменений ограничено). Надо проверить, что ω не принадлежит U и что $f(\omega) = \alpha$.

Если ω принадлежит U, то ω попадает в некоторый интервал U, который рано или поздно обнаружится. После этого момента ω_t не могут лежать в U, что противоречит сходимости.

Покажем, что $f(\omega) = \alpha$. Пусть z — произвольное конечное начало α; покажем, что $f(\omega)$ начинается на z. Пусть k — длина z. На каждом шаге имеется некоторая ветвь в дереве (большого ветвления), состоящая из k блоков и отображающаяся в z. Эта ветвь стабилизируется с ростом t, и после стабилизации в f добавляется часть, гарантирующая, что $f(\omega)$ начинается на z. $\triangleright$

138 Выведите из доказательства теоремы, что для всякой последовательности α найдётся случайная последовательность ω, к которой она сводится по Тьюрингу, причём для получения n битов последовательности α используется не более $2n \log n$ первых битов последовательности ω.

[Вроде как эту оценку можно улучшить (так учит Лоран); одна из возможностей состоит в том, чтобы и α разбивать на блоки, но можно ли достичь наилучшего известного на этом пути, неясно.]

В порядке философических спекуляций можно объяснить смысл утверждения этой теоремы так. Для любой последовательности можно представить объяснение, почему её появление не удивительно: про случайные последовательности это как бы подразумевается определением случайности, и остаётся пристроить к датчику случайных чисел машину, реализующую вычислимое отображение f.

6. Общая классификация сложностей

[class]

6.1. Сложность разрешения

[class-decision]

Начав с простой колмогоровской сложности *KS*, мы рассматривали затем также префиксную сложность *KP* и монотонную сложность *KM*. Все три определялись с помощью кратчайших описаний, но способы описаний были из разных классов. Для простой сложности это были просто вычислимые функции, для префиксной сложности — вычислимые непрерывные отображения $\Sigma \to \mathbb{N}_\perp$, для монотонной — вычислимые непрерывные отображения $\Sigma \to \Sigma$.

Для единообразия можно считать, что простая колмогоровская сложность определяется с помощью непрерывных вычислимых отображений $\mathbb{N}_\perp \to \mathbb{N}_\perp$, должным образом определив это понятие. Топология на множестве $\mathbb{N}_\perp$ и само это множество были описаны в разделе 4.4.3 (с. 86). Несложно проверить, что непрерывные отображения $\mathbb{N}_\perp \to \mathbb{N}_\perp$ бывают двух видов: если $\perp$ отображается в некоторое натуральное число (а не в $\perp$), то отображение постоянно; если же $\perp$ отображается в $\perp$, то натуральные числа могут отображаться куда угодно независимо друг от друга: отображения второго типа находятся во взаимно однозначном соответствии с частичными функциями из $\mathbb{N}$ в $\mathbb{N}$, если значение $\perp$ ставить в соответствие точкам, на которых функция не определена. Как и раньше, естественно назвать отображение $f\colon \mathbb{N}_\perp \to \mathbb{N}_\perp$ ***вычислимым***, если множество пар $\langle x, y\rangle$, для которых $y \preccurlyeq f(x)$, перечислимо. Ясно, что все постоянные отображения вычислимы, а среди непостоянных вычислимы как раз те, которым соответствуют вычислимые в обычном смысле функции (вычислимость функции равносильна перечислимости графика). Поэтому ничего нового мы не получаем: добавляются постоянные функции, отображающие элемент $\perp$ и все натуральные числа в некоторое c, но с точки зрения сложности они ничего не дают. (Формально говоря, такие функции следует отличать от функций, которые соответствуют постоянным всюду определённым функциям $\mathbb{N} \to \mathbb{N}$, поскольку первые отображают $\perp$ в натуральное число, а вторые — в $\perp$.)

Весь этот формализм, впрочем, нужен лишь в качестве мотивировки следующей схемы, объясняющей происхождение рассмотренных нами сложностей (рис. 13): Каждая из

		объекты	
		$\mathbb{N}_\perp$	Σ
описания	$\mathbb{N}_\perp$	*KS*	?
	Σ	*KP*	*KM*

Рис. 13. Происхождение трёх видов сложности.

[class-1]

трёх сложностей получается, если в качестве декомпрессоров (способов описания) рассматривать вычислимые непрерывные отображения соответствующих пространств (описаний в объекты).

Одна из клеток в этой таблице пока не заполнена; она соответствует способам описания, которые являются вычислимыми непрерывными отображениями $\mathbb{N}_\perp$ в Σ. Сейчас мы определим этот вид сложности, который называется *сложностью разрешения* и будет обозначаться *KR* (иногда его ещё обозначают *KD*, но в последнее время это обозначение стало применяться в ином смысле, для так называемой distinguishing complexity, так что мы будем его избегать).

Будем рассматривать в качестве способов описания машины, которые на вход получают двоичное слово (как завершённый объект, скажем, написанным на ленте с маркером конца после этого слова) и на выходе печатают бит за битом. При этом они могут и не останавливаться, так что для каждого входного слова x получается конечная или бесконечная последовательность. (Если эта последовательность бесконечна, то она, очевидно, вычислима.)

Таким образом, машине такого типа соответствует отображение множества всех двоичных слов (которые можно отождествить с натуральными числами, входящими в $\mathbb{N}_\perp$) в множество Σ конечных и бесконечных последовательностей. Если M — такая машина, то сложностью $\mathit{KR}_M(x)$ слова x относительно способа описания M называется длина кратчайшего слова y, для которого $M(y)$ (выходная последовательность машины, получившей y на входе) начинается на x.

139 Проверьте, что среди всех машин указанного вида имеется оптимальная машина M, для которой KR_M минимально с точностью до $O(1)$.

140 Дайте определение вычислимых непрерывных отображений $\mathbb{N}_\perp \to \Sigma$. Чем они отличаются от рассмотренных нами машин? Почему это различие несущественно для определения сложности? [Указание: значение на $\perp$ может быть непустым.]

Таким образом, пустая клеточка в нашей таблице заполняется (рис. 14):

описания \ объекты	$\mathbb{N}_\perp$	Σ
$\mathbb{N}_\perp$	*KS*	*KR*
Σ	*KP*	*KM*

Рис. 14. Четыре вида сложности.

[class-2]

В следующей теореме перечислены основные свойства сложности разрешения.

Theorem 102. [decision-complexity]

(а) *Если слово x является началом слова y, то $\mathit{KR}(x) \leqslant \mathit{KR}(y)$.*

(б) *Сложность начальных отрезков последовательности $\omega \in \Omega$ (неубывающая с ростом длины отрезка) ограничена тогда и только тогда, когда последовательность ω вычислима.*

(*Назвав предел сложности начальных отрезков сложностью разрешения последовательности* ω, *можно сказать, что сложность разрешения бесконечной последовательности конечна тогда и только тогда, когда последовательность вычислима.*)

(в) $KR(x) \leqslant KS(x) + O(1)$ *для любого слова* x.

(г) $KR(x) \leqslant KM(x) + O(1)$ *для любого слова* x.

(д) $KM(x) \leqslant KR(x) + O(\log KR(x))$ *для любого слова* x.

(е) $KS(x|l(x)) \leqslant KR(x) + O(1)$ *для любого слова* x.

(ё) *Если* $f\colon \Sigma \to \Sigma$ *— вычислимое непрерывное отображение, то* $KR(f(x)) \leqslant KR(x) + O(1)$ (*константа зависит от* f, *но не от* x).

(ж) *Если* $f\colon \Sigma \to \mathbb{N}_\perp$ *— вычислимое непрерывное отображение, то* $KS(f(x)) \leqslant KR(x) + O(1)$ (*константа зависит от* f, *но не от* x).

(з) *Если* $f\colon \mathbb{N}_\perp \to \Sigma$ *— вычислимое непрерывное отображение, то* $KR(f(x)) \leqslant KS(x) + O(1)$ (*константа зависит от* f, *но не от* x).

(и) *Любой набор попарно несравнимых слов* (*ни одно не является началом другого*), *все из которых имеют сложность разрешения меньше* n, *содержит менее* 2^n *слов.*

(й) *Функция* KR *перечислима сверху.*

(к) *Функция* KR *является минимальной* (*с точностью до константы*) *функцией, обладающей двумя предыдущими свойствами: если* (i) *функция* K *перечислима сверху и* (ii) *любое множество несравнимых слов, у которого* $K(x) < n$ *для всех элементов, содержит* $O(2^n)$ *слов, то* $KR(x) \leqslant K(x) + O(1)$.

(л) $KR(x) \leqslant KA(x) + O(1)$ *для всех слов* x.

$\lhd$ (а) Непосредственно следует из определения (описание слова есть описание любого его начала).

(б) Пусть последовательность ω вычислима. Возьмём в качестве способа описания машину, которая независимо от входа печатает ω на выходе бит за битом. Сложность начальных отрезков последовательности ω при таком способе равна нулю (пустой вход является их описанием), и потому при оптимальном способе описания ограничена. Обратно, если сложность начальных отрезков ограничена, то некоторое описание годится для бесконечного числа начальных отрезков, и потому последовательность вычислима.

(в) Всякая вычислимая частичная функция, аргументами и значениями которой являются двоичные слова, может рассматриваться как способ описания в нашем смысле (пока вычисление не закончилось, ничего не печатаем, как только закончилось — печатаем результат бит за битом).

(г) Всякое вычислимое непрерывное отображение $\Sigma \to \Sigma$ может рассматриваться как способ описания в нашем смысле (после ограничения на конечные аргументы; можно сказать, что мы с самого начала подаём описание на вход корректной машины и более никаких клавиш не нажимаем).

(д) Пусть $R : \mathbb{N} \to \Sigma$ — оптимальный способ описания при определении сложности разрешения. Рассмотрим вычислимое отображение $\hat{R} : \Sigma \to \Sigma$, положив $\hat{R}(\hat{x}u) = R(x)$, где $\hat{x}$ — самоограниченный код слова x (скажем, само x, предварённое двоичной записью его длины с удвоенными битами и разделителем 01), а u — любое слово (последнее необходимо для монотонности).

(е) Пусть снова $R : \mathbb{N} \to \Sigma$ — оптимальный способ описания для сложности разрешения. Определим способ условного описания S, положив $S(y, n)$ равным n первым битам последовательности $R(y)$ (если n больше длины последовательности, то $S(y,n)$ не определено).

(ё) Рассмотрим новый способ описания, являющийся композицией оптимального способа описания и отображения f, и сравним его с оптимальным.

(ж) Рассмотрим композицию оптимального способа описания для сложности разрешения и отображения f как способ описания для простой колмогоровской сложности.

(з) Рассмотрим композицию оптимального способа описания для простой колмогоровской сложности и отображения f как способ описания для сложности разрешения.

(и) Попарно несравнимые слова не могут иметь общего описания (в этом случае они были бы началами описываемой им последовательности). Если сложности попарно несравнимых слов меньше n, то их описания — различные слова длины меньше n, а таких слов меньше 2^n.

(й) Применяя параллельно оптимальный способ описания ко всем словам, мы получаем постепенно улучшающиеся верхние оценки для функции KR (постепенно обнаруживая новые описания).

(к) Это — первое содержательное утверждение теоремы (до сих пор были лишь простые вариации на знакомые темы).

Пусть дана функция K, обладающая двумя указанными свойствами. Увеличивая её на константу, можно считать без ограничения общности, что существует не более 2^n несравнимых слов x, для которых $K(x) < n$.

Мы построим способ описания, при котором всякое слово x с $K(x) < n$ будет иметь описание длины ровно n. Это делается параллельно и независимо для каждого n. А именно, наблюдая за уменьшающимися верхними оценками для K, мы пополняем список слов x, для которых $K(x) < n$. Можно считать, что в каждый момент этот список конечен и с течением времени растёт. Рассмотрим поддерево двоичного дерева, состоящее из слов этого списка и всех их начал. Со временем это поддерево растёт. В каждый момент у него не более 2^n листьев, поскольку листья являются несравнимыми словами x, для которых $K(x) < n$. Каждому листу мы присвоим метку, которая является словом длины n. При добавлении нового слова к дереву это новое слово либо продолжает какой-то из листьев (который перестаёт быть листом), либо ответвляется от дерева во внутренней вершине. В первом случае новое слово становится листом, который наследует метку прежнего листа. Во втором случае для нового листа мы выделяем новое слово в качестве метки (что всегда возможно, так как листьев меньше 2^n).

Фиксируем метку и посмотрим, что происходит с листьями с этой меткой. Вначале таких листьев нет (метка не использована). Возможно, так и останется навсегда (до этой метки дело не дойдёт), но если дойдёт, то метка будет сдвигаться по дереву вверх (каждое следующее её положение будет продолжением предыдущего), и потому ей соответствует некоторая конечная или бесконечная ветвь в дереве (последовательность нулей и единиц). Возникает отображение слов длины n в Σ (при этом меткам, до которых дело не дойдёт, соответствуют последовательности нулевой длины).

Объединив эти отображения при всех n, получим способ описания, при котором (для любого n) сложность всех слов x с $K(x) < n$ не превосходит n, что и требовалось.

(л) Если x_i — несравнимые двоичные слова, то $\sum 2^{-KA(x_i)} \leqslant 1$ (поскольку $2^{-KA(x_i)}$ есть априорная вероятность множества Σ_{x_i}, а эти множества не пересекаются). Поэтому несравнимых слов, у которых $KA(x_i) < n$, не может быть больше 2^n, и осталось восползоваться предыдущим утверждением теоремы. ▷

141 Покажите, что сложность разрешения можно определять так: для каждой вычислимой функции S двух аргументов (первый — слово, второй — натуральное число; значения — нули и единицы) определим $KR_S(x)$ для любого слова $x = x_0 \ldots x_{n-1}$ как наименьшую длину слова y, при котором $S(y, i) = x_i$ при всех $i = 0, 1, \ldots, n-1$, после чего среди всех функций S выберем оптимальную.

142 Покажите, что сложность разрешения слова x равна (с точностью до $O(1)$) минимуму $KS(p)$ по всем программам p (данного языка программирования, скажем, паскаля), которые не имеют входа и печатают на выходе слово x или его продолжение.

[Если вместо *KS* в этой задаче взять *KP*, получится верхняя оценка для монотонной сложности. Будет ли она точна? наверно, нет, но хорошо бы это проверить!]

6.2. Сравнение сложностей

[classcomp]

Сложности, входящие в рассмотренную нами таблицу (с двумя вариантами для пространств описаний и двумя вариантами для описываемых объектов), можно изобразить в виде ромба, стороны которого соответствуют неравенствам между сложностями (с точностью до $O(1)$), как показано на рис. 15:

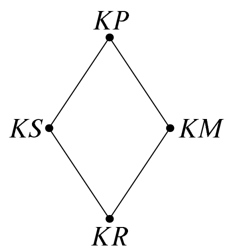

Рис. 15. Неравенства между сложностями.

[class-3]

Если мы хотим избежать упоминания топологических понятий и определения вычислимых непрерывных отображений для разных пространств (хотя они здесь по существу, как учит теория абстрактных типов данных в смысле Скотта и близкая к ней теория f_0-пространств Ершова, см. [70]), то для построения указанных четырёх сложностей по единой схеме можно обойтись следующей упрощённой конструкцией.

На множестве $\Xi = \mathbb{B}^*$ всех слов рассмотрим два бинарных отношения: равенство слов $(=)$ и сравнимость слов (наличие общего продолжения, обозначение $\asymp$). Пусть α и β — любые из этих двух отношений (так что всего есть четыре варианта).

Будем говорить, что множество $S \subset \Xi \times \Xi$ является α-β-корректным, если следующее условие выполнено для всех слов x_1, x_2, y_1, y_2:

$$(x_1, y_1) \in S,\ (x_2\, y_2) \in S,\ x_1 \alpha x_2 \quad \Rightarrow \quad y_1 \beta y_2$$

Например, $=$-$=$-корректные отношения — это униформные множества, то есть графики функций.

143 [computable-sets] (а) Покажите, как по $\asymp$-$=$-корректному отношению построить непрерывное отображение $\Sigma \to \mathbb{N}_\perp$.

(б) Покажите, как по $\asymp$-$\asymp$-корректному отношению построить непрерывное отображение $\Sigma \to \Sigma$.

(в) Покажите, как по $=$-$\asymp$-корректному отношению построить непрерывное отображение $\mathbb{N}_\perp \to \Sigma$.

Будем называть *α-β-способом описания* перечислимое α-β-корректное отношение на $\Xi \times \Xi$. Для всякого способа описания S определим сложность как функцию, сопоставляющую с каждым словом x длину его кратчайшего описания, то есть длину кратчайшего слова y, для которого $\langle y, x\rangle$ принадлежит S.

Theorem 103. *Для каждой из четырёх комбинаций $\alpha, \beta \in \{=, \asymp\}$ существует оптимальный α-β-способ описания, и соответствующая сложность совпадает (с точностью до $O(1)$) с одной из четырёх сложностей KS, KP, KM, KR.*

$\lhd$ В каждом из случаев по α-β-корректному способу описания легко строится вычислимое отображение соответствующих пространств (см. задачу 143), задающее ту же самую сложность, и наоборот. $\rhd$

Таким образом, в таблице сложностей можно поменять названия строк и столбцов (рис. 16):

описания \ объекты	$=$	$\asymp$
$=$	*KS*	*KR*
$\asymp$	*KP*	*KM*

Рис. 16. α-β-сложности.

[class-4]

144 Покажите, как определить для пары слов:

(а) монотонную сложность, рассматривая в качестве декомпрессоров вычислимые непрерывные отображения $\Sigma \to \Sigma \times \Sigma$. (Такие отображения можно отождествить с парами вычислимых отображений $\Sigma \to \Sigma$.)

(б) априорную вероятность, рассматривая машины с датчиком случайных чисел и двумя выходными лентами, где можно печатать бит за битом.

(в) сложность разрешения, рассматривая вычислимые непрерывные отображения $\mathbb{N}_\perp \times \Sigma \to \Sigma$.

[Про свойства этих мер сложности вроде бы почти ничего не известно: скажем, верно ли, что $KM(x, y) \leqslant l(x) + l(y)$?]

Другая схема (восходящая к Левину, см. [31]) классификации сложностей состоит в описании их как наименьших перечислимых сверху функций, удовлетворяющих некоторым ограничениям. Соберём соответствующие ограничения (на перечислимую сверху функцию K):

- число различных слов x, при которых $K(x) < n$, есть $O(2^n)$ (простая сложность *KS*, теорема 8, с. 25);
- ряд $\sum_x 2^{-K(x)}$ сходится (префиксная сложность *KP*, теорема 56, с. 98);
- число различных несравнимых слов x, при которых $K(x) < n$, есть $O(2^n)$ (сложность разрешения *KR*, теорема 102, с. 164);
- $\sum_x 2^{-K(x_i)} \leqslant 1$ для любого множества попарно несравнимых слов x_i (априорная сложность *KA*, теорема 73, с. 126).

В этой схеме фигурируют те же четыре сложности, за одним исключением: вместо монотонной сложности фигурирует априорная. (Для случая префиксной сложности такого расхождения нет, поскольку она совпадает с логарифмом максимальной перечислимой снизу полумеры на $\mathbb{N}$.)

Поэтому при объединении этих двух схем у нас получается уже не ромб, а пятиугольник (рис. 17):

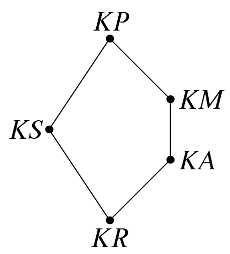

Рис. 17. Пять сложностей

[class-5]

Напомним основные результаты о сравнении сложностей в этом пятиугольнике. Прежде всего, все сложности отличаются не более чем на $O(\log n)$ для слов длины n. В самом деле, $\mathit{KP}(x) \leqslant \mathit{KS}(x) + O(\log \mathit{KS}(x))$ (теорема 59, с. 100). С другой стороны, $\mathit{KS}(x) \leqslant \mathit{KS}(x|l(x)) + \mathit{KS}(l(x)) \leqslant \mathit{KR}(x) + O(\log n)$. Поэтому две наиболее различающиеся сложности в этом пятиугольнике (верхняя и нижняя) отличаются не более чем на $O(\log n)$ для слов длины n.

Более сложная ситуация возникает, если мы хотим ограничить разность между двумя сложностями константой, умноженной на логарифм одной из них. Это можно сделать для неравенств вдоль линий, идущих в «северо-восточном» направлении:

$$\mathit{KP}(x) \leqslant \mathit{KS}(x) + O(\log \mathit{KS}(x))$$

(см. теорему 59) и

$$\mathit{KM}(x) \leqslant \mathit{KR}(x) + O(\log \mathit{KR}(x))$$

(теорема 102). Тем более это верно, если заменить *KM* на *KA* (как мы уже упоминали в задаче 110, с, 134). Для «северо-западных» неравенств это сделать нельзя: *KM* и *KR* ограничены для начальных отрезков вычислимых последовательностей (в частности, для слов из одних нулей), а *KS* и *KP* — нет (для слова из n нулей сложность равна сложности n, и достигает $\log n$ при некоторых n). Мы уже обсуждали это в теореме 79, где отмечали, что разница между *KP* и *KM* может быть обоих знаков порядка $\log n$ для некоторых слов длины n (при бесконечно многих n).

Там же мы упоминали без доказательства теорему Гача о том, что разница между *KM* и *KA* не ограничена (теорема 80).

Среди результатов, связывающих различные сложности, есть и довольно тонкие. Например, ни из одного до сих пор доказанного результата не следует, что разность между $KP(x)$ и $KS(x)$ стремится к бесконечности. Это можно вывести из следующей теоремы, доказанной Р. Соловеем (R. Solovay) в 1970-е годы (рукопись с этим доказательством так и не была опубликована, хотя была доступна многим специалистам).

Theorem 104. `[solovay-bound]`

$$KS(x) \leqslant KP(x) - KP(KP(x)) + KP(KP(KP(x))) + O(1)$$

$\triangleleft$ Мы видели (теорема 58, с. 99), что логарифм числа слов, для которых $KP(x) \leqslant n$, не превосходит $n - KP(n) + c$ при некотором c и при всех n. Кроме того, такие слова можно перечислять (при известном n), так что каждое такое слово можно восстановить, зная n и порядковый номер слова в перечислении. Этот порядковый номер мы будем записывать в виде двоичного слова ровно из $n - KP(n) + c$ битов (добавив слева нули, если нужно). Имея такое слово, мы уже знаем $n - KP(n)$ (константу c считаем фиксированной), так что для задания n достаточно задать ещё $KP(n)$, что можно сделать с помощью самоограниченного кода длиной $KP(KP(n))$. Приписав к этому коду упомянутый выше порядковый номер, мы получим (для любого слова x с $KP(x) \leqslant n$ в том числе для всех слов префиксной сложности n) слово длиной $KP(KP(n)) + n - KP(n) + c$, по которому мы можем эффективно восстановить слово x. Что и требовалось. $\triangleright$

[Другие результаты Соловея?

Что можно сказать про разницу между *KM* и *KA*?

Результат Мучника про последовательности x_i и y_i, для которых $KS(x_i) - KS(y_i) \to +\infty$, но $KP(x_i) - KP(y_i) \to -\infty$.

Miller: *KS* vs *KP*?

A priori сложность для пар отличается от монотонной?]

6.3. Условные сложности

`[classcond]`

Мы уже упоминали несколько видов условной сложности (когда определяется сложность одного слова относительно другого). В разделе 2.2 мы определяли условную сложность $KS(x|y)$ как длину кратчайшего описания p слова x при известном y, то есть минимальную длину слова p, для которого $S(p, y) = x$. Здесь S — способ описания; способами описания были произвольные частичные вычислимые функции двух аргументов.

В разделе 4.7 определялась условная префиксная сложность $KP(x|y)$. При этом требовалось, чтобы способ описания был префиксно корректным *относительно* p: это значит, что если слово p является описанием данного x при данном y, то любое продолжение слова p также является таким описанием.

Наконец, при доказательстве теоремы 85 мы упоминали условную монотонную сложность $KM(x|y)$. При её определении в качестве способа описания берётся вычислимое семейство вычислимых непрерывных отображений $D_y\colon \Sigma \to \Sigma$, индексированное словами y. Вычислимость семейства понимается в том смысле, что множество троек $\langle y, u, v\rangle$, для которых $v \preccurlyeq D_y(u)$, перечислимо.

Аналогичным образом можно было бы определить и условную сложность разрешения.

Во всех этих определениях условия рассматриваются как законченные слова, никак не связанные друг с другом: если p является описанием x при известном y, при другом y то же самое p может быть описанием совершенно другого x (даже если новое y отличается от старого добавлением одного бита).

Другими словами, способ описания (скажем, для условной префиксной сложности) мы считаем вычислимым отображением

$$D\colon \Sigma \times \mathbb{N} \to \mathbb{N}_\perp;$$

при этом первый член пары $\langle p, y\rangle \in \Sigma \times \mathbb{N}$ считается описанием (требуется монотонность по p), а второй условием (монотонность по y не требуется).

Если рассматривать условия как вершины дерева и требовать монотонности по этому аргументу, получатся четыре других вида условной сложности, которые, однако, практически не рассматривались (работа [10] — одно из редких исключений).

Таким образом, всего возникает 8 видов условных сложностей (каждый из трёх компонентов — условия, описания и объекты — можно понимать двумя способами). Технически проще всего определить эти условные сложности так. Пусть $\alpha, \beta, \gamma \in \{=, \asymp\}$ (см. раздел 6.2). Определим $(\alpha, \beta)|\gamma$-способ описания как перечислимое множество S троек слов $\langle p, x, y\rangle$, удовлетворяющее следующему требованию:

$$\langle p_1, x_1, y_1\rangle \in S, \quad \langle p_2, x_2, y_2\rangle \in S, \quad p_1\alpha p_2, \quad y_1\gamma y_2 \Rightarrow x_1\beta x_2$$

Далее сложность $K_S(x|y)$ определяется как наименьшая длина слова p, при котором $\langle p, x, y\rangle \in S$.

Theorem 105. *В каждом из восьми случаев среди способов описания существует оптимальный (для которого сложность минимальна в этом классе способов описания с точностью до $O(1)$).*

$\boxed{\textbf{145}}$ Проведите аккуратное доказательство этой теоремы (по существу ничем не отличающееся от случая обычной или префиксной условной сложности).

Соответствующую оптимальному $(\alpha, \beta)|\gamma$ способу описания сложность можно обозначить $K_{(\alpha,\beta)|\gamma}$. При таком обозначении $KP(x|y)$ есть $K_{(\asymp,=)|=}$, а $KS(x|y)$ есть $K_{(=,=)|=}$.

$\boxed{\textbf{146}}$ Покажите, что от замены $=$ на $\asymp$ в качестве γ условная сложность может только возрасти. [Указание: на способ описания накладываются дополнительные ограничения, поэтому их становится меньше. По этой же причине обычная сложность не превосходит префиксной.]

[Задача (наверно, несложная): показать, что увеличение может быть больше чем на $O(1)$. И вообще исследовать эти сложности.]

Вот пример утверждения, включающего в себя рассмотренные только что модификации условных сложностей:

147 Докажите, что

$$\mathit{KS}(x) \leqslant K_{(=,=)|\asymp}(x|y) + \mathit{KR}(y) + O(\log \mathit{KR}(y)).$$

Опишем ещё один подход к определению условной сложности, восходящий к колмогоровской интерпретации логических связок как операций над задачами [21]. Условную сложность x при известном y можно интерпретировать как сложность задачи «указать x по данному y»; в свою очередь, эту задачу можно считать множеством всех функций, переводящих y в x (любая такая функция считается «решением» этой задачи).

Это можно уточнить, например, так. Рассмотрим пространство $\mathbb{F}$, элементами которого являются все частичные функции с натуральными аргументами и значениями. На этом множестве определим порядок: $f_1 \preccurlyeq f_2$, если функция f_2 продолжает функцию f_1. *Конечными элементами* этого множества будем считать функции с конечной областью определения. Для каждого конечного элемента f рассмотрим конус над ним, то есть множество всех его продолжений (конечных и бесконечных). Непрерывное отображение $T\colon \mathbb{N}_\perp \to \mathbb{F}$ будем называть *вычислимым*, если перечислимо множество пар $\langle a, f\rangle$, где $a \in \mathbb{N}_\perp$, f — конечный элемент $\mathbb{F}$ и $f \preccurlyeq T(a)$. Такие вычислимые отображения будем считать способами описания (функций). Для каждой функции f определим сложность (относительно данного способа описания T) как минимальную длину слова (точнее, логарифм числа, ведь мы отождествляем числа и двоичные слова) a, для которого $f \preccurlyeq T(a)$.

148 Докажите, что при таком определении существует оптимальный способ описания и что сложность функции $y \mapsto x$ (область определения состоит из единственного слова y, значение на котором равно x) равна $\mathit{KS}(x|y) + O(1)$.

В этом же духе можно интерпретировать и все восемь указанных выше условных сложностей (для пространств $Y, X \in \{\mathbb{N}_\perp, \Sigma\}$ определив пространство функций $(Y \to X)$, а затем в качестве способов описания рассматривая вычислимые отображения пространства описаний $P \in \{\mathbb{N}_\perp, \Sigma\}$ в пространство $(Y \to X)$. При этом удобно использовать домены Скотта или f_0-пространства в смысле Ершова (подробности см. в [70]).

Несколько другая интерпретация условной сложности как сложности задачи «получить x из y», не использующая вычислимость в пространствах функций, рассмотрена в главе 13.

Родственное (хотя и несколько другое) понятие сложности функции было рассмотрено Шнорром [67,68]. Напомним, что в теории рекурсии *нумерацией* называется некоторое отображение ν, сопоставляющее с каждым натуральным числом n некоторую функцию ν_n (частичную) из $\mathbb{N}$ в $\mathbb{N}$. Нумерация ν называется *вычислимой*, если частичная функция двух аргументов

$$\langle n, x\rangle \mapsto \nu_n(x)$$

является вычислимой, и *гёделевой*, или (в русскоязычной литературе) *главной*, если для любой другой вычислимой нумерации μ существует вычислимая всюду определённая функция h, *сводящая* μ к ν в том смысле, что $\mu_n = \nu_{h(n)}$ при всех n. (В частности, среди ν_n должны встречаться все вычислимые функции.)

Следуя Шнорру, усилим это условие и дополнительно потребуем, чтобы $h(n) = O(n)$ (длина слова $h(n)$ превосходила бы длину n не более чем на константу, если отождествить натуральные числа с двоичными словами). Если для всякой вычислимой нумерации μ существует функция h с таким свойством, то нумерацию ν будем называть *оптимальной*.

Theorem 106. *Оптимальные нумерации существуют.*

$\triangleleft$ Выберем какой-либо естественный язык программирования двуместных функций и будем считать слово $\hat{u}v$ номером функции, которая получится, если в функции с программой u фиксировать первый аргумент равным v. (Здесь $\hat{u}$ — самоограниченное описание слова u, получаемое, например, удвоением каждого бита и приписыванием 01 в конец.) $\triangleright$

Шнорр показал (см. [67,68]), что любые две оптимальные нумерации ν_1 и ν_2 отличаются вычислимой перестановкой номеров π, увеличивающей объём в ту и другую сторону не более чем на $O(1)$. (Это значит, что $\nu_1(n) = \nu_2(\pi(n))$ при всех n, причём $\pi(n) = O(n)$ и $\pi^{-1}(n) = O(n)$.)

6.4. Сложность относительно оракула

[classlim]

В теории вычислимых функций хорошо известен следующий метод, называемый *релятивизацией*: берётся какое-то определение или утверждение, касающееся класса вычислимых функций, и всюду в нём вычислимые функции заменяются на функции, вычислимые относительно некоторого *оракула*. В качестве такого оракула берётся некоторая всюду определённая функция α, аргументами и значениями которой являются натуральные числа или двоичные слова (обычно характеристическая функция некоторого множества A). Алгоритмам разрешается в качестве элементарного действия обращаться к «внешней процедуре», возвращающей значение $\alpha(n)$ для переданного ей значения параметра n. (Для характеристической функции множества A это означает, что про любое натуральное число можно узнать, принадлежит ли оно множеству A или нет.) Если сама функция α не вычислима (множество A неразрешимо), то класс вычислимых функций расширяется; вычисляемые такими алгоритмами функции называются α-вычислимыми.

После этого практически все понятия и теоремы теории вычислимых функций переносятся на α-вычислимые функции. Скажем, можно говорить об α-перечислимых множествах, или об α-вычислимых действительных числах, или (подходя ближе к колмогоровской сложности) об α-перечислимых снизу мерах. При этом (практически все) теоремы остаются верными и в таком «релятивизованном» варианте.

В частности, для любого множества A можно определить понятие колмогоровской сложности, релятивизованной относительно множества A, рассматривая способы описания (декомпрессоры) с оракулом A. Это можно сделать и для обычной, и для префиксной, и для всех других видов рассмотренных нами сложностей (условных и безусловных).

Обычно использование оракула отмечают верхним индексом, так что, скажем, $KP^A(x)$ есть префиксная сложность слова x с оракулом A.

На самом деле мы делаем даже несколько большее: мы не просто для фиксированного оракула A определяем релятивизованную сложность (с точностью до $O(1)$), а определяем (с этой точностью) некоторую функцию двух аргументов: $\langle x, A\rangle \mapsto K^A(x)$ (где K — один из четырёх видов сложности).

149 Проверьте, что это действительно можно сделать, и что получающиеся сложности являются пределами условных сложностей, описанных в разделе 6.3:

$$KP^A(x) = K_{\asymp,=}(x) = \lim_{n\to\infty} K_{(\asymp,=|\asymp)}(x|A_n),$$

где A_n — начальный отрезок длины n характеристической функции множества A. (Аналогично и для других видов сложностей.)

Заметим, что релятивизованные сложности (с точностью $O(1)$) не больше обычных (алгоритм может не обращаться к оракулу, потому среди A-способов описания есть и обычные способы описания).

При некоторых A функция сложности с оракулом A может существенно уменьшаться (по сравнению со сложностью без оракула). Например, рассмотрим в качестве оракула A универсальное перечислимое неразрешимое множество. (Такой оракул обычно обозначают $\mathbf{0}'$.) Другими словами, $\mathbf{0}'$-оракул есть внешний оракул для проблемы остановки: ему можно послать для анализа программу (без оракула), и он скажет, завершится её работа или нет. Имея такой оракул, можно для каждого слова x найти его кратчайшее описание (поскольку мы можем с помощью оракула проверять, закончит декомпрессор работу на данном слове или нет). Поэтому функция KS является $\mathbf{0}'$-вычислимой (это же относится к функции KP, условным сложностям и т.п.), а список всех слов сложности меньше n, который без оракула имел сложность $n+O(1)$, а также числа $B(n)$ и $BB(n)$ (см. раздел 1.2) теперь имеют $\mathbf{0}'$-сложность $O(\log n)$.

С другой стороны, большинство слов длины n имеют $\mathbf{0}'$-сложность $n - O(1)$, и потому для них $\mathbf{0}'$-сложность близка к обычной (нерелятивизованной).

6.4.1. Сложность при условии больших чисел

Определим новый вид условной сложности: сложность числа (слова x) относительно множества A. Эта сложность, говоря неформально, есть сложность такой задачи: «получить x, если задан некоторый элемент множества A» (неизвестно какой).

Такую сложность можно определить несколькими эквивалентными способами.

Пусть фиксирован некоторый разумный язык программирования. (Формально говоря, нужно, чтобы соответствующая ему нумерация была главной, то есть чтобы была возможна эффективная трансляция программ с других языков [77].) Определим условную сложность объекта (слова, натурального числа) x относительно множества объектов A как минимальную (простую колмогоровскую) сложность программы, переводящей *любой* из элементов множества A в x. (В более общем виде это определение рассматривается в главе 13.)

Возможность трансляции гарантирует, что это определение корректно, то есть что сложность (с точностью до $O(1)$) не зависит от выбора главной нумерации.

Важно отличать эту сложность от сложности x относительно конечного множества A, заданного списком своих элементов (известен не список элементов A, а некоторый его элемент — неважно какой). Мы будем обозначать только что определённую сложность $KS(x\|A)$, сохраняя обозначение $KS(x|A)$ для сложности x относительно конечного множества A, заданного списком своих элементов.

Другой (эквивалентный) способ определения $KS(x\|A)$ состоит в следующем. Пусть D — вычислимая частичная функция двух аргументов (декомпрессор), x — двоичное слово, а A — множество двоичных слов. Определим $KS_D(x\|A)$ как наименьшую длину такого слова p, что $D(p,y)=x$ для всех $y \in A$.

$\boxed{\mathbf{150}}$ `[relative-complexity]` Докажите, что среди всех вычислимых частичных функций имеется оптимальная функция (для которой функция KS_D минимальна с точностью до $O(1)$). Докажите, что для оптимальной функции D величина $KS_D(x\|A)$ совпадает с ранее определённой сложностью $KS(x\|A)$ с точностью до $O(1)$.

Для случая одноэлементного множества $A=\{a\}$ обе величины $KS(x|A)$ и $KS(x\|A)$ совпадают (с точностью до $O(1)$) с обычной условной сложностью $KS(x|a)$; см. задачу 23.

Рассмотрим в качестве A множество всех чисел, больших некоторого числа n. (Как всегда, числа мы отождествляем со словами.) Сложность x относительно этого множества будем обозначать $KS(x\| \geqslant n)$. Очевидно, эта сложность не превосходит $KS(x)$ и убывает с возрастанием n (вообще $KS(x\|A)$ очевидным образом убывает при уменьшении множества A, превращаясь в $O(1)$ для пустого A). Поэтому у неё есть некоторый предел при $n\to\infty$.

Theorem 107. `[large-number-condition1]`

$$\lim_{n\to\infty} KS(x\| \geqslant n) = KS^{\mathbf{0}'}(x) + O(1).$$

$\lhd$ Пусть указанный предел равен k. Тогда существует программа p сложности k, переводящая все достаточно большие числа в x. Имея дополнительно оракул $\mathbf{0}'$, можно использовать эту программу как описание объекта x. Именно, надо путём перебора найти такие N и y, что программа *p не переводит никакое $n \geqslant N$ ни в какой объект, отличный от y*. Это можно сделать с помощью оракула, поскольку выделенное свойство пары N, y имеет перечислимое дополнение. При этом из нашего предположения следует, что $y=x$. Поэтому

$$KS^{\mathbf{0}'}(x) \leqslant \lim_{n\to\infty} KS(x\| \geqslant n) + O(1).$$

Напротив, пусть имеется описание y объекта x относительно $\mathbf{0}'$-оптимального способа описания, имеющее длину k. Рассмотрим следующую программу: получив некоторое N, сделать N шагов порождения универсального перечислимого множества $\mathbf{0}'$ и затем использовать полученное множество в качестве оракула, применяя оптимальный способ описания к y. Эта программа эффективно строится по y и потому её сложность не превосходит $KS(y)+O(1) \leqslant l(y)+O(1) = k+O(1)$. С другой стороны, для достаточно больших N эта программа порождает x (поскольку при вычислении значения оптимального способа описания на y используется лишь конечное число запросов к оракулу, при достаточно большом N все эти запросы получат правильные ответы даже и при замене настоящего оракула на его N-шаговое приближение). $\rhd$

Оказывается, что аналогичное утверждение верно, если заменить $KS(x\| \geqslant n)$ на величину $\sup_{m\geqslant n} KS(x|m)$. Сразу же ясно, что

$$\sup_{m\geqslant n} KS(x|m) \leqslant KS(x\| \geqslant n),$$

поскольку оптимальная программа в правой части годится и для любого m из левой. Удивительным образом оказывается, что это уменьшение не меняет предела при $n \to \infty$:

Theorem 108. [large-number-condition2]

$$\limsup_{n\to\infty} KS(x|n) = KS^{\mathbf{0}'}(x) + O(1).$$

$\triangleleft$ Нужно доказать, что если для данного слова x выполнено свойство

сложность $KS(x|n)$ меньше k при всех достаточно больших n,

то $\mathbf{0}'$-сложность x не превосходит $k + O(1)$. Сложность здесь — в отличие от предыдущей теоремы — в том, что эти программы (длины меньше k для разных n могут быть разными).

Заметим, что слов x с указанным свойством (для данного k) меньше 2^k: если бы их было больше, то при достаточно бóльших n на них не хватило бы программ длины меньше k.

Поэтому было бы достаточно доказать, что множество таких слов является $\mathbf{0}'$-перечислимым и соответствующий алгоритм перечисления эффективно строится по k (другими словами, что функция $x \mapsto \limsup KS(x|n)$ является $\mathbf{0}'$-перечислимой сверху). Однако естественное описание этого множества,

$$\exists N\, (\forall n \geqslant N)\, [KS(x|n) < k],$$

доказывает лишь его принадлежность к Σ_3 (условие в квадратных скобках перечислимо, и есть два дополнительных квантора), так что мы поступим иным образом.

Заметим, что нам не нужна $\mathbf{0}'$-перечислимость этого множества, нам достаточно включить его в некоторое $\mathbf{0}'$-перечислимое множество, содержащее менее 2^k элементов с данным k. Это делается следующим образом.

Рассмотрим двумерное перечислимое множество пар $\langle n, x\rangle$, для которых $KS(x|n) < k$. Это множество «тонкое» в том смысле, что все вертикальные сечения этого множества содержат менее 2^k элементов.

Пусть выбрана некоторая точка $\langle n, x\rangle$. Попробуем добавить горизонтальный луч, выходящий из этой точки, к нашему множеству (добавить в него все пары $\langle m, x\rangle$ при $m \geqslant n$). При этом множество может перестать быть тонким, а может и не перестать, и эти два случая можно различить с помощью $\mathbf{0}'$-оракула: наружение тонкости есть перечислимое свойство (найдётся сечение, в котором обнаружатся 2^k различных элементов — вместе с добавленным).

Будем выполнять эту операцию по очереди для всех лучей (для всех пара $\langle n, x\rangle$ в некотором порядке. (При этом, если какой-то луч удалось добавить без нарушения тонкости, то в дальнейшем его элементы учитываются наравне с элементами исходного множества при добавлении следующих лучей.) Этот процесс является $\mathbf{0}'$-вычислимым, и потому x-координаты всех добавленных лучей образуют $\mathbf{0}'$-перечислимое множество.

Оно содержит менее 2^k элементов (поскольку добавление лучей не нарушает тонкости), а также содержит все x, у которых $\limsup K(x|n) < k$. В самом деле, для такого x найдётся луч, изначально лежавший в множестве, и потому его добавление заведомо возможно. $\triangleright$

(Это доказательство является упрощённым вариантом изложенного в статье [76].)

Теоремы, аналогичные теоремам 107 и 108, имеют место и для префиксной сложности. Определение префиксной сложности относительно множества требует специального обсуждения, которое мы отложим и начнём со второй из этих теорем.

Theorem 109. [large-number-condition3]

$$\limsup_{n\to\infty} \mathit{KP}(x|n) = \mathit{KP}^{\mathbf{0}'}(x) + O(1)$$

$\triangleleft$ Переходя к априорным вероятностям (условным и безусловным), эту теорему можно переписать так:

$$\liminf_{n\to\infty} m(x|n) = m^{\mathbf{0}'}(x)$$

(с точностью до ограниченного и отделённого от нуля множителя).

Покажем сначала, что левая часть не меньше правой (точнее, меньше не более чем в $O(1)$ раз). В самом деле, рассмотрим машину с оракулом $\mathbf{0}'$, выход которой имеет распределение $m^{\mathbf{0}'}$. Теперь будем для каждого n запускать эту машину с конечной частью оракула, соответствующей n шагам его перечисления. При этом вероятность получить x на выходе может и увеличиться и измениться, но нижний предел этих вероятностей не меньше соответствующей вероятности для настоящего оракула. В самом деле, вероятность получить x для настоящего оракула есть мера открытого множества (объединения конусов), и каждый из этих конусов использует лишь конечное число вопросов к оракулу и на некотором шаге появится.

Докажем теперь обратное неравенство. Тут ситуация очень похожа на доказательство теоремы 108. Имеется перечислимое снизу семество полумер; при каждом n функция $x \mapsto m(x|n)$ является полумерой (сумма ряда по x не больше 1). Отсюда легко следует, что для нижнего предела

$$m'(x) = \liminf_{n\to\infty} m(x|n)$$

сумма ряда $\sum_x m'(x)$ также не превышает единицы. Если бы функция m' оказалась $\mathbf{0}'$-перечислимой снизу, то всё было бы доказано — но запись этого свойства в виде

$$r < \liminf_{n\to\infty} m(x|n) \Leftrightarrow (\exists q > r)\, \exists N\, (\forall n > N)\, [q < m(x|n)]$$

содержит лишний квантор (свойство в квадратных скобках перечислимо, а не разрешимо). Но, как и в предыдущем доказательстве, нам достаточно построить $\mathbf{0}'$-перечислимую мажоранту.

Для этого будем рассматривать тройки $\langle N, x, \varepsilon$ (где ε — положительное рациональное число) и будем пытаться увеличивать функцию $m(\cdot|\cdot)$ до ε на луче, состоящем из пар $\langle n, x\rangle$ при данном x и при всех $n \geqslant N$, если это не нарушит свойство полумеры (для каждого n сумма по всем x не превышает 1).

Как и раньше, возможность такого добавления можно проверить с помощью $\mathbf{0}'$-оракула (поскольку невозможность есть перечислимое свойство). Будем производить такие добавления для всех троек (учитывая уже сделанные увеличения при определении возможности следующих). Оставляя от каждой добавленной тройки x и ε (то есть рассматривая для каждого x точную верхную грань всех ε, с которыми это x добавлено), получим $\mathbf{0}'$-перечислимую снизу полумеру, мажорирующую m' (ведь если m' где-то больше ε, то на некотором луче функция уже больше ε и тем самым добавление заведомо возможно). $\triangleright$

Чтобы сформулировать аналогичное утверждение для $KP(x\| \geqslant n)$, надо прежде всего определить эту сложность (префиксную сложность с условием, являющимся множеством). Тут есть несколько возможностей, и не вполне ясно, какую из них следует считать «правильной».

Можно пытаться определить $KP(x\|A)$ как минимальную префиксную сложность программы, которая даёт x на любом входе из множества A. Однако, как показывает задача 79 (с. 102) тогда для одноэлементных множеств $A = \{a\}$ не получится $KP(x|a)$ и потому, видимо, такое определение надо считать неудачным.

Можно действовать по аналогии с задачей 150. Рассмотрим произвольную вычислимую функцию $\langle p, x\rangle \mapsto D(p, x)$ префиксно корректную по первому аргументу (при фиксированном втором). Для любого x и любого множества A определим $KP_D(k\|A)$ как наименьшую длину такого p, что $f(p, n) = k$ для всех $n \in A$. Разница с определением для обычной сложности состоит в том, что мы требуем от декомпрессора префиксной корректности по первому аргументу. Среди таких декомпрессоров есть оптимальный (в этом классе), и соответствующую функцию KP_D можно принять за определение префиксной сложности с множеством в качестве условия.

151 Покажите, что та же самая сложность (с точностью до $O(1)$) получится, если в качестве декомпрессоров рассматривать вычислимые непрерывные отображения $\Sigma \to \mathbb{F}$ (пространства конечных и бесконечных последовательностей нулей и единиц в пространство частичных функций из $\mathbb{N}$ в $\mathbb{N}$) и рассматривать кратчайшее слово p, образ которого есть некоторая частичная функция, равная x на всех элементах A.

Можно определять префиксную сложность относительно множества, используя беспрефиксные функции вместо префиксно корректных. В классе вычислимых беспрефиксных по первому аргументу функций существует функция, для которой относительная сложность $KP_f(x\|A)$ минимальна. Таким образом мы получаем определение $KP'(x\|A)$, аналогичное условной сложности $KP'(k|n)$ и совпадающее с ней при $A = \{n\}$ (с точностью до аддитивной константы).

Наконец, можно определить априорную вероятность $m(x\|A)$, взяв вероятностную машину с входом y и изучая меру множества всех последовательностей ω, которые (будучи использованы в качестве случайных битов) засталяют машину напечатать x при любом входе $y \in A$. Здесь также существует оптимальная машина (при которой эта вероятность наибольшая) и для одноэлементных множеств A это определение сводится к ранее известному.

Как и раньше, для всех k, A выполнены неравенства

$$-\log m(k\|A) \leqslant KP(k\|A) + O(1) \leqslant KP'(k\|A) + O(1),$$

но авторам неизвестно, являются ли они строгими. Однако все эти величины превышают

$$-\log \inf_{x\in A} m(k|x) = \sup_{x\in A} \mathit{KP}(x|a),$$

и потому любая из них может быть использована в теореме, аналогичной теореме 107. В частности, для (видимо, наиболее естественной) величины $\mathit{KP}(x\|A)$, определённой выше с помощью префиксно корректных функций, получаем такое утверждение:

Theorem 110. [large-number-condition4]

$$\lim_{n\to\infty} \mathit{KP}(x\| \geqslant n) = \mathit{KP}^{\mathbf{0}'}(x) + O(1).$$

[Хорошо бы выяснить, действительно ли неравенства строгие]

6.4.2. Пределы частот и априорная вероятность, релятивизованная $\mathbf{0}'$

В заключение приведём результат из статьи [57], связывающий частоты появления в вычислимых последовательностях с релятивизованной относительно $\mathbf{0}'$ префиксной сложностью.

Пусть $f(0), f(1), \ldots$ — вычислимая последовательность натуральных чисел. Для данных натуральных n и k подсчитаем, сколько раз встречается k среди $f(0), \ldots, f(n-1)$, и поделим результат на n. Полученное частное можно назвать *частотой* k среди первых n членов последовательности.

Теперь при фиксированном k рассмотрим *нижний предел* частоты k среди первых n членов при $n \to \infty$, который будем называть *нижней частотой* числа k в последовательности f.

Пусть p_k — нижняя частота числа k в данной последовательности. Легко видеть, что $\sum_k p_k \leqslant 1$. В самом деле, если какая-то конечная частичная сумма этого ряда была бы больше 1, то сумма соответствующих нижних пределов была бы больше 1, и потому сумма достаточно далёких допредельных значений превзошла бы 1, что невозможно (сумма частот в любом начальном отрезке не больше 1.

Кроме того, выполнено следующее утверждение (для любой вычислимой последовательности f и нижних частот p_k появления в ней различных чисел k):

Theorem 111. [zero-prime-semimeasure] *Функция $k \mapsto p_k$ $\mathbf{0}'$-перечислима снизу.*

Определение перечислимой снизу функции мы давали в разделе 4.1; сейчас мы рассматриваем релятивизованный относительно $\mathbf{0}'$ вариант этого понятия.

$\triangleleft$ В самом деле, утверждение $r < p_k$ (где r — рациональное число) можно записать так:

> существует такое рациональное $p > r$ и такое N, что *частота появления k в любом начальном отрезке последовательности длины больше N превосходит p.*

Набранное курсивом свойство является коперечислимым (имеет перечислимое дополнение): если оно неверно, то это можно обнаружить, предъявив соответствующий начальный отрезок. Поэтому это свойство $\mathbf{0}'$-разрешимо (применим оракул к алгоритму поиска этого отрезка). А потому свойство $r < p_k$ в целом $\mathbf{0}'$-перечислимо. $\triangleright$

По существу мы тут используем такой общий факт:

152 Пусть r_n — вычислимая последовательность рациональных чисел. Покажите, что число $\liminf r_n$ является $\mathbf{0}'$-перечислимым снизу и соответствующий $\mathbf{0}'$-алгоритм можно построить, зная алгоритм для r_n.

Отметим кстати, что верно и обратное утверждение:

153 `[liminf-criterion]` Всякое $\mathbf{0}'$-перечислимое снизу действительное число является нижним пределом вычислимой последовательности рациональных чисел.

[Указание. Это число является точной верхней гранью $\mathbf{0}'$-вычислимой последовательности рациональных чисел r_n, каждое из которых является предельным значением стабилизирующейся последовательности $r_{n,k}$. Положим s_k равным максимуму из чисел $r_{0,k}, \ldots, r_{t-1,k}$, где t — минимальное число, для которого $r_{t,k} \ne r_{t,k-1}$.]

Оказывается, что для подходящей последовательности f функция $k \mapsto p_k$ является максимальной $\mathbf{0}'$-перечислимой снизу полумерой. Это вытекает из следующего утверждения:

Theorem 112. `[zero-prime-maximal]` *Для всякой $\mathbf{0}'$-перечислимой снизу последовательности неотрицательных действительных чисел $q_0, q_1, \ldots$ с $\sum_i q_i \leqslant 1$ найдётся вычислимая последовательность $f(0), f(1), \ldots$, нижняя частота появления любого натурального k в которой не меньше q_k.*

Это позволяет дать эквивалентное определение $\mathbf{0}'$-релятивизованной префиксной сложности числа k как минус логарифма нижней частоты числа k в оптимальной последовательности f (той, в которой нижние частоты максимальны с точностью до $O(1)$-множителя).

$\triangleleft$ Перечислимость снизу означает, что множество пар $\langle r, k\rangle$, где r — рациональное число, меньшее q_k, перечислимо (с оракулом $\mathbf{0}'$). Как известно из теории вычислимых функций (см., например, [77]), $\mathbf{0}'$-перечислимые множества составляют класс Σ_2, и потому можно найти разрешимое свойство R, для которого

$$r < q_k \Leftrightarrow \exists u\, \forall v\, R(r, k, u, v)$$

Нам будет удобно иметь дело с несколько другим представлением, а именно, мы выберем вычислимую всюду определённую функцию $\langle r, k, n\rangle \mapsto S(r, k, n)$ со значениями 0 и 1, для которой $r < q_k$ тогда и только тогда, когда в последовательности $S(r, k, 0), S(r, k, 1) \ldots$ конечное число нулей. Последовательность $S(r, k, 0), S(r, k, 1) \ldots$ можно построить так: ищем последовательно для каждого из $u = 0, 1, 2, \ldots$ значение v, при котором $R(r, k, u, v)$ ложно, одновременно дописывая в последовательность единицы, как только такое значение обнаруживается, перемежаем единицы нулём. Конечность числа нулей означает, что для некоторого u такого v не найдётся, то есть что $r < q_k$.

Удобна такая метафора: время от времени для некоторых пар $\langle r, k\rangle$ появляется запрос «хочу, чтобы q_k было больше r», которые время от времени могут «сбрасываться», одновременно возобновляясь. При этом для пар $\langle r, k\rangle$, у которых $r < q_k$, число сбросов конечно (и возобновлённый в какой-то момент запрос уже не будет сброшен), а для остальных пар ($r \geqslant q_k$) запросы сбрасываются бесконечно много раз. (Моменты сбросов запросов соответствуют нулям в последовательности S.) Весь этот процесс вычислим и в каждый момент имеется лишь конечное число запросов (так нам удобно считать, а роли это не играет, поскольку конечность числа нулей в последовательности не зависит от её начального отрезка).

Напомним, что мы хотим построить вычислимую последовательность $f(0), f(1), \ldots$, в которой нижняя частота появления числа k не меньше q_k.

Для этого достаточно представить исходную $\mathbf{0}'$-перечислимую снизу полумеру как нижний предел вычислимой последовательности мер с рациональными значениями, то есть построить вычислимую двумерную таблицу из рациональных чисел

$$\begin{array}{cccc} p_0^0 & p_1^0 & p_2^0 & \ldots \\ p_0^1 & p_1^1 & p_2^1 & \ldots \\ p_0^2 & p_1^2 & p_2^2 & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{array}$$

с такими свойствами: в каждой строке лишь конечное число ненулевых элементов, в сумме равных единице, а нижний предел в k-столбце не меньше q_k. В самом деле, пусть дана такая таблица. Можно считать, что все числа в i-й строке кратны $1/i$ (заменив их на приближения, что не повлияет на предел). Теперь строим последовательность f так: вначале некоторое время руководствуемся первой строкой как таблицей частот, затем переходим ко второй строке и пользуемся ей гораздо большее время (так, чтобы вклад первой строки в частоты стал мал), затем переходим к третьей строке и следуем ей ещё дольше (чтобы забить вклад первой и второй строки) и так далее.

Итак, осталось построить таблицу с таким свойством: если запрос «сделать q_k больше r» с некоторого момента не сбрасывается, то в k-ом столбце нижний предел не меньше q_k. Это делается так: строя n-ю строку таблицы в момент времени n, мы удовлетворяем все имеющиеся к данному моменту запросы (для всех k) в порядке их «стажа» (сколько времени этот запрос уже не сброшен), увеличивая соответствующие p_k до соответствующего r пока это возможно (пока сумма не превзошла единицу; можно считать, что запросов достаточно много и это рано или поздно случится, в этот момент мы обрезаем последний запрос, делая сумму равной единице, и завершаем построение n-й строки).

Почему это гарантирует выполнение условия на нижний предел? Предположим, что $r < q_k$ на самом деле. Тогда запрос «сделать q_k больше r» в некоторый момент появится и уже более никогда не будет сброшен. Посмотрим на запросы, которые появились до него. Некоторые из них будут сброшены, а некоторые — так и не будут. Дождёмся такого момента, когда все эти сбросы произойдут. Тогда среди запросов с большим стажем останутся лишь такие пары r', k', которые никогда не будут сброшены, и потому для них $r' < q_{k'}$ и потому сумма всех таких r' вместе с нашим r меньше единицы. Поэтому запросы с бóльшим стажем не помешают удовлетворить наш запрос, что и требовалось доказать. $\triangleright$

154 Покажите, что можно построить вычислимую последовательность, в которой нижние пределы частот в точности равны q_k.

[Указание. Надо скомбинировать приведённую конструкцию с решением задачи 153.]

155 Докажите, что теорема 112 останется верной, если разрешить рассматривать частичные вычислимые функции f из $\mathbb{N}$ в $\mathbb{N}$ как «последовательности с пробелами»: для любой вычислимой частичной функции f из $\mathbb{N}$ в $\mathbb{N}$ существует вычислимая последовательность $g(0), g(1), \ldots$ (без пробелов) с теми же или бо́льшими нижними частотами: нижняя частота любого числа k в g не меньше его нижней частоты в $f(0), f(1), \ldots$ (определяемой как нижний предел количества появлений числа k среди $f(0), \ldots, f(N-1)$, делённого на N). [Указание. Для каждого N частоты появления в начальном отрезке длины N образуют перечислимую снизу полумеру (вместо вычислимой меры для всюду определённых последовательностей); конструкция из доказательства теоремы 109 позволяет мажорировать нижний предел $\mathbf{0}'$-перечислимой снизу полумерой, от которой уже можно перейти ко всюду определённой функции.]

[Сюда можно было бы дописать про оракулы, не меняющие *KP*, и новое доказательство теоремы Поста о неполных множествах?]

7. Шенноновская энтропия и колмогоровская сложность

[entropy]

7.1. Шенноновская энтропия

[entropy-def]

Пусть нам нужно закодировать буквы некоторого алфавита A, состоящего из k букв $a_1, \ldots, a_k$, двоичными словами (наподобие азбуки Морзе, только вместо точки и тире используются нуль и единица). Пусть буква a_i кодируется некоторым словом c_i. Естественно требовать, чтобы все слова c_i были различны. Но этого мало, если мы хотим записывать коды подряд. Скажем, если алфавит состоит из букв А, Б и В, имеющих коды 0, 1 и 01, то мы не сможем отличить слово АБАБ от слова АБВ: в обоих случаях будет последовательность 0101. (В азбуке Морзе, кстати, такой проблемы нет, поскольку между отдельными буквами делается перерыв — больший, чем между точками и тире внутри буквы). Поэтому надо отдельно позаботиться об однозначности декодирования.

Другой предмет заботы при построении кода — его экономность. Полезно выбрать кодовые слова c_i по возможности более короткими (насколько это возможно при сохранении однозначности декодирования). Более того, если не удаётся сделать короткими все кодовые слова, разумно в первую очередь позаботиться о наиболее часто встречающихся буквах. (Это обстоятельство учитывалось и при составлении азбуки Морзе.)

7.1.1. Коды

[prefix-codes]

Перейдём к формальным определениям. *Кодом* для алфавита A, состоящего из k букв $a_1, \ldots, a_k$, называется набор из k двоичных слов $c_1, \ldots, c_k$. Они называются *кодовыми словами* данного кода; слово c_i называется *кодом* буквы a_i; всякое слово в алфавите A кодируется двоичным словом, получаемым соединением кодов соответствующих букв.

Будем называть код *инъективным*, если коды различных букв различны, и *однозначно декодируемым*, если коды любых двух различных слов различны. Код называется *префиксным*, если ни одно из кодовых слов (соответствующих буквам алфавита A) не является началом (префиксом) другого. (Это название стало традиционным, хотя более логичное — *беспрефиксный* код — также используется.)

Theorem 113. *Всякий префиксный код является однозначно декодируемым.*

$\triangleleft$ Первое кодовое слово (код первой буквы) отщепляется однозначно (в силу префиксности), затем отщепляется код второй буквы и т.п. $\triangleright$

$\boxed{\textbf{156}}$ Покажите, что не всякий однозначно декодируемый код является префиксным. [Указание. Он может быть, например, «суффиксным».]

$\boxed{\textbf{157}}$ Укажите явно взаимно однозначное соответствие между множеством бесконечных последовательностей цифр $0, 1, 2$ и множеством бесконечных последовательностей нулей и единиц. [Указание. Используйте префиксный код $0 \mapsto 00$, $1 \mapsto 01$, $2 \mapsto 1$.]

$\boxed{158}$ Пусть слова $c_1, \ldots, c_k$ и $d_1, \ldots, d_k$ образуют префиксный код (по отдельности). Покажите, что kl слов c_id_j (приписываем одно слово к другому без разделителя) также образуют префиксный код.

Чтобы сравнивать коды по их экономности, нужно фиксировать частоты букв. Пусть даны неотрицательные числа $p_1, \ldots, p_k$, в сумме равные единице; число p_i будем называть *частотой* (или *вероятностью*) буквы a_i. Для каждого кода $c_1, \ldots, c_k$ (для букв $a_1, \ldots, a_k$) определим *среднюю длину* кода как сумму

$$\sum_i p_i l(c_i)$$

Возникает задача: для данных $p_1, \ldots, p_k$ найти код по возможности меньшей средней длины (в том или ином классе кодов).

$\boxed{159}$ Как найти минимальный (с точки зрения средней длины) инъективный код для данного набора $p_1, \ldots, p_n$? [Указание: все буквы надо упорядочить по убыванию частот, а все двоичные слова — в порядке возрастания длин, начиная с пустого.]

7.1.2. Определение шенноновской энтропии

`[entropy-code-length]`

Что можно сказать о минимально возможной длине префиксного кода для данных частот $p_1, \ldots, p_k$? Для ответа на этот вопрос полезна *шенноновская энтропия*. Она определяется для данных $p_1, \ldots, p_k$ (неотрицательных и в сумме равных единице) как величина

$$H = p_1(-\log p_1) + p_2(-\log p_2) + \ldots + p_k(-\log p_k)$$

(при $p = 0$ мы полагаем $p \log p = 0$, доопределяя тем самым функцию $p \mapsto p \log p$ по непрерывности).

Мотивировка этой формулы такова: буква a_i появляется с частотой p_i, а каждое её появление несёт $-\log p_i$ битов информации; в среднем получается H битов на букву. Нужно только объяснить, почему мы считаем, что появление буквы с вероятностью p несёт $-\log p$ битов информации. Это можно сделать так. Пусть задумано одно из 2^n равновозможных чисел. Чтобы его отгадать, надо задать n вопросов (типа да/нет), каждый из которых даёт 1 бит информации. Значит, вероятность $1/2^n$ соответствует n битам информации в событии.

Конечно, последний абзац — это всего лишь метафорический комментарий, позволяющий легче запомнить формулу. Зато следующее утверждение является вполне точным.

Пусть фиксированы неотрицательные числа $p_1, \ldots, p_k$, в сумме равные единице.

Theorem 114. `[prefix-code-length]` (a) *Для любого префиксного кода с кодовыми словами $c_1, \ldots, c_k$ выполняется неравенство*

$$\sum_i p_i l(c_i) \geqslant H$$

(средняя длина кода не меньше энтропии).

(**б**) *Существует префиксный код, для которого*

$$\sum_i p_i l(c_i) < H + 1$$

$\lhd$ Заметим, что в этой теореме реально фигурируют не сами кодовые слова, а их длины. Поэтому важно знать, какие наборы чисел могут быть длинами кодовых слов префиксного кода. Ответ даёт такая лемма:

Лемма (неравенство Крафта).[kraft-lemma] *Пусть фиксированы целые неотрицательные числа $n_1, \ldots, n_k$ и требуется найти двоичные слова $c_1, \ldots, c_k$ указанных длин ($l(c_i) = n_i$), причём так, чтобы ни одно из этих слов не было началом другого. Это возможно тогда и только тогда, когда $\sum_i 2^{-n_i} \leqslant 1$.*

Это утверждение нам уже встречалось, см. леммы в доказательстве теорем 50 (с. 90) и 52 (с. 91). Коротко говоря, если ни одно из слов c_i не является началом другого, то соответствующие интервалы длин 2^{-n_i} не пересекаются, и потому сумма их длин не превосходит единицы. (В других терминах: случайная последовательность нулей и единиц начинается на c_i с вероятностью 2^{-n_i}; эти события несовместны, так как слова несравнимы, и потому сумма вероятностей не превосходит единицы.)

В обратную сторону можно воспользоваться даже более простым способом, чем при доказательстве теоремы 52, поскольку число слов конечно их длины заранее известны. Нужно просто выкладывать соответствующие интервалы длин 2^{-n_i} слева направо на отрезке $[0, 1]$, причём делать это в порядке убывания длин. Тогда каждый интервал будет правильно «выровнен» и ему будет соответствовать двоичное слово длины n_i.

Вернёмся к доказательству теоремы. Можно считать, что все p_i положительны, поскольку нулевые p_i не дают вклада ни в среднюю длину кода, ни в энтропию. В пункте (а) нам надо доказать, что если n_i — неотрицательные целые числа и $\sum_i 2^{-n_i} \leqslant 1$, то сумма $\sum p_i n_i$ не меньше шенноновской энтропии H. Это удобнее доказывать сразу для произвольных n_i (не обязательно целых) и перейдя к другим координатам. Обозначим через q_i величину 2^{-n_i}. В этих координатах утверждение таково: если $q_i > 0$ и $\sum q_i \leqslant 1$, то

$$\sum p_i(-\log q_i) \geqslant \sum p_i(-\log p_i)$$

Это неравенство иногда называют *неравенством Гиббса*[gibbs-inequality]. Чтобы доказать его, заметим, что разница между правой и левой частью равна

$$\sum_i p_i \log \frac{q_i}{p_i}$$

и в силу выпуклости логарифма (взвешенная сумма логарифмов не превосходит логарифма взвешенной суммы: $\sum p_i \log u_i \leqslant \log(\sum_i p_i u_i)$; это верно для любых положительных u_i) не превосходит

$$\log\left(\sum_i p_i \frac{q_i}{p_i}\right) = \log\left(\sum q_i\right) \leqslant \log 1 = 0.$$

Утверждение (а) доказано.

Отметим кстати, что неотрицательную величину

$$\sum_i p_i \log \frac{p_i}{q_i}$$

называют *расстоянием Кульбака – Лейблера* (Kullback – Leibler distance) между распределениями вероятностей p_i и q_i (при этом предполагается, что $\sum q_i = 1$), хотя это «расстояние» и не симметрично. Выпуклость логарифма (отрицательность второй производной) гарантирует, что это расстояние неотрицательно и обращается в нуль, лишь если $p_i = q_i$ при всех i.

Чтобы доказать утверждение (б), рассмотрим числа $n_i = \lceil -\log_2 p_i \rceil$ (где $\lceil u \rceil$ обозначает наименьшее целое число, большее или равное u). Тогда

$$\frac{p_i}{2} < 2^{-n_i} \leqslant p_i$$

Неравенство $2^{-n_i} \leqslant p_i$ гарантирует, что выполнены условия леммы (и потому можно найти кодовые слова соответствующих длин). Неравенство $p_i/2 < 2^{-n_i}$ означает, что n_i превосходит $(-\log p_i)$ менее чем на 1, что сохраняется и после усреднения: средняя длина кода $(\sum p_i n_i)$ превосходит $H = \sum p_i(-\log p_i)$ менее чем на 1. $\triangleright$

Кратко доказательство теоремы можно резюмировать так: если забыть, что длины кодовых слов должны быть целыми, и разрешать любые числа n_i, только бы сумма 2^{-n_i} не превышала единицы, то выгоднее всего взять $n_i = -\log p_i$ (следует из выпуклости логарифма). Требование же целочисленности приводит к увеличению n_i, но не более чем на единицу.

Theorem 115. *Энтропия распределения $p_1, \ldots, p_n$ с n значениями не превосходит $\log n$ и равна $\log n$ в единственном случае, когда все p_i равны.*

$\triangleleft$ Если n есть степень двойки, то неравенство $H \leqslant \log n$ прямо следует из теоремы 114, поскольку можно рассмотреть префиксный код, в котором n кодовых слов имеют длину $\log n$. В общем случае надо применить неравенство Гиббса с $q_i = 1/n$ при всех i и вспомнить, что это неравенство обращается в равенство при $p_i = q_i$. $\triangleright$

7.1.3. Код Хаффмана

`[huffman-encoding]`

Мы доказали, что средняя длина оптимального префиксного кода (для данных $p_1, \ldots, p_k$) заключена между H и $H + 1$. Попробуем разобраться, как найти этот код.

Пусть $n_1, \ldots, n_k$ — длины кодовых слов оптимального кода для данных $p_1, \ldots, p_k$. Будем предполагать (переставив буквы), что

$$p_1 \leqslant p_2 \leqslant \ldots \leqslant p_k.$$

В этом случае

$$n_1 \geqslant n_2 \geqslant \ldots \geqslant n_k$$

(если бы более частая буква кодировалась длиннее, чем более редкая, то обмен кодов уменьшил бы среднюю длину).

Заметим, что для оптимального кода $n_1 = n_2$ (две наиболее редкие буквы всегда имеют одну и ту же длину кода). В самом деле, если $n_1 > n_2$, то n_1 больше всех остальных n_i. Поэтому в сумме $\sum_i 2^{-n_i}$ первое слагаемое меньше всех других, неравенство $\sum_i 2^{-n_i} \leqslant 1$ не может обратиться в равенство по соображениям чётности, и левая часть его меньше правой по крайней мере на 2^{-n_1}. А значит, n_1 можно уменьшить на единицу, не нарушив неравенства $\sum_i 2^{-n_i} \leqslant 1$, и код не является оптимальным.

Поэтому при выборе оптимального кода достаточно ограничиться кодами с $n_1 = n_2$, и оптимальный среди них соответствует минимуму выражения

$$p_1 n_1 + p_2 n_2 + p_3 n_3 + \ldots + p_k n_k = (p_1 + p_2) n + p_3 n_3 + \ldots + p_k n_k$$

(если через n обозначить общее значение n_1 и n_2) по всем $n, n_3, \ldots, n_k$, для которых

$$2^{-n} + 2^{-n} + 2^{-n_3} + \ldots + 2^{-n_k} \leqslant 1.$$

Перепишем это неравенство как

$$2^{-(n-1)} + 2^{-n_3} + \ldots + 2^{-n_k} \leqslant 1,$$

а выражение, подлежащее минимизации, как

$$(p_1 + p_2) + (p_1 + p_2)(n-1) + p_3 n_3 + \ldots + p_k n_k.$$

Член $(p_1 + p_2)$ постоянен и не влияет на поиск минимума, так что задача сводится к поиску оптимального префиксного кода для $k-1$ букв с вероятностями $p_1 + p_2, p_3, \ldots, p_k$.

Получаем рекурсивный алгоритм: соединить две наиболее редкие буквы в одну (сложив вероятности), найти оптимальный префиксный код для этого случая (рекурсивный вызов), а потом вместо одного кодового слова x для соединённой буквы взять два кодовых слова на единицу длиннее ($x0$ и $x1$); ясно, что префиксность кода при этом не нарушится.

Построенный с помощью такого алгоритма оптимальный префиксный код называется *кодом Хаффмана* для данного набора вероятностей p_i.

7.1.4. Неравенство Крафта – Макмиллана

[kraft-mcmillan]

До сих пор мы изучали в основном префиксные коды. Оказывается, что переход к произвольным однозначно декодируемым кодам ничего не даёт (с точки зрения сокращения кода):

Theorem 116 (неравенство Макмиллана). [mcmillan-inequality] *Пусть $c_1, \ldots, c_k$ — кодовые слова однозначно декодируемого кода, а $n_i = l(c_i)$ — их длины. Тогда*

$$\sum_i 2^{-n_i} \leqslant 1.$$

Тем самым (лемма Крафта) можно построить и префиксный код с теми же длинами слов.

$\triangleleft$ Будем считать, что в кодовых словах вместо цифр 0 и 1 используются буквы u и v. (Скажем, кодовые слова 0, 01 и 11 мы запишем как u, uv и vv.) Напишем формальную сумму $(c_1 + \ldots + c_k)$ всех кодовых слов, возведём её в N-ю степень (число N мы потом выберем) и раскроем скобки, не переставляя u и v (как если бы они не коммутировали). Например, для $N = 2$ и для приведённого выше примера получится

$$(u + uv + vv)(u + uv + vv) = uu + uuv + uvv + uvu + uvuv + uvvv + vvu + vvuv + vvvv.$$

Каждое слагаемое в правой части есть соединение некоторых кодовых слов, причём все слагаемые различны (свойство однозначности декодирования). Теперь подставим вместо u и v число $1/2$. В левой части $(c_1 + \ldots + c_k)^N$ превратится при этом в $(2^{-n_1} + \ldots + 2^{-n_k})^N$. Правую часть оценим сверху: если бы в неё входили все возможные слова данной длины t, то получилось бы 2^t членов, каждый из которых равен 2^{-t}, и сумма равнялась бы 1 (для каждой длины). Поэтому сумма в правой части не превосходит максимальной длины слагаемых, то есть, не больше $N \max(n_i)$.

Теперь видно, что если $\sum 2^{-n_i} > 1$, то при больших N левая часть (растущая экспоненциально) становится больше правой (растущей линейно). $\triangleright$

Это доказательство производит впечатление искусственного (хотя и красивого) трюка. Более естественное доказательство (или, если угодно, более естественное изложение того же доказательства) будет приведено ниже (с. 193).

7.2. Энтропия пары и условная энтропия

[entropy-pair]

7.2.1. Энтропия пары случайных величин

[entropy-pair-definition] При обсуждении энтропии удобно использовать стандартную для теории вероятностей терминологию. Пусть ξ — случайная величина, принимающая конечное число значений $\xi_1, \ldots, \xi_k$ с вероятностями $p_1, \ldots, p_k$. Тогда её *шенноновская энтропия* определяется формулой

$$H(\xi) = p_1(-\log p_1) + \ldots + p_k(-\log p_k)$$

Это определение позволяет говорить об энтропии пары случайных величин ξ и η (определённых на одном и том же вероятностном пространстве), поскольку такая пара сама образует случайную величину. Следующая теорема утверждает, что энтропия пары не превосходит суммы энтропий:

Theorem 117. [entropy-pair-bound]

$$H(\langle \xi, \eta \rangle) \leqslant H(\xi) + H(\eta)$$

Мы предполагаем, что величины ξ и η принимают конечное число значений, поэтому эта теорема представляет собой некоторое неравенство с суммами логарифмов. Именно, пусть ξ принимает k значений $\xi_1, \ldots, \xi_k$, а η принимает l значений $\eta_1, \ldots, \eta_l$. Тогда величина $\langle \xi, \eta \rangle$ может принимать, вообще говоря, kl значений $\langle \xi_i, \eta_j \rangle$ (некоторые из значений могут не встречаться или встречаться с вероятностью нуль). Распределение вероятностей для пары $\langle\xi,\eta\rangle$, таким образом, задаётся таблицей из k строк и l столбцов: число p_{ij}, стоящее в i-й строке и j-м столбце, представляет собой вероятность события «$(\xi=\xi_i)$ и $(\eta=\eta_j)$» (здесь $i=1,\ldots,k$ и $j=1,\ldots,l$). Все числа p_{ij} неотрицательны и в сумме равны единице (некоторые из них могут равняться нулю).

Сложив числа в строках, мы получим распределение вероятностей для величины ξ: она принимает значение ξ_i с вероятностью $\sum_j p_{ij}$; эту сумму удобно обозначить p_{i*}; аналогичным образом η принимает значение η_j с вероятностью p_{*j}, которая есть сумма чисел в j-м столбце.

Таким образом, сформулированная теорема представляет собой неравенство, справедливое для любой прямоугольной таблицы с неотрицательными числами, в сумме равными единице:

$$\sum_{i,j} p_{ij}(-\log p_{ij}) \leqslant \sum_i p_{i*}(-\log p_{i*}) + \sum_j p_{*j}(-\log p_{*j})$$

(где p_{i*} и p_{*j} определяются как суммы по строкам и столбцам).

Это неравенство в конечном счёте сводится к выпуклости логарифма, но полезно понимать его интуитивный смысл. Если отождествить (забыв про разницу порядка единицы) энтропию с длиной кратчайшего префиксного кода, то теорему можно доказать так: пусть имеются короткие префиксные коды для ξ и η (с кодовыми словами $c_1,\ldots,c_k$ и $d_1,\ldots,d_l$). Тогда можно рассмотреть код для пары $\langle\xi,\eta\rangle$, кодируя значения $\langle\xi_i,\eta_j\rangle$ словом c_id_j (приписываем d_j справа к c_i без разделителя). Это будет, как легко проверить, префиксный код (чтобы отщепить кодовое слово от бесконечной последовательности, надо сначала отщепить c_i, а потом d_j; в обоих случаях это делается однозначно). Средняя длина этого кода будет равна сумме средних длин кодов для ξ и η. Он не обязан быть оптимальным (ведь и неравенство может быть строгим), но даёт оценку сверху для оптимального кода.

$\lhd$ Это рассуждение можно превратить в строгое доказательство, если вспомнить, что при доказательстве теоремы 114 (с. 184) мы установили, что энтропия равна минимуму величины $\sum_i p_i(-\log_2 q_i)$ по всем наборам неотрицательных чисел q_i с единичной суммой. В частности, энтропия пары (левая часть неравенства) есть минимум сумм

$$\sum_{i,j} p_{ij}(-\log q_{ij})$$

по всем наборам q_{ij} неотрицательных чисел с суммой единица. Будем рассматривать не все наборы, а лишь наборы «ранга 1», которые получаются как произведения

$$q_{ij} = q_{i*}\cdot q_{*j}$$

для некоторых наборов неотрицательных чисел q_{i*} и q_{*j}, каждый из которых имеет сумму 1. Тогда $(-\log q_{ij})$ распадётся в сумму $(-\log q_{i*})+(-\log q_{*j})$, а вся сумма — в две суммы, которые (после суммирования по одному из индексов) окажутся равными

$$\sum_i p_{i*}(-\log q_{i*})$$

и

$$\sum_j p_{*j}(-\log q_{*j})$$

соответственно. Минимумы этих сумм равны $H(\xi)$ и $H(\eta)$.

Таким образом, левая часть неравенства есть минимум некоторой величины по всем наборам, а правая — по наборам ранга 1, откуда и вытекает требуемое неравенство. $\triangleright$

7.2.2. Условная энтропия

`[conditional-entropy-definition]` *Условной вероятностью* некоторого события B при условии события A называют отношение вероятности события «A и B» к вероятности события A. Это определение имеет смысл, если вероятность события A отлична от нуля. Мотивировка понятна: мы рассматриваем долю исходов, где произошло B, не среди всех исходов, а только среди тех, где произошло A.

Если A — событие, а ξ — случайная величина с конечным числом значений $\xi_1, \ldots, \xi_k$, то можно рассмотреть (помимо вероятностей $\Pr[\xi = \xi_i]$) и условные вероятности $\Pr[(\xi = \xi_i)|A]$. Их сумма тоже равна единице, и получается некоторое новое распределение вероятностей. Его энтропия называется *условной энтропией величины ξ при условии A* и обозначается $H(\xi|A)$, а само это распределение вероятностей можно обозначить $(\xi|A)$.

$\boxed{\mathbf{160}}$ Покажите, что величина $H(\xi|A)$ может быть и больше, и меньше величины $H(\xi)$. [Указание: распределение $(\xi|A)$ (особенно при малой вероятности события A) мало связано с распределением вероятностей для ξ.]

Неформально говоря, $H(\xi|A)$ — это минимально возможная средняя длина кода, если нас интересуют лишь случаи, когда произошло событие A.

Пусть теперь (как и в прошлом разделе) даны две случайные величины ξ и η. Будем предполагать, что для каждой из них все значения имеют ненулевую вероятность (нулевые можно выбросить). Для каждого значения η_j величины η рассмотрим событие $\eta = \eta_j$ (его вероятность мы обозначали p_{*j}). Рассмотрим условную энтропию величины ξ при условии этого события. Она соответствует распределению вероятностей $i \mapsto p_{ij}/p_{*j}$. Далее усредним эти энтропии с весами, равными вероятностям событий $\eta = \eta_j$. Полученное среднее называют *условной энтропией ξ при известном η* и обозначают $H(\xi|\eta)$. Формально говоря,

$$H(\xi|\eta) = \sum_j \Pr[\eta = \eta_j] H(\xi|\eta = \eta_j)$$

или, в наших обозначениях,

$$H(\xi|\eta) = \sum_j p_{*j} \sum_i \frac{p_{ij}}{p_{*j}} \left(-\log \frac{p_{ij}}{p_{*j}}\right)$$

Основные свойства условной энтропии перечислены в следующей теореме, справедливой для любых случайных величин ξ, η:

Theorem 118. `[conditional-entropy]`

(а) *$H(\xi|\eta) \geqslant 0$;*

(б) *$H(\xi|\eta) = 0$ тогда и только тогда, когда $\xi = f(\eta)$ с вероятностью 1 для некоторой функции f (оговорка про «вероятность 1» означает, что мы пренебрегаем значениями, которые имеют нулевую вероятность).*

(в) *$H(\xi|\eta) \leqslant H(\xi)$*

(г) *$H(\langle\xi, \eta\rangle) = H(\eta) + H(\xi|\eta)$*

$\lhd$ Первое из этих утверждений очевидно: все $H(\xi|\eta = \eta_j)$ неотрицательны, потому неотрицательна и их взвешенная сумма.

(б) Если взвешенная сумма равна нулю, то все слагаемые с ненулевыми коэффициентами равны нулю, то есть при каждом значении η_j величина $(\xi|\eta = \eta_j)$ имеет нулевую энтропию (и потому принимает лишь одно значение с точностью до событий нулевой вероятности).

Утверждение (в) можно объяснить так: $H(\xi|\eta)$ будет средней длиной оптимального кода для ξ, если разрешить кодировать значения ξ по-разному, в зависимости от значения величины η (в каждом случае свой код, который оптимизируется с учётом условных вероятностей). Ясно, что это облегчает построение оптимального кода, поэтому средняя длина получается меньше $H(\xi)$.

Более формально: при каждом j величина $H(\xi|\eta = \eta_j)$ равна минимуму суммы

$$\sum_i \frac{p_{ij}}{p_{*j}}(-\log q_{ij})$$

по всем неотрицательным $q_{1j} + q_{2j} + \ldots + q_{kj} = 1$ (мы используем свой набор переменных для каждого j) и потому $H(\xi|\eta)$ равна минимуму суммы

$$\sum_j p_{*j} \sum_i \frac{p_{ij}}{p_{*j}}(-\log q_{ij})$$

по всем таблицам, составленным из неотрицательных чисел q_{ij}, у которых сумма каждого столбца равна единице. Если мы теперь ограничимся таблицами, у которых все столбцы одинаковы, $q_{ij} = q_i$, то сумма превратится в

$$\sum_j p_{*j} \sum_i \frac{p_{ij}}{p_{*j}}(-\log q_i) = \sum_j \sum_i p_{ij}(-\log q_i) = \sum_i p_{i*}(-\log q_i)$$

и её минимум станет равным $H(\xi)$. Поэтому $H(\xi|\eta) \leqslant H(\xi)$.

Наконец, пункт (г) представляет собой равенство, которое непосредственно следует из определений:

$$\begin{gathered}\sum_{i,j} p_{ij}(-\log p_{ij}) = \sum_j p_{*j} \sum_i \frac{p_{ij}}{p_{*j}}(-\log \frac{p_{ij}}{p_{*j}} - \log p_{*j}) = \\ \sum_j p_{*j} \sum_i \frac{p_{ij}}{p_{*j}}(-\log \frac{p_{ij}}{p_{*j}}) + \sum_j p_{*j} \sum_i \frac{p_{ij}}{p_{*j}}(-\log p_{*j}) = \\ \sum_j p_{*j} H(\xi|\eta = \eta_j) + \sum_j p_{*j}(-\log p_{*j}) = H(\xi|\eta) + H(\eta).\end{gathered}$$

Теорема доказана. $\rhd$

Из этой теоремы немедленно вытекает теорема 117 (с. 188). Кроме того, из неё видно, что энтропия пары не меньше энтропии любого из её членов (поскольку условная энтропия неотрицательна). Отсюда легко вытекает такое утверждение:

Theorem 119. [no-new-entropy] *Пусть ξ — случайная величина с конечным числом значений, а f — функция, определённая на множестве значений величины ξ. Тогда*

$$H(f(\xi)) \leqslant H(\xi),$$

где $f(\xi)$ — случайная величина, получающаяся применением f к ξ (формально говоря, композиция f и ξ).

С точки зрения наборов чисел переход от ξ к $f(\xi)$ означает, что мы объединяем некоторые значения (складывая соответствующие вероятности).

$\triangleleft$ В самом деле, величина $\langle \xi, f(\xi) \rangle$ имеет в точности то же распределение, что и ξ, поэтому её энтропия не меньше энтропии второго члена пары. $\triangleright$

$\boxed{161}$ Укажите прямое доказательство в терминах кодирования и поиска минимума.

$\boxed{162}$ В каких случаях неравенство теоремы 119 обращается в равенство?

7.2.3. Независимость и энтропия

[independence-and-entropy] Понятие независимости случайных величин легко выражается в терминах энтропии. Напомним, что величины ξ и η *независимы*, если вероятность события «$\xi = \xi_i$ и $\eta = \eta_j$» равна произведению вероятностей событий $\xi = \xi_i$ и $\eta = \eta_j$ по отдельности. (Переформулировка: если распределение вероятностей ξ относительно условия $\eta = \eta_j$ совпадает с исходным; аналогично и для η относительно $\xi = \xi_i$.) В наших обозначениях независимость означает, что $p_{ij} = p_{i*}p_{*j}$ (матрица вероятностей имеет ранг 1).

Theorem 120. [independence] *Величины ξ и η независимы тогда и только тогда, когда*

$$H(\langle \xi, \eta \rangle) = H(\xi) + H(\eta).$$

Другими словами, критерий независимости состоит в том, что неравенство теоремы 117 обращается в равенство. Используя теорему 118, можно переписать это равенство в виде $H(\xi) = H(\xi|\eta)$ или $H(\eta) = H(\eta|\xi)$.

$\triangleleft$ Логарифм является строго выпуклой функцией: неравенство

$$\log\left(\sum p_i x_i\right) \geqslant \sum p_i \log x_i,$$

справедливое для неотрицательных p_i с единичной суммой и произвольных положительных x_i, обращается в равенство, лишь если все x_i равны (за исключением тех, которые входят с нулевыми коэффициентами p_i).

Отсюда следует, что для любых неотрицательных p_i, в сумме равных единице, минимум выражения

$$\sum p_i(-\log q_i)$$

который берётся по всем неотрицательным q_i, в сумме равным 1, достигается в единственной точке, когда $q_i = p_i$. (Надо уточнить, что при $p_i = 0$ мы полагаем $p_i(-\log q_i) = 0$ и разрешаем q_i быть нулевым.)

Теперь вспомним, что делалось при доказательстве теоремы 117. Минимум по матрицам ранга 1 (при котором правая часть равна сумме энтропий) достигался при

$$q_{ij} = p_{i*} \cdot p_{*j}$$

Если он совпадает с минимумом по всем наборам q_{ij}, который достигается при $q_{ij} = p_{ij}$, то это значит, что имеет место равенство

$$p_{ij} = p_{i*} \cdot p_{*j}$$

и величины ξ и η независимы. $\triangleright$

163 Проведите аналогичное рассуждение, основываясь на теореме 118.

164 Докажите, что величины α, β, γ независимы в совокупности (вероятность события $(\alpha = \alpha_i, \beta = \beta_j, \gamma = \gamma_k)$ равна произведению трёх отдельных вероятностей) тогда и только тогда, когда

$$H(\langle \alpha, \beta, \gamma \rangle) = H(\alpha) + H(\beta) + H(\gamma).$$

Теоремы 117 и 120 показывают, что разность $H(\xi) + H(\eta) - H(\langle \xi, \eta \rangle)$ всегда неотрицательна и обращается в нуль тогда и только тогда, когда величины ξ и η независимы. Тем самым логично считать её числовой мерой «степени зависимости» между ξ и η. Эту разность обозначают $I(\xi : \eta)$ и называют *взаимной информацией* случайных величин ξ и η. Теорема 118 позволяет переписать выражение для $I(\xi : \eta)$ так:

$$I(\xi : \eta) = H(\eta) - H(\eta|\xi) = H(\xi) - H(\xi|\eta).$$

(взаимная информация показывает, «насколько знание одной величины уменьшает энтропию другой»).

В качестве применения рассмотренных понятий снова докажем неравенство Макмиллана.`[macmillan-revisited]` Мы будем действовать немного в другом порядке и сначала докажем, что для однозначно декодируемого кода случайной величины ξ средняя длина кодового слова не меньше $H(\xi)$.

Чтобы сделать это, заметим, что для инъективного кода, у которого все кодовые слова имеют длину меньше c, средняя длина не меньше $H(\xi) - \log c$. В самом деле, если n_i — длины кодовых слов такого кода, то сумма величин 2^{-n_i} не превосходит c (для каждой длины сумма по словам этой длины не превосходит единицы), и потому неравенство теоремы 114 ухудшается не более чем на $\log c$.

Теперь рассмотрим N независимых одинаково распределённых копий случайной величины ξ. Возникающая случайная величина, которую мы обозначим ξ^N, имеет энтропию в N раз больше. Если использовать для кодирования каждой из копий наш однозначно декодируемый код, а потом все слова соединить, то получится код для ξ^N, средняя длина которого в N раз больше средней длины исходного кода для ξ. Этот код будет инъективным (по определению однозначного декодирования), и максимальная длина его не превосходит cN, где c — максимальная длина кодовых слов рассматриваемого нами однозначно декодируемого кода. Применяя утверждение предыдущего абзаца, получаем, что

$$N \cdot (\text{средняя длина однозначно декодируемого кода}) \geqslant NH(\xi) - \log(cN)$$

Теперь, поделив на N и устремив N к бесконечности, получаем искомое утверждение (поскольку $\log(cN)/N$ стремится к 0).

Теперь уже легко доказать собственно неравенство Макмиллана. Пусть однозначно декодируемый код состоит из слов длиной $n_1, \ldots, n_k$ и $\sum 2^{-n_i} > 1$. Положим сначала $p_i = 2^{-n_i}$, а затем пропорционально уменьшим p_i, сделав их сумму равной единице. Рассмотрим случайную величину с распределением вероятностей p_i и её кодирование с помощью нашего кода. Средняя длина кода будет $\sum p_i n_i$ и будет меньше $H = \sum p_i(-\log p_i)$, поскольку $n_i < -\log p_i$ за счёт уменьшения величин p_i.

165 Проведите подробно это доказательство неравенства Макмиллана и установите его соответствие с ранее приведённым.

7.2.4. «Релятивизация» и базисные неравенства

[entropy-relativization] Доказанные нами утверждения имеют свои «условные» варианты. Например, неравенство

$$H(\langle\xi,\eta\rangle) \leqslant H(\xi) + H(\eta)$$

при добавлении случайной величины α в качестве условия превращается в

$$H(\langle\xi,\eta\rangle|\alpha) \leqslant H(\xi|\alpha) + H(\eta|\alpha)$$

Нового доказательства по существу не требуется, так как при каждом значении α_i случайной величины α выполнено неравенство

$$H(\langle\xi,\eta\rangle|\alpha=\alpha_i) \leqslant H(\xi|\alpha=\alpha_i) + H(\eta|\alpha=\alpha_i)$$

(применяем теорему 117 к условным распределениям вероятностей случайных величин ξ и η), и остаётся сложить эти неравенства с весами $\Pr[\alpha=\alpha_i]$.

Теперь можно выразить условные энтропии через безусловные, используя формулу $H(\beta|\gamma) = H(\langle\beta,\gamma\rangle) - H(\gamma)$, и привести подобные члены. Получится такое утверждение:

Theorem 121 (базисное неравенство). [basic-shannon]

$$H(\xi,\eta,\alpha) + H(\alpha) \leqslant H(\xi,\alpha) + H(\eta,\alpha)$$

Для краткости мы опускаем угловые скобки и пишем $H(\xi,\eta,\alpha)$ вместо $H(\langle\xi,\eta,\alpha\rangle)$ или ещё более подробного $H(\langle\langle\xi,\eta\rangle,\alpha\rangle)$.

Аналогичную «релятивизацию» (добавление случайных величин как условий) можно применить и к понятию взаимной информации и определить, скажем, $I(\alpha:\beta|\gamma)$ как

$$H(\alpha|\gamma) + H(\beta|\gamma) - H(\langle\alpha,\beta\rangle|\gamma).$$

Базисное неравенство (теорема 121) утверждает, что $I(\alpha:\beta|\gamma) \geqslant 0$ для любых случайных величин α,β,γ.

166 Докажите, что $I(\langle\alpha,\beta\rangle:\gamma) \geqslant I(\alpha:\gamma)$

$\boxed{\textbf{167}}$ Докажите, что

$$I(\langle \alpha, \beta \rangle : \gamma) = I(\alpha : \gamma) + I(\beta : \gamma | \alpha).$$

Если $I(\alpha : \gamma | \beta) = 0$, говорят, что α и γ *независимы при известном* β, а также что α, β, γ образуют *марковскую цепь* («прошлое» α связано с «будущим» γ лишь через «настоящее» β).

$\boxed{\textbf{168}}$ Докажите, что в этом случае $I(\alpha : \gamma) \leqslant I(\alpha : \beta)$, и потому $I(\alpha : \gamma) \leqslant H(\beta)$.

При решении этих задач полезно использовать диаграммы, аналогичные диаграммам для колмогоровской сложности в главе 2. Диаграмма для двух величин состоит из трёх областей, каждой из которых соответствует неотрицательное значение; суммы значений в двух областях слева равна $H(\alpha)$, а в двух областях справа — $H(\beta)$ (рис.18).

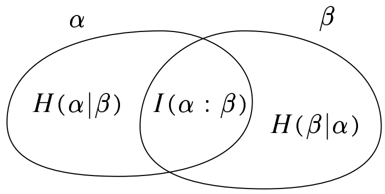

Рис. 18. Энтропии двух случайных величин.

[entropy.1]

Диаграмма для трёх величин α, β, γ показана на рис. 19. Центральной области соответствует значение, которое мы обозначили через $I(\alpha : \beta : \gamma)$; его можно определить как $I(\alpha : \beta) - I(\alpha : \beta | \gamma)$, а также как $I(\alpha : \gamma) - I(\alpha : \gamma | \beta)$ и т.п. — при переходе к безусловным энтропиям получается выражение

$$I(\alpha : \beta : \gamma) = H(\alpha) + H(\beta) + H(\gamma) - H(\alpha, \beta) - H(\alpha, \gamma) - H(\beta, \gamma) + H(\alpha, \beta, \gamma)$$

В отличие от шести других величин на рисунке, $I(\alpha : \beta : \gamma)$ может быть отрицательной. Так будет, например, если величины α и β независимы, но зависимы при известном γ.

$\boxed{\textbf{169}}$ Приведите пример таких величин α, β, γ. [Указание. По аналогии с примером на с. 52 можно рассмотреть независимые величины α и β, равномерно распределённые на $\{0, 1\}$, и положить $\gamma = \alpha + \beta \bmod 2$.]

$\boxed{\textbf{170}}$ Докажите, что если величины α и β различаются с вероятностью $\varepsilon < 1/2$, и число различных значений величины α равно a, то выполняется *неравенство Фано*:

$$H(\alpha|\beta) \leqslant \varepsilon \log a + h(\varepsilon),$$

где $h(\varepsilon)$ — энтропия случайной величины с двумя значениями, имеющими вероятности ε и $1 - \varepsilon$. [Указание. Введём величину γ, которая принимает два значения 0 и 1 при $\alpha \neq \beta$ и $\alpha = \beta$ соответственно. Тогда $H(\alpha|\beta) \leqslant H(\gamma) + H(\alpha|\beta, \gamma)$. Первое слагаемое равно $h(\varepsilon)$, а второе надо записать как

$$\Pr[\gamma = 0] H((\alpha|\beta)|\gamma = 0) + \Pr[\gamma = 1] H((\alpha|\beta)|\gamma = 1),$$

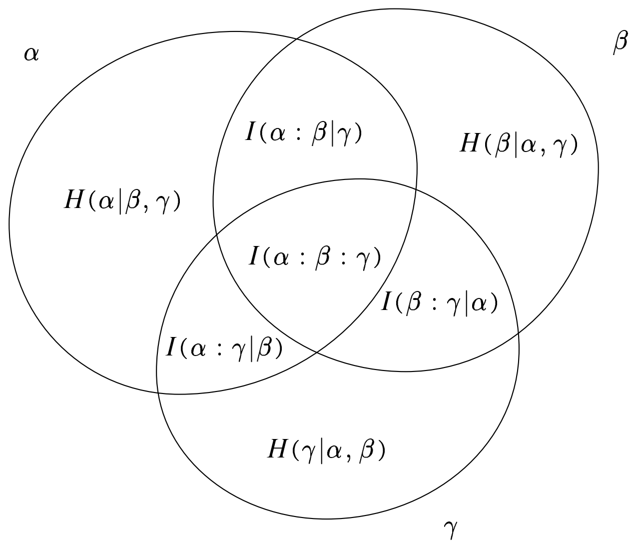

Рис. 19. Энтропии трёх случайных величин.

[entropy.2]

то есть

$$\Pr[\alpha \neq \beta] H((\alpha|\beta)|\alpha \neq \beta) + \Pr[\alpha = \beta] H((\alpha|\beta)|\alpha = \beta),$$

что не превосходит $\varepsilon \log a + 0$.]

171 Пусть $H(\alpha|\beta,\gamma) = 0$ и $I(\beta:\alpha) = 0$. Докажите, что $H(\gamma) \geqslant H(\alpha)$. (Если агент хочет передать в Центр секретное сообщение α в виде открытого текста β с помощью заранее согласованного с Центром ключа γ, причём так, чтобы враги, не знающие γ, не получили никакой информации об α, то энтропия ключа должна быть не меньше энтропии сообщения. Это утверждение называют иногда *теоремой Шеннона об идеальном шифре*.)

172 [condit-triple-entropy] Докажите, что для любых трёх случайных величин α, β, γ выполнено неравенство

$$2H(\alpha,\beta,\gamma) \leqslant H(\alpha,\beta) + H(\beta,\gamma) + H(\alpha,\gamma).$$

[Указание: см. доказательство аналогичного утверждения для колмогоровской сложности в теореме 26, (с. 50).]

7.3. Сложность и энтропия

[complexity-and-entropy] Рассмотренные нами свойства шенноновской энтропии, условной энтропии и взаимной информации (для случайных величин) сходны с аналогичными утверждениями о колмогоровской сложности (см. главу 2). Можно ли каким-либо образом уточнить и формализовать эти аналогии?

Вопрос этот можно понимать двояко. Во-первых, можно доказывать, что колмогоровская сложности и шенноновская энтропия похожи по своим свойствам (скажем, для них верны одни и те же неравенства, см. раздел 10.6, с. 290). С другой стороны, можно сравнивать численные значения колмогоровской сложности и шенноновской энтропии, и этим мы сейчас займёмся. Правда, тут сразу же возникает проблема: шенноновская энтропия определена для случайных величин, а колмогоровская сложность — для двоичных слов, как же их сравнивать? Тем не менее иногда такое сравнение возможно, и мы сейчас разберём несколько утверждений такого вида. Общее и весьма расплывчатое описание соответствующих результатов звучит так: шенноновская энтропия учитывает лишь частотные закономерности, а колмогоровская сложность — все алгоритмические закономерности, поэтому в общем случае колмогоровская сложность меньше. Однако в тех случаях, если объект порождается случайно по вероятностной мере, с большой вероятностью других (не частотных) закономерностей нет, и сложность близка к энтропии.

Перейдём к конкретным уточнениям этого общего тезиса.

7.3.1. Колмогоровская сложность и энтропия частот

[frequencies-entropy] Будем рассматривать произвольный (не обязательно двухбуквенный) конечный алфавит A и слова в этом алфавите. Колмогоровская сложность для них определяется естественным образом, и мы будем говорить о ней без особых оговорок. (Мы нигде не использовали, что объекты, сложность которых определяется, являются именно двоичными словами. Но важно, что мы рассматриваем двоичные слова в качестве описаний: измеренная в байтах сложность будет в восемь раз меньше битовой!)

Пусть фиксирован алфавит A, содержащий k букв. Рассмотрим произвольное слово x некоторой длины N в этом алфавите. Пусть $p_1, \ldots, p_k$ — частоты появления букв в слове x. Все они являются дробями со знаменателем N; сумма частот равна единице. Пусть $h(p_1, \ldots, p_k)$ — шенноновская энтропия соответствующего распределения.

Theorem 122. [complexity-le-entropy]

$$\frac{\mathit{KS}(x)}{N} \leqslant h(p_1, \ldots, p_k) + \frac{O(\log N)}{N}$$

Имеется в виду, что $O(\log N)$ не больше $c \log N$, причём константа c не зависит ни от N, ни от слова x, ни от частот $p_1, \ldots, p_k$, но может зависеть от k, так что в этой теореме размер алфавита считается фиксированным.

$\lhd$ По существу это чисто комбинаторное утверждение. В самом деле, $\mathit{KS}(x|N, p_1, \ldots, p_k)$ не превосходит $\log C(N, p_1, \ldots, p_k) + O(1)$, где

$$C(N, p_1, \ldots, p_k) = \frac{N!}{(p_1 N)!(p_2 N)! \ldots (p_k N)!}$$

есть количество слов длины N в алфавите A, имеющих данные частоты $p_1, \ldots, p_k$. Это следует из того, что каждое слово с такими частотами может быть (при известных параметрах) задано своим порядковым номером в (скажем, алфавитном) списке таких слов, для чего достаточно $\log C(N, p_1, \ldots, p_k)$ битов.

Оценка числа $C(N, p_1, \ldots, p_k)$ выполняется по формуле Стирлинга. Если отбросить полиномиальные по N множители (возникающие из-за $\sqrt{2\pi k}$ в формуле для $k!$), то останется как раз $2^{Nh(p_1, \ldots, p_k)}$. Подробно это вычисление для случая $k = 2$ проводилось при доказательстве усиленного закона больших чисел (теорема 27, с. 57); общий случай аналогичен.

Остаётся заметить, что для задания $N, p_1, \ldots, p_k$ достаточно примерно $k \log N$ битов (нужно указать k целых чисел $p_1 N, \ldots, p_k N$, в сумме дающих N), поэтому при удалении условия в выражении $KS(x|N, p_1, \ldots, p_k)$ сложность возрастёт не более чем на $O(\log N)$ (с константой, примерно равной k). $\triangleright$

Другой способ доказательства того же утверждения использует верхнюю оценку для монотонной сложности (теорема 81, с. 134). Рассмотрим распределение вероятностей на последовательностях букв алфавита A, соответствующее независимым испытаниям с частотами $p_1, \ldots, p_k$. Вероятность появления последовательности, начинающейся на слово z длины N, в котором частоты букв равны $q_1, \ldots, q_k$, равна

$$p_1^{q_1 N} \ldots p_k^{q_k N}$$

(буква a_i имеет вероятность p_i и появляется $q_i N$ раз), а её двоичный логарифм равен

$$-N \cdot (q_1(-\log p_1) + \ldots + q_k(-\log p_k))$$

В частности, если $q_i = p_i$, то логарифм вероятности равен $-Nh(p_1, \ldots, p_k)$ и потому монотонная сложность оценивается сверху числом $Nh(p_1, \ldots, p_k)$. Остаётся заметить, что монотонная сложность отличается от обычной на величину порядка $O(\log N)$ для слов длины N.

Это рассуждение, однако, не вполне корректно. Дело в том, что в оценке для монотонной сложности предполагалось, что мера фиксирована и оцениваются сложности различных слов. Здесь же, интересуясь оценкой для сложности слова x, мы рассматриваем меру, которая соответствует частотам появления букв в само́м слове x, поэтому формально утверждение теоремы 81 применить нельзя. Если вспомнить её доказательство, то видно, что получается оценка для «условной» монотонной сложности при известных $p_1, \ldots, p_k$; переход от неё к безусловной сложности может привести к увеличению оценки на $O(\log N)$, так что таким способом мы получаем другое доказательство теоремы 122.

$\boxed{173}$ Оцените зависимость константы в $O(\log N)$ от k. [Указание: оба доказательства дают $k(1 + o(1)) \log N$.]

$\boxed{174}$ Покажите, что при отделённых от нуля частотах $p_1, \ldots, p_k$ оценку предыдущей задачи можно улучшить до $(k/2 + O(1)) \log N$. [Указание. В первом доказательстве надо вспомнить об опущенных квадратных корнях в формуле Стирлинга, которые в основном попадают в знаменатель. Второе доказательство также можно модифицировать, рассмотрев меру, в которой вместо точных значений p_k берутся их приближения с точностью до $1/\sqrt{N}$. При этом оценка ухудшится, но поскольку в окрестности минимума приращение гладкой функции пропорционально квадрату аргумента, то это ухуждение будет не более чем на $O(1)$; на этом мы сэкономим половину битов при задании чисел $p_1, \ldots, p_k$.]

[Тут требуются какие-то уточнения: если k велико, то что означает, что все частоты отделены от нуля?!]

Заметим, что неравенство теоремы 122 может быть как угодно далеко от равенства: если, скажем, алфавит состоит из двух букв, и в слове x они чередуются, то правая часть равна 1, а левая имеет порядок $(\log N)/N$. Что и не удивительно — правая часть учитывает только частоты, а не другие (не-частотные) закономерности. В следующем разделе мы покажем, что при случайном порождении слова сложность близка к энтропии.

7.3.2. Математическое ожидание сложности

Пусть фиксирован k-буквенный алфавит A и k неотрицательных чисел $p_1, \ldots, p_k$ с суммой 1 (которые мы для простоты считаем рациональными).

Рассмотрим случайную величину ξ, значениями которой являются буквы алфавита A, принимаемые с вероятностями $p_1, \ldots, p_k$. Для каждого N рассмотрим случайную величину ξ^N, соответствующую N независимым копиям случайной величины ξ. Её значениями являются слова длины N в алфавите A (каждая буква порождается независимо, причём вероятность появления i-ой буквы равна p_i). Нас будет интересовать математическое ожидание сложности значений случайной величины ξ^N (взвешенное среднее с весами, равными вероятностям).

Theorem 123. [expected-complexity] *Математическое ожидание величины* $\mathit{KP}(\xi^N|N)$ *равно* $NH(\xi) + O(1)$ *(константа в* $O(1)$ *может зависеть от* ξ*, но не зависит от* N*).*

Заметим, что (при положительных p_i) среди значений величины ξ^N будут встречаться все слова длины N в алфавите A; некоторые из них имеют сложность много больше NH (если только распределение вероятностей не равномерное, поскольку в последнем случае таких слов нет), а другие имеют сложность много меньше NH.

$\triangleleft$ Для каждого слова длины N (то есть для каждого значения случайной величины ξ^N рассмотрим его кратчайшее описание при известном N (относительно оптимального префиксно корректного способа описания). Полученные слова образуют префиксный код в смысле раздела 7.1.1. Средняя длина этого кода как раз равна математическому ожиданию сложности значений величины ξ^N. Поэтому теорема 114 (с. 184) гарантирует, что это математическое ожидание не меньше $H(\xi^N) = NH(\xi)$. Нижняя оценка (и даже без константы $O(1)$) доказана.

Та же самая теорема позволяет получить и верхнюю оценку. В самом деле, она утверждает существование префиксного кода, имеющего среднюю длину кодового слова не выше $H + 1$. Такой код можно алгоритмически построить, зная число N (и числа p_i, предполагаемые фиксированными). Например, можно использовать конструкцию из доказательства теоремы 114, или применить код Хаффмана, или даже просто перебирать все коды, пока не найдётся подходящий.

Так или иначе построенный код можно рассматривать как некоторый способ условного префиксного описания (при известном N), для которого средняя префиксная сложность значения величины ξ^N не превосходит $H(\xi^N) + 1 = NH(\xi) + 1$. Переход к оптимальному префиксно корректному способу описания увеличит среднюю сложность не более чем на константу. $\triangleright$

$\boxed{175}$ Покажите, что можно немного усилить верхнюю оценку и показать, что среднее значение монотонной сложности $\mathit{KM}(\xi^N)$ не превосходит $NH(\xi) + O(1)$. [Указание. Примените теорему 81 к распределению вероятностей, соответствующему независимым испытаниями величины с распределением ξ.]

Мы предполагали, что величины $p_1, \ldots, p_k$ рациональны и фиксированы заранее. Если стремиться получить «равномерную» оценку, имеющую место для всех наборов $p_1, \ldots, p_n$, то следует добавить их в условие и оценивать среднее значение величины $\mathit{KP}(\xi^N|N, p_1, \ldots, p_k)$. На нижнюю оценку это вообще не повлияет, поскольку она верна для любого префиксного кода, а для получения верхней оценки (построения кода) этой информации достаточно. (Сказанное относится к случаю рациональных значений p_i; для произвольных действительных значений p_i можно взять их достаточно точные приближения.)

$\boxed{\textbf{176}}$ Сформулируйте и докажите соответствующие точные утверждения.

Доказанная теорема показывает, что *в среднем* сложность равна энтропии, хотя она бывает и больше, и меньше её. На самом деле верно и более сильное утверждение: почти что с единичной вероятностью сложность близка к энтропии, и вероятность того, что случайно выбранное значение величины ξ^N будет иметь сложность, сильно отличающуюся от энтропии, мала. Это утверждение является алгоритмическим аналогом теоремы Шеннона о пропускной способности канала без шума, и мы вернёмся к этому вопросу в разделе 7.3.4

7.3.3. Сложность начальных отрезков случайных последовательностей

[complexity-initial-segments] Покажем, как связана сложность с энтропией для начальных отрезков случайных (в смысле Мартин-Лёфа) последовательностей. Пусть по-прежнему фиксирован некоторый алфавит A из k букв и распределение вероятностей $p_1, \ldots, p_k$ на буквах этого алфавита. Будем считать, что $p_1, \ldots, p_k$ являются вычислимыми действительными числами.

Рассмотрим пространство A^∞ бесконечных последовательностей букв алфавита A и распределение вероятностей на нём, соответствующее независимым появлениям каждой буквы с распределением $p_1, \ldots, p_k$. Получается вычислимое распределение вероятностей на A^∞, и можно говорить о случайных по Мартин-Лёфу элементах A^∞. (Соответствующая теория, которая была у нас для двухбуквенного алфавита, переносится безо всяких изменений.)

Theorem 124. [complexity-limit] *Пусть ω — случайная по Мартин-Лёфу последовательность (относительно указанного распределения), $(\omega)_N$ — её начальный отрезок длины N. Тогда*
$$\lim \frac{KS((\omega)_N)}{N} = H,$$
где H — шенноновская энтропия, то есть $H = \sum p_i(-\log p_i)$.

$\boxed{\textbf{177}}$ Покажите, что для равномерного распределения это утверждение является непосредственным следствием критерия случайности (теорема 82, с. 136).

(Редкий случай, когда равномерность распределения действительно существенна.)

В формулировке теоремы мы использовали простую колмогоровскую сложность KS, но это не имеет значения, поскольку разные варианты отличаются на $O(\log N) = o(N)$. В доказательстве будет удобнее использовать монотонную сложность.

$\lhd$ По критерию случайности Левина – Шнорра (теорема 82, с. 136) сложность близка к минус логарифму вероятности появления начального отрезка $(\omega)_N$. Вероятность берётся относительно рассматриваемого распределения вероятностей на A^∞, и упомянутый минус логарифм равен $N\sum q_i(-\log p_i)$, где q_i — частота появления i-й буквы в слове $(\omega)_N$. Остаётся воспользоваться тем, что для случайных последовательностей справедлив усиленный закон больших чисел, который гарантирует, что q_i стремится к p_i при $N \to \infty$. $\rhd$

Из доказательства видно, что отклонение удельной сложности (в расчёте на букву) от энтропии складывается из трёх частей: дефект случайности (который имеет порядок $O(1)/N$), различие между простой сложностью и монотонной (порядок $O(\log N)/N$) и отклонение частот от вероятностей (оно даёт наибольший вклад; закон повторного логарифма в теории вероятностей говорит, что это отклонение чуть больше $O(\sqrt{N})/N$.

Мы предполагали, что числа p_i вычислимы, поскольку в противном случае нельзя говорить о случайности по Мартин-Лёфу. Тем не менее и в этом случае можно утверждать, что множество тех последовательностей, у которых удельная сложность начальных отрезков не стремится к шенноновской энтропии соответствующего распределения, имеет меру 0 (относительно данного распределении вероятностей).

$\boxed{\mathbf{178}}$ Докажите это. [Указание. При получении верхней оценки для сложности можно использовать не сами числа p_i, а приближения к ним (с некоторым запасом точности, скажем, с точностью до $1/N^2$ для слов длины N); при этом для задания этих приближений потребуется дополнительная информация объёма $O(\log N)$. Нижняя же оценка вообще не использует алгоритмической природы p_i (скажем, можно получить оценку для сложности с оракулом, при котором p_i вычислимы).]

7.3.4. Вероятность отклонения сложности от энтропии

`[complexity-deviation]`

Теорема 124 носит «предельный» характер; интересно, однако, получить и оценки вероятностей больших отклонений энтропии от сложности для конечных последовательностей. (Поясним это аналогией: помимо усиленного закона больших чисел, говорящего о пределах частот, в теории вероятностей есть оценки вероятностей больших уклонений на конечных отрезках.)

Пусть фиксировано некоторое распределение вероятностей $p_1, \ldots, p_k$ на буквах алфавита A. Будем снова для простоты предполагать, что числа p_i рациональны (или хотя бы вычислимы). Рассмотрим распределение вероятностей на A^N, считая буквы во всех позициях независимыми. Для каждого слова, порождаемого таким процессом, рассмотрим его сложность. Мы уже знаем (теорема 123), что среднее значение этой сложности равно энтропии NH, где $H = \sum p_i(-\log p_i)$. Но интересно знать также, как именно колеблется сложность вокруг этого среднего значения.

В самом простом (но, как мы увидим, совсем нетипичном) случае, когда имеются две равновероятные буквы, распределение вероятностей равномерно на всех словах длины N. Мы знаем, что в этом случае все слова имеют сложность не больше $N + O(1)$, а подавляющее большинство слов (за вычетом доли 2^{-c}) имеют сложность не меньше $N - c$, то есть сколько-нибудь большие отклонения сложности от энтропии весьма маловероятны.

Примерно так же обстоит дело и в случае равновероятных букв для алфавита любого размера. Однако если не все буквы равновероятны, ситуация меняется.

Ключевое наблюдение здесь состоит в следующем. Для каждого слова x, помимо фиксированных вероятностей p_i, рассмотрим также фактические частоты $q_i(x)$ появления соответствующих букв в x. Оказывается, что с большой вероятностью сложность случайно выбранного (в соответствии с рассматриваемым распределением p_i) слова близка к $k(x) = N \sum_i q_i(x)(-\log p_i)$. В самом деле, монотонная сложность по теореме 81 (с. 134) превосходит $k(x)$ не более чем на константу. С другой стороны, рассуждение при доказательстве теоремы Левина – Шнорра (с. 136, лемма 1) показывает, что вероятность события $KM(x) < k(x) - c$ (в соответствии с рассматриваемым нами распределением вероятности на словах длины N) не превосходит 2^{-c} для любого c.

Таким образом, вопрос о распределении сложностей сводится к хорошо известному в теории вероятностей вопросу о распределении частот. Известно (теорема Муавра – Лапласа), что оно близко к нормальному (гауссовскому); матемическое ожидание частоты есть вероятность, а среднеквадратическое отклонение убывает пропорционально $1/\sqrt{N}$. По сравнению с таким отклонением разницей порядка $O(\log N)$ (различные виды сложности, использование N в качестве условия и т.д.) можно пренебречь. Сказанное при некотором уточнении превращается в доказательство следующей теоремы:

Theorem 125. [square-root-deviation] *Пусть фиксирована случайная величина ξ с k значениями. Для любого положительного $\varepsilon > 0$ существует такое число c, что при всех N вероятность события $NH(\xi) - c\sqrt{N} < KS(x) < NH(\xi) + c\sqrt{N})$, если все буквы слова x независимы и распределены как ξ, не меньше $1 - \varepsilon$.*

Строго говоря, наши рассуждения использовали вычислимость чисел p_i. Но от этого предположения можно избавиться, если заменить числа p_i достаточно точными приближениями (скажем, с точностью до $1/N^2$, что требует $O(\log N)$ дополнительных битов).

7.3.5. Теоремы Шеннона о кодировании

[shannon-coding-theorem]

Доказанное нами утверждение об отклонении сложности от энтропии является естественным переводом на язык теории сложности известных результатов Шеннона о длине кода, который позволяет передавать блоки из N независимых случайных символов с данным распределением с малой вероятностью ошибки.

Пусть вновь ξ — случайная величина с k значениями (буквами алфавита A) и некоторым распределением. Пусть N — некоторое число, а ξ^N — случайная величина со значениями в A^N, у которой буквы в каждой позиции распределены как ξ и независимы. Мы хотим кодировать значения случайной величины m-битовыми словами по схеме рис. 20:

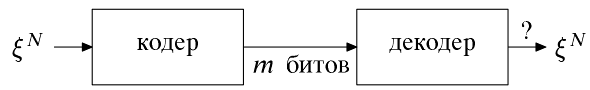

Рис. 20. Кодирование величины ξ^N с помощью m битов.

[entropy.3]

Здесь в качестве кодера рассматривается произвольное отображение множества A^N в множество $\mathbb{B}^m$ двоичных слов длины m, а в качестве декодера — произвольное отображение $\mathbb{B}^m$ в A^N. Говорят, что происходит ошибка, если выходное слово в алфавите A не равно входному (значению величины ξ^N), и вероятность ошибки измеряется относительно описанного распределения случайной величины ξ^N. Нас интересует, при каких m и ε существует код длины m с вероятностью ошибки не более ε. Начнём со следующего очевидного замечания:

Theorem 126. *Такой код существует тогда и только тогда, когда 2^m наиболее вероятных значений случайной величины ξ^N имеют суммарную вероятность не меньше $1-\varepsilon$.*

$\triangleleft$ В самом деле, при кодировании m битами можно передать без ошибок 2^m значений, но не больше. Ясно, что вероятность ошибки будет минимальной, если в качестве этих 2^m значений выбрать 2^m наиболее вероятных значений. $\triangleright$

В следующей теореме алфавит A и величина ξ фиксированы.

Theorem 127. *Для всякого $\varepsilon > 0$ существует такое число c, что:*

(а) *для ξ^N существует код длины не больше $NH(\xi) + c\sqrt{N}$ с вероятностью ошибки не больше ε;*

(б) *любой код длины не больше $NH(\xi) - c\sqrt{N}$ для ξ^N имеет вероятность ошибки не меньше $1-\varepsilon$ (вероятность правильной передачи не больше ε).*

$\triangleleft$ (а) Мы знаем, что при подходящем c значение случайной величины ξ^N имеет сложность менее $m = NH(\xi) + c\sqrt{N}$ с вероятностью $1-\varepsilon$. Поэтому для таких значений можно использовать кратчайшие описания (в смысле колмогоровской сложности) в качестве кодов. (Заметим, что в теореме ничего не говорится про алгоритм кодирования, нам важно лишь существование кода. Формально говоря, в качестве закодированных сообщений у нас выступают слова длины меньше m, но их менее 2^m штук и можно их заменить на слова длины m.)

(б) Здесь нам понадобится небольшая хитрость. Если существует код указанной длины с указанной вероятностью ошибки, то можно построить такой код эффективно (воспользовавшись предыдущей теоремой, или просто полным перебором) и рассматривать декодирующую функцию этого кода как способ условного описания (при известных параметрах p_i и N). Поэтому все значения величины ξ^N, которые раскодируются правильно, имеют (условную — относительно указанных параметров) сложность не более $NH(\xi) - c\sqrt{N} + O(\log N)$ (последний член соответствует сложности параметров и поглощается при увеличении константы c). По предыдущей теореме (при подходящем c) вероятность этого не превосходит ε. $\triangleright$

$\boxed{179}$ Как и раньше, в этом рассуждении предполагается, что мы знаем числа p_i точно, и если они невычислимы, возникают дополнительные трудности. Покажите, как можно исправить это рассуждение, заменив p_i на их достаточно точные приближения.

$\boxed{180}$ Сформулируйте и докажите аналогичным способом теорему об условном кодировании и условной энтропии. [Указание. Пусть имеются две зависимые случайные величины ξ и η, проводится N независимых испытаний, значение величины η^N известно и отправителю, и получателю, и отправитель хочет передать m битов так, чтобы получатель смог восстановить значение величины ξ^N. При каких m это возможно, а при каких нет?]

7.4. Марковские цепи

с точки зрения колмогоровской сложности и энтропии (что здесь известно? Андрей?) сжатие lempel-ziv? как найти энтропии марковской цепи?

8. Некоторые приложения

[appl]

8.1. Бесконечность множества простых чисел

[appl-primes]

Начнём с совсем «игрушечного» применения: докажем, что множество простых чисел бесконечно.

Пусть это не так и существует всего m различных простых чисел $p_1, \ldots, p_m$. Тогда любое целое число x разлагаясь на простые множители, представляется в виде

$$x = p_1^{k_1} p_2^{k_2} \ldots p_m^{k_m}$$

и тем самым может быть задано набором степеней $k_1, \ldots, k_m$. Каждое из чисел k_i не превосходит $\log x$ (основания степеней не меньше 2) и потому имеет сложность не более $O(\log\log x)$ (двоичная запись содержит $O(\log\log x)$ битов). Поскольку m фиксировано, то сложность набора $\langle k_1, k_2, \ldots, k_m\rangle$ есть $O(\log\log x)$ и потому сложность любого числа x (которое алгоритмически получается из этого набора) есть $O(\log\log x)$. А для «случайного» n-битового числа x сложность примерно равна n, а не $O(\log n)$, как получается по этой формуле (логарифм n-битового числа не превосходит n).

Можно ли считать это «честным» применением колмогоровской сложности? Скептик скажет, что здесь всего лишь производится обычный подсчёт числа возможностей (counting argument, как говорят): если существует лишь m различных простых чисел, то различных чисел от 1 до x существует не более $(\log x)^m$, поскольку каждое такое число задаётся показателями степеней при простых множителях, и каждый показатель меньше $\log x$. После чего мы приходим к противоречию, поскольку $x > (\log x)^m$ при больших x.

Возразить на это нечего — именно это рассуждение и пересказано с помощью колмогоровской сложности (и стало немного более громоздким за счёт различных асимптотических обозначений). Тем не менее в подобном переводе может быть некоторый смысл, поскольку новый язык формирует новую интуицию, и она может быть полезна, даже если потом можно всё пересказать на старом языке.

К этому обсуждению мы ещё вернёмся, разобрав другие примеры.

8.2. Перенос информации по ленте

[appl-tape]

Следующий пример тоже модельный — мы докажем, что копирование слова длины n на одноленточной машине Тьюринга требует (в худшем случае) не менее εn^2 шагов. Этот классический результат был получен в 1960-ые годы с помощью так называемого «метода следов»; наше доказательство является переводом классического на язык колмогоровской сложности. (Мы предполагаем известными основные понятия, связанные с машинами Тьюринга; см., например, [77]).

Пусть на ленте машины Тьюринга выделена «буферная зона» некоторого размера b; нас интересует скорость переноса информации через эту буферную зону, скажем, слева направо, из области L в область R (рис. 21)

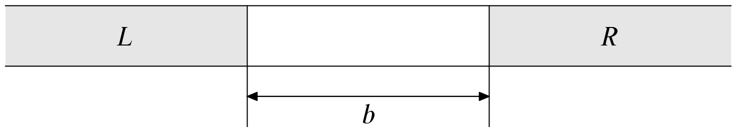

Рис. 21. Буферная зона размера b.

[tape-buffer]

В некоторый момент ленты буферная область и R пусты, а область L содержит произвольную информацию. Нас интересует, какова может быть сложность слова R через t шагов после этого. Мы докажем, что она не больше $(t\log m)/b+O(\log t)$, где m — число состояний машины Тьюринга, а b — ширина буферной зоны. В самом деле, каждое из m состояний машины несёт $\log m$ битов информации, и за шаг информация переносится на одну клетку, а перенос её на b клеток требует в b раз больше времени.

Осталось лишь уточнить формулировку и доказательство.

Theorem 128. *Пусть фиксирована машина Тьюринга с m состояниями. Тогда существует такая константа c, что для любого b и для любого вычисления с буферной зоной b (вначале эта зона и лента справа от неё пусты, головка машины слева от зоны) сложность правой части ленты $R(t)$ после t шагов вычисления не превосходит*

$$\frac{t\log m}{b}+4\log t+c.$$

$\triangleleft$ Проведём границу где-нибудь посередине буферной зоны, и при каждом пересечении границы головкой машины Тьюринга слева направо (при «выезде за границу») будем записывать, в каком состоянии она её пересекла (мечта ЧК-ГБ!) Записанная последовательность состояний называется *следом* машины. Зная след, можно восстановить работу машины справа от границы (без использования информации о ленте слева от границы). В самом деле, надо искусственно поместить машину в первое состояние, указанное в следе, и выпустить её за границу; по возвращении перевести её во второе состояние следа и снова выпустить и так далее. При этом за границей машина будет вести себя так же, как и раньше (ведь ничего, кроме состояния, при пересечении границы у ней нет). В частности, в некоторый момент t' состояние ленты будет содержать $R(t)$. При этом t' не превосходит t, хотя может быть и меньше (поскольку время, проведённое машиной слева от границы, теперь не учитывается). Таким образом, можно алгоритмически восстановить $R(t)$, зная след, t', а также расстояние b' от границы до начала зоны R. Поэтому найдётся такая константа c (зависящая от машины, но не от b и t), что для любого следа S

$$KS(R(t))\leqslant l(S)\log m+4\log t+c$$

(мы умножили длину $l(S)$ следа S на $\log m$, поскольку след является словом в m-буквенном алфавите и на каждую букву приходится $\log m$ битов; приписывание к нему чисел b' и t' (в самоограниченной записи) требует не более $2\log b+2\log t$ битов, причём можно считать, что $t>b$ (иначе R пуста, головка туда не дошла); константа c возникает при переходе к оптимальному способу описания).

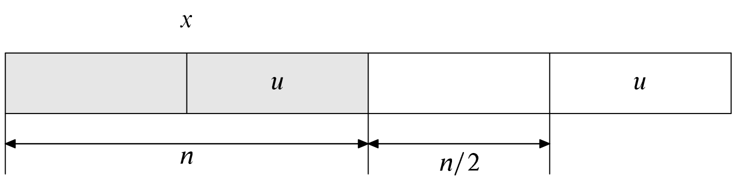

Рис. 22. Буферная зона при копировании.
[tape-copy-buffer]

Это неравенство верно для любого начального содержимого части L и для любого положения границы: если теперь для данного L взять самый короткий из следов, то его длина меньше t/b (всего разных положений границы $b+1$, и в каждый момент пересечение происходит лишь в одном месте, так что сумма длин следов не больше t). Получаем требуемое неравенство. ▷

$\boxed{181}$ Покажите, что оценку теоремы можно улучшить, заменив b в знаменателе на $2b$. [Указание. Въезд суммарно потребует почти столько же шагов, сколько и выезд (не более чем на b меньше)]

Теперь сразу же получается квадратичная оценка для задачи копирования.

Пусть одноленточная машина M копирует любое входное слово: если вначале на ленте написано слово x из нулей и единиц, а дальше лента пуста, то в конце работы справа от x появляется его копия, и на ленте написано xx.

Theorem 129. *Существует такая константа $\varepsilon > 0$, что для любого n существует слово длины n, копирование которого с помощью M занимает не менее εn^2 шагов.*

◁ Будем для простоты считать, что n чётно, и возьмём в качестве x слово, у которого вторая половина u имеет сложность не меньше длины (то есть $n/2$). Применим теорему о скорости переноса информации, считая буферной зоной участок длины $n/2$ справа от x (рис. 22).

Пусть копирование заняло t шагов, тогда сложность зоны R после этого не меньше $n/2$; с другой стороны, она не превосходит $t\log m/b + 4\log t + c$ по доказанной теореме, где b — ширина буферной зоны, то есть $n/2$. Получаем, что

$$\frac{n}{2} \leqslant \frac{t\log m}{n/2} + 4\log t + c.$$

Без ограничения общности считаем, что $t < n^2$ (иначе всё и так хорошо) и заменяем $4\log t$ на $8\log n$. Получаем, что

$$t \geqslant \frac{n^2}{4\log m} - O(n\log n).$$

Второй член мал по сравнению с первым при больших n (и формально можно распространить неравенство на все n за счёт уменьшения коэффициента при n^2). ▷

Насколько существенна в этом доказательстве колмогоровская сложность? Скептик может вновь сказать, что по существу мы просто оценивали число слов, которые можно скопировать за данное время, используя тот факт, что разные слова должны давать разные следы (иначе справа от границы поведение машины было бы одно и то же). Действительно, именно так и выглядело первоначальное доказательство этой оценки (уточним, что в первоначальных работах рассматривалось не копирование, а распознавание симметрии, но вся техника та же самая). Насколько идея этого доказательства становится понятнее при переводе его на язык колмогоровской сложности — вопрос вкуса.

Многие оценки в теории сложности вычислений могут быть доказаны аналогичным образом, когда в качестве «трудного случая» рассматривается слово максимальной сложности в некотором классе и затем показывается, что если бы оценка не соблюдалась, то это слово имело бы меньшую сложность. Много таких приложений (со ссылками на оригинальные работы) приведено в книге Ли и Витаньи [38], которые сами сыграли большую роль в развитии этого метода (называемого *Incompressibility method*). Отметим, что во многих случаях оценка впервые была доказана как раз с использованием колмогоровской сложности.

В следующем разделе мы приведём ещё один пример использования этого метода, а затем перейдём к некоторым другим приложениям. Они интересны не сами по себе, а как примеры разнообразных способов, с помощью которых можно применять теорию колмогоровской сложности для доказательства утверждений, с ней не связанных.

8.3. Конечные автоматы с несколькими головками

Конечный автомат с k головками отличается от обычного (с которым, как мы предполагаем в этом разделе, читатель знаком) тем, что вместо одной читающей головки он имеет k односторонних (движущихся слева направо) читающих головок. В начальный момент k читающих головок головки видят первый (крайний слева) символ входного слова. На каждом шаге, в зависимости от состояния автомата и от k символов, читаемых k головками, автомат меняет своё состояние и продвигает некоторые из k головок (не менее одной) вправо, после чего процесс повторяется. Вслед за словом на ленте записан специальный символ конца слова; работа автомата заканчивается, когда все головки попадают на этот символ. Автомат допускает слово, если по окончании работы он находится в заключительном состоянии. Множество всех слов, допускаемых автоматом, называется языком, распознаваемым этим автоматом.

Пример. Рассмотрим множество всех слов вида $x\#x$, где x — всевозможные двоичные слова. Этот язык не распознаётся конечным автоматом с одной головкой (классический результат теории конечных автоматов), но распознаётся автоматом с двумя головками. В самом деле, одну головку надо отправить на поиски символа $\#$, после чего синхронно сдвигать обе головки и проверять, что под ними проходят одинаковые символы.

Тем самым две головки лучше, чем одна (можно распознавать больше языков). Оказывается, что вообще $k+1$ головок лучше, чем k:

Theorem 130. *Для всякого k существует язык, который распознаётся некоторым автоматом с $k+1$ головками, но не распознаётся никаким автоматом с k головками.*

$\triangleleft$ Для каждого $m \geqslant 1$ рассмотрим язык L_m, состоящий из всех слов

$$w_1\#\dots w_m\#w_m\#\dots w_1$$

(для всех двоичных слов $w_1, \ldots, w_m$). Каждое из слов w_i повторяется дважды, причём в правой половине слова идут в обратном порядке (это очень существенно).

Автомат с k головками может распознавать этот язык так: одна из головок отправляется в правую половину слова, а остальные $k-1$ головок встают на начала слов $w_1, \ldots, w_{k-1}$. Каждая из этих $k-1$ головок сверяет порученное ей слово с его копией, когда первая головка проходит мимо этой копии. После этого первые $k-1$ слов проверены, первая головка полностью использована (она дошла до конца слова и больше ни на что не пригодна), а остальные головки стоят левее ещё не проверенных слов $w_k, w_{k+1}, \ldots$. Процесс повторяется: ещё одна головка отправляется через всё слово, а $k-2$ головок сверяют каждая своё слово и так далее. Всего таким образом можно сверить

$$(k-1)+(k-2)+\ldots+1=\frac{k(k-1)}{2}=C_k^2$$

слов. (Заметим, что число m у нас фиксировано, поэтому поиск слова с заданным номером требует ограниченного числа состояний автомата.)

Таким образом, *язык L_m распознаётся k-головочным автоматом при $m \leqslant C_k^2$.*

Осталось показать, что *при $m > C_k^2$ язык L_m не распознаётся никаким k-головочным автоматом.* Пусть это не так и автомат M его распознаёт. Чтобы придти к противоречию, рассмотрим независимые случайные слова $w_1, \ldots, w_m$ достаточно большой длины N. Более формально, рассмотрим слово длины mN и сложности не меньше mN, и разрежем его на m слов длины N, которые обозначим $w_1, \ldots, w_m$. По предположению слово

$$W = w_1\# \ldots w_m\# w_m\# \ldots w_1$$

допускается автоматом M; мы придём к противоречию, показав, что либо слово $w_1 \ldots w_m$ имеет меньшую сложность, чем мы предположили, либо автомат M работает неверно (не распознаёт L_m).

Будем говорить, что пара головок автомата M *проверила* слово w_i, если при работе автомата M на данном слове был момент, когда эти головки находились внутри левой и правой копии слова w_i. Ключевое наблюдение: *никакая пара головок не может проверить два разных слова.* В самом деле, если она проверила слово w_i, то до этого левая головка читала слова с меньшими номерами, а правая — с большими, а после этого наоборот.

По предположению $m > C_k^2$; поэтому существует некоторое слово w_i, не проверенное ни одной парой головок. Покажем, что либо это слово проще, чем нужно, либо одну из его копий можно изменить незаметно для автомата M (тем самым дискредитировав этот автомат).

Будем наблюдать за работой автомата M на слове W, обращая особое внимание на те моменты, когда одна из головок автомата входит в слово w_i (в любую из копий) или выходит из неё, и записывать положения всех головок и состояние автомата в эти моменты. Полученный протокол P имеет сложность $O(\log N)$ (константа зависит от k, m и от числа состояний автомата, но не от N), поскольку моментов входа и выхода не более $4k$ (по четыре для каждой головки), и в каждый момент надо записать состояние автомата и положение всех головок, что требует $O(\log N)$ битов.

Покажем, что (если автомат M распознаёт L_m) слово w_i восстанавливается однозначно по всем остальным словам w_j с $j \neq i$ и протоколу P. В этом случае сложность слова

$w_1 \dots w_m$ не превосходит $(m-1)N$ (число битов в остальных словах) плюс $O(\log N)$ (число битов в протоколе) плюс $O(1)$, что меньше mN при больших N, и полученное противоречие завершает доказательство.

Восстанавливать будем так: подставляя по очереди всевозможные слова w данной длины на место w_i (при тех же w_j с $j \ne i$), мы запускаем автомат M и смотрим, не получился ли протокол P. При этом возможны разные случаи.

(1) Если хотя при при одном w автомат M не допустит соответствующего слова, он заведомо неправильный.

(2) Автомат M допускает все такие слова, при этом протокол P встречается только однажды, при $w = w_i$. (В этом случае w_i восстанавливается по P и остальным словам.)

(3) Автомат M допускает все такие слова, при этом протокол P встречается и при некотором $w \ne w_i$. Покажем, что в этом случае автомат M допускает неправильное слово (не из L_m), а именно слово W', в котором в левой половине записано w_i, а в правой на месте w_i стоит слово w.

В самом деле, у нас есть два допускающих поведения автомата M: когда с обеих сторон w_i и когда с обеих сторон w. Разрежем каждое из них на отрезки по тем моментам, когда одна из головок входит или выходит из w_i (или w). Положения остальных головок в эти моменты, а также состояния автомата, зафиксированы в протоколе P, и потому одни и те же в обоих случаях. (Сами моменты времени могут быть различны, поскольку их мы — хотя могли бы — не включили в протокол.) Поэтому можно склеивать отрезки поведения автомата для двух разных случаев; покажем, что это можно сделать так, чтобы в левой половине было w_i, а в правой w.

По предположению в ходе обработки слова W никакая пара головок одновременно не попадает в левую и правую копии слова w_i; поскольку прохождение границы фиксируется в протоколе, это же верно и после замены w_i на w. Поэтому для каждого отрезка протокола есть три возможности: (а) головка в левой копии; (б) головка в правой копии; (в) ни то, ни другое. Если теперь для случая (а) брать кусок поведения автомата M на слове W, для случая (б) брать кусок поведения на изменённом слове (когда w_i заменено на w), а для случая (в) брать любой из двух вариантов (они совпадают), то получится работа автомата M на смешанном слове W', и потому автомат допускает W' в противоречии с предположением. $\triangleright$

8.4. Усиленный закон больших чисел

[appl-lln]

Усиленный закон больших чисел был доказан нами в разделе 3.2 (теорема 27, с. 57) без использования колмогоровской сложности (прямым подсчётом). Подробно был разобран случай равномерной меры: в этом случае закон утверждает, что множество тех последовательностей $\omega = \omega_0\omega_1\dots$, у которых предел частот

$$p_n = \frac{\omega_0 + \omega_1 + \ldots + \omega_{n-1}}{n}$$

при $n \to \infty$ существует и равен $1/2$, имеет меру 1 относительно равномерной меры на пространстве Ω. Другими словами, дополнение этого множества (те последовательности, где предел не существует или не равен $1/2$), является нулевым. Впоследствии, в теореме 32 (с. 67), мы установили, что это нулевое множество является и эффективно нулевым, откуда заключили, что для всех случайных по Мартин-Лёфу последовательностей предел частот равен $1/2$ (теорема 33, с. 67).

Однако можно действовать и в другом направлении. А именно, можно сначала доказать, что для всех случайных по Мартин-Лёфу последовательностей предел частот равен $1/2$, используя критерий случайности в терминах сложности (теорема 82, с. 136). Этот критерий говорит, что для случайной по равномерной мере последовательности ω монотонная сложность её начального отрезка $(\omega)_n$ длины n равна $n + O(1)$. Из этого можно заключить, что частота единиц p_n единиц в $(\omega)_n$ стремится к $1/2$. В самом деле, по теореме 122, сложность (обычная или монотонная, разница между ними логарифмическая) слова $(\omega)_n$ не превосходит $nh(p_n, 1 - p_n) + O(\log n)$, и потому величина $h(p_n, 1 - p_n)$ стремится к 1 при $n \to \infty$ (разница между ними не превосходит $O(\log n)/n$). Отсюда и из графика рис. 8 (с. 58) можно заключить, что $p_n \to 1/2$ при $n \to \infty$ для любой случайной последовательности, а такие последовательности образуют множество меры 1.

Возникает вопрос, является ли это рассуждение действительно новым или просто мы повторили по существу те же аргументы на несколько другом языке? Скорее второе: если вспомнить, как доказывалась теорема 122, то видно, что там использовалась та же оценка с формулой Стирлинга, что и в прямом доказательстве закона больших чисел. (Другое рассуждение, с оценкой через логарифм меры по теореме 81, также имеет прямой перевод в терминах оценок вероятностей, который приводился нами в разделе 3.2, после доказательства теоремы 27.)

Что же мы выигрываем от перехода к колмогоровской сложности? Можно указать несколько преимуществ. Во-первых, можно расширить класс последовательностей, для которых справедлив усиленный закон больших чисел:

Theorem 131. [lln-complexity-bound] *Пусть последовательность ω такова, что $KS((\omega)_n) = n + o(n)$. Тогда последовательность частот единиц в её начальных отрезках стремится к $1/2$.*

◁ Доказательство остаётся по существу без изменений: в этом случае $h(p_n, 1 - p_n)$ по прежнему есть $1 + o(1)$. ▷

Во-вторых, мы можем не ограничиваться утверждением $p_n \to 1/2$, а оценивать скорость этой сходимости. Соответствующее утверждение в теории вероятностей называется *законом повторного логарифма*, и колмогоровская сложность позволяет дать простое доказательство верхней оценки в этом законе.

Theorem 132. [iterated-logarithm-upper] *Пусть ω — случайная по Мартин-Лёфу последовательность относительно равномерной меры, p_n — частота единиц в её начальном отрезке длины n. Тогда, каково бы ни было $\varepsilon > 0$, неравенство*

$$|p_n - 1/2| \leqslant (1 + \varepsilon)\sqrt{\frac{\ln \ln n}{2n}}$$

выполнено для всех достаточно бóльших n.

◁ Посмотрим, какую оценку даёт нам (описанное выше) использование колмогоровской сложности. Мы знаем, что

$$n - O(1) \leqslant KM((\omega)_n) \leqslant nh(p_n, 1-p_n) + O(\log n),$$

откуда

$$h(p_n, 1-p_n) \geqslant 1 - O(\log n/n).$$

Функция

$$p \mapsto h(p, 1-p) = p(-\log p) + (1-p)(-\log(1-p))$$

имеет максимум в точке $p = 1/2$, при этом вторая производная в точке максимума отлична от нуля (и равна $-4/\ln 2$). Поэтому по формуле Тейлора получаем, что

$$h(1/2+\delta) = 1 - \frac{2}{\ln 2}\delta^2 + o(\delta^2)$$

при $\delta \to 0$. Следовательно, для $\delta_n = p_n - 1/2$ имеем

$$\delta_n^2 = O(\log n/n),$$

то есть

$$|p_n - 1/2| = O\left(\sqrt{\frac{\log n}{n}}\right).$$

Это уже кое-что, хотя до нужной нам оценки ещё далеко. (Заметим в скобках, что и в теории вероятностей оценка с повторным логарифмом была получена не сразу: сначала Хаусдорф (1913) доказал оценку $O(n^\varepsilon/\sqrt{n})$, затем Харди и Литлвуд (1914) её улучшили, заменив n^ε на $\sqrt{\log n}$, затем Штейнгауз (1922) доказал оценку $(1+\varepsilon)\sqrt{(2\ln n)/n}$, и лишь потом Хинчин (1924) получил более сильную оценку доказываемой нами теоремы 132, заменив логарифм на повторный логарифм. Так что пока что мы находимся на уровне Харди и Литлвуда.)

Посмотрим, как можно улучшить верхнюю оценку для $\mathit{KM}((\omega)_n)$. (Нижнюю оценку улучшать особенно некуда.) Она получалась сравнением $\mathit{KM}((\omega)_n)$ с логарифмом вероятности последовательности $(\omega)_n$ относительно бернуллиевой меры с параметром p_n. Этот логарифм был в точности равен $nh(p_n, 1-p_n)$, но сама бернуллиева мера зависела от n, и потому конструкция из доказательства теоремы 81 давала оценку с дополнительным членом, равным $\mathit{KP}(p_n)$ (мы должны начать с самоограниченного кода для p_n). При этом $\mathit{KP}(p_n)$ можно оценить как $(2+\varepsilon)\log n$, поскольку числитель и знаменатель дроби p_n не превосходят n. Всё это вместе даёт оценку

$$\frac{2}{\ln 2}(p_n - 1/2)^2 \approx 1 - h(p_n, 1-p_n) \leqslant (2+\varepsilon)\log n/n,$$

что по-прежнему недостаточно.

Что же ещё можно сэкономить? Заметим, что мы уже знаем, что p_n достаточно близко к $1/2$: если знаменателем считать n, то числитель отклоняется от $n/2$ на величину чуть больше $\sqrt{n}$. Поэтому (при известном знаменателе) количество битов для описания числителя можно сократить примерно вдвое, и тогда вместо 2 в правой части будет $1{,}5$. Но мы хотим большего.

Решающая идея состоит в следующем. Пусть для определённости $p_n = 1/2 + \delta_n > 1/2$. Будем вместо p_n использовать для оценки монотонной сложности величину $1/2 + \delta'_n$, где δ'_n — приближение к δ_n снизу с небольшой (фиксированной) относительной погрешностью. Пусть, например, $0{,}9\delta_n < \delta'_n \leqslant \delta_n$. Такое число δ'_n можно найти среди членов геометрической прогрессии $(0{,}9)^k$, и его сложность равна примерно $\log k$, то есть $\log(-\log \delta_n / \log 0{,}9) = \log(-\log \delta_n) + c$. При этом, если $\delta_n < 1/\sqrt{n}$, то доказывать нечего, поэтому сложность δ_n можно оценить как $(1+\varepsilon)\log\log n$ (при сколь угодно малом ε и достаточно больших n).

Это хорошо, но зато ухудшилась верхняя оценка на сложность $(\omega)_n$. А именно, вместо $h(p_n, 1-p_n)$ получается более сложное выражение

$$p_n(-\log p'_n) + (1-p_n)(-\log(1-p'_n)),$$

где $p'_n = 1/2 + \delta'_n$; можно сказать, что фактические частоты нулей и единиц в $(\omega)_n$ равны p_n и $1-p_n$, но при «кодировании» используются не они, а приближённые (и упрощённые) p'_n и $1-p'_n$. Ясно, что это выражение лишь увеличится, если заменить в нём p_n на p'_n (поскольку $p_n > p'_n > 1/2$, вторая скобка с логарифмом больше первой, и увеличение её веса за счёт первой увеличивает всю сумму).

В итоге получаем оценку

$$n - O(1) \leqslant nh(p'_n, 1-p'_n) + (1+\varepsilon)\log\log n$$

для любого $\varepsilon > 0$ (при достаточно бо́льших n) и отсюда, как и раньше, получаем

$$\delta'_n \leqslant (1+\varepsilon)\sqrt{\ln 2 \cdot \log\log n / 2n}.$$

Оценка для «истинного» δ_n лишь ненамного больше (в $1/0{,}9$ раза в нашем примере); поскольку вместо $0{,}9$ можно было взять сколь угодно близкое к единице число, получаем искомое утверждение (множитель $\ln 2$ превращает первый логарифм в натуральный, а второй можно заменить на натуральный, поскольку это поглощается множителем $(1+\varepsilon)$). ▷

182 Приведённое доказательство теоремы 132 не полностью использует случайность последовательности ω: покажите, что оно проходит для последовательностей, у которых $n - \mathit{KM}((\omega)_n) = o(\log\log n)$.

8.5. Последовательности без запрещённых подслов

[appl-1111]

8.5.1. Запрещённые и простые слова

Доказываемое в этом разделе утверждение интересно не само по себе, а как неожиданное применение колмогоровской сложности.

Theorem 133. [no-forbidden-strings] *Пусть $\alpha < 1$ — некоторое положительное действительное число и для каждого n некоторые двоичные слова, общим числом не более $2^{\alpha n}$, объявлены* запрещёнными. *Тогда существуют число c и бесконечная последовательность нулей и единиц, не содержащая запрещённых подслов длиннее c.*

Например, можно считать запрещёнными слова, у которых (простая) колмогоровская сложность меньше αn, их как раз не больше $2^{\alpha n}$. Получаем такое следствие:

Theorem 134. `[no-simple-strings]` *Пусть $\alpha < 1$ — некоторое положительное действительное число. Тогда существует бесконечная последовательность нулей и единиц, в которой все подслова достаточно большой длины n имеют сложность не меньше αn.*

Поучительно сравнить это утверждение с критерием случайности для равномерной меры (теорема 86, с. 140). Там речь шла лишь о начальных отрезках, а не всех подсловах, зато монотонная сложность была не меньше $n - O(1)$ (для обычной сложности это даёт оценку $n - O(\log n)$). Как показывает следующая задача, для всех подслов (а не только для начал) такую более сильную оценку получить нельзя. (Это и не удивительно: в настоящей случайной последовательности должны появляться все подслова, в том числе и малой сложности.)

183 Для всякой последовательности ω нулей и единиц найдётся $\alpha < 1$ и бесконечно много различных подслов, у которых отношение сложности к длине не больше α. [Указание. Рассмотрим два случая. Если последовательность содержит в качестве подслов все двоичные слова, то доказывать нечего. Если же в неё не входит некоторое слово u длины k, то разбивая длинные слова на блоки длины k и кодируя их оптимальным образом (блок u никогда не встречается и код ему не нужен), мы получаем отношение сложности к длине не больше $(\log(2^k - 1))/k$.]

Доказательство теоремы 133 происходит в два этапа. Сначала мы докажем её частный случай, теорему 134, а затем удивительным образом окажется, что из него следует и общий случай.

$\triangleleft$ Для доказательства теоремы 134 рассмотрим произвольное β, для которого $\alpha < \beta < 1$. Пользуясь теоремой 65 (с. 110), найдём число N с таким свойством: ко всякому слову x можно приписать N битов так, чтобы префиксная сложность возросла от этого не менее чем на βN. Будем использовать это свойство многократно, начав с пустого слова: получится бесконечная последовательность блоков длины N, от добавления каждого блока префиксная сложность возрастает как минимум на βN. Отсюда следует, что сложность любой группы из подряд идущих k блоков не меньше $\beta k N - O(1)$. В самом деле, от добавления этой группы сложность растёт не менее чем на $\beta k N$, а $\mathit{KP}(xy) \leqslant \mathit{KP}(x) + \mathit{KP}(y) + O(1)$. Отсюда следует, что и для любого подслова u (не обязательно начинающегося и кончающегося на границе блока) сложность не меньше $\beta l(u) - O(1)$, поскольку изменение сложности и длины из-за граничных эффектов (обрезание по границе блока) есть $O(1)$. Остаётся вспомнить, что β мы взяли по сравнению с α некоторым запасом, который покрывает и граничные эффекты, и разницу между обычной и префиксной сложностями. $\triangleright$

184 Проведите аналогичное рассуждение, используя обычную сложность вместо префиксной. [Указание. См. задачу 34 на с. 45.]

$\triangleleft$ Теперь докажем теорему 133; проще всего это сделать, если воспользоваться релятивизованным вариантом сложности. Будем считать, что множество F запрещённых слов дано нам в качестве оракула, то есть всем рассматриваемым алгоритмам разрешается (в качестве внешней процедуры) узнавать про любое слово, запрещено оно или нет. Как обычно в теории алгоритмов, такая релятивизация ничему не мешает, и мы будем считать, что теорема 134 верна и для релятивизованной относительно F сложности.

Заметим, что при таком оракуле все запрещённые слова длины n имеют сложность (при известном n) не более $\alpha n+O(1)$, поскольку каждое слово может быть задано своим порядковым номером в списке запрещённых слов. Поэтому безусловная сложность запрещённого слова длины n не больше $\alpha n+O(\log n)$, и достаточно взять последовательность, у которой нет длинных подслов с удельной сложностью меньше β для некоторого $\beta > \alpha$. $\triangleright$

Любопытно, что при желании можно вывести теорему 133 из теоремы 134, не используя релятивизации. А именно, справедливо следующее утверждение:

Theorem 135. *Если для некоторого рационального α и некоторого множества F запрещённых слов утверждение теоремы 133 не выполняется, то оно не выполняется и для некоторого разрешимого множества F.*

Ясно, что для разрешимого множества F релятивизация ничего не меняет и не нужна; ограничение рациональными α тоже, очевидно, несущественно.

$\triangleleft$ Пусть для некоторого $\alpha < 1$ и некоторого множества F утверждение теоремы 133 неверно. Тогда для всякого c можно указать множество F_c с такими свойствами:

(а) F_c содержит только слова длиннее c;

(б) F_c содержит не более $2^{\alpha k}$ слов длины k (при любом k);

(в) всякая бесконечная последовательность содержит хотя бы одно подслово из F_c.

(В самом деле, можно в качестве F_c взять множество всех слов из F, имеющих длину больше c.)

Обычное рассуждение (лемма Кёнига, компактность) показывает, что и всякая достаточно длинная последовательность тогда содержит хотя бы одно слово из F_c, и потому множество F_c без ограничения общности можно считать конечным. Более того, его (точнее, какое-либо множество с указанными свойствами) можно найти перебором, так что оно вычислимо строится по c. (Переход к конечному множеству нужен, чтобы воспользоваться перебором.)

Построим теперь последовательность c_i, беря c_{i+1} больше длин всех слов в множестве F_{c_i}. Объединение соответствующих множеств F_{c_i} и будет разрешимым множеством, для которого не выполнено утверждение теоремы 133. $\triangleright$

Обратите внимание на схему рассуждения: мы предполагали, что объект с некоторым свойством существует, а затем замечали, что тогда можно перебором построить (возможно, другой) объект с таким свойством. Эта схема в той или иной форме ещё не раз нам встретится.

8.5.2. Лемма Ловаса

Мы показали, как от сложностного утверждения перейти к комбинаторному аналогу. Удивительным образом оказывается, что можно двигаться и в обратную сторону, начав с комбинаторного утверждения (локальной леммы Ловаса) и выведя из него следствия, касающиеся колмогоровской сложности. Но вначале несколько общих слов.

Вероятностные доказательства существования. Часто существование объекта, удовлетворяющего некоторому набору условий, доказывают так: на множестве всех объектов вводят распределение вероятностей, и для каждого условия подсчитывают вероятность того, что оно будет нарушено. Если эти вероятности малы — настолько малы, что их сумма меньше 1, — то случайный объект удовлетворяет всем условиям с положительной вероятностью и потому искомый объект существует.

Свойство вероятностей, используемое в этом рассуждении: если вероятность события A_i не больше ε_i, то вероятность объединения событий $A_1, \ldots, A_n$ не больше $\sum \varepsilon_i$, а вероятность того, что не произойдёт ни одного из них, не меньше

$$1 - \varepsilon_1 - \varepsilon_2 - \ldots - \varepsilon_n.$$

Вычисление вероятностей часто сводится к подсчёту количества «плохих» элементов в некотором множестве: если общее число плохих элементов меньше числа всех элементов, то есть и хорошие элементы. Как мы уже говорили, это можно перевести и на сложностной язык: если плохих элементов мало, то они имеют малую сложность, а потому несжимаемый элемент (элемент максимальной сложности) обязательно будет хорошим.

Если попытаться доказать теорему 133 с помощью такого рода подсчётов, то ничего не получится: вероятность обнаружить плохую строку в данной позиции действительно мала ($2^{(\alpha-1)n}$), но если позиций достаточно много, то сумма этих вероятностей уже станет больше единицы — а нам надо доказать существование последовательностей любой длины без плохих подслов. Спасительное наблюдение: если два участка случайной последовательности не пересекаются, то появления в них запрещённых подслов — независимые события.

Случай независимых событий. Для начала рассмотрим случай полностью независимых событий. Если события A_i независимы, и вероятность A_i равна ε_i, то вероятность того, что ни одно из A_i не произойдёт, равна

$$(1-\varepsilon_1) \cdot (1-\varepsilon_2) \cdot \ldots \cdot (1-\varepsilon_n).$$

Отсюда, кстати, можно заключить, что это произведение больше $1 - \varepsilon_1 - \varepsilon_2 - \ldots$ (вариант неравенства Бернулли).

Видно, что для независимых событий эта вероятность будет положительной, даже если сумма всех ε_i больше единицы; достаточно, чтобы каждое из ε_i было меньше единицы.

Лемма Ловаса относится к промежуточной ситуации, когда событий слишком много (и потому не годится первая оценка), и они не все независимы (и потому не годится вторая оценка).

Пусть имеется n событий $A_1, \ldots, A_n$ и для каждого i от 1 до n фиксировано некоторое множество $N(i) \subset \{1, \ldots, n\}$, не содержащее i. Его элементы будем называть *соседями* i. (Никаких условий симметрии не предполагается: вершина может не быть соседом своего соседа.)

Предположим, что событие A_i независимо со всеми событиями, кроме себя и своих соседей (мы допускаем вольность речи и отождествляем событие A_i с индексом i). (Имеется в виду именно независимость от набора всех несоседних событий, а не от каждого из них в отдельности). Тогда справедлива такая оценка:

Theorem 136. `[lovasz-lemma]` *Если для каждого i выбрано положительное число $\varepsilon_i < 1$ и*

$$\Pr[A_i] \leqslant \varepsilon_i \prod_{j \in N(i)} (1 - \varepsilon_j),$$

то вероятность того, что не произошло ни одного из событий A_i, не меньше

$$(1-\varepsilon_1)\cdot(1-\varepsilon_2)\cdot\ldots\cdot(1-\varepsilon_n).$$

Таким образом, мы получаем в точности ту же оценку, что и для независимых событий, но условие сильнее: за каждое событие-соседа j, с которым независимости нет, мы расплачиваемся множителем $(1-\varepsilon_j)$ в правой части.

$\lhd$ Доказательство леммы Ловаса довольно странно — в нём вроде бы ни в какой момент ничего нетривиального не происходит, но придумать или даже запомнить его не так просто. Поэтому начнём для тренировки с простых замечаний:

(а) Если A и B — два события, то

$$\Pr[A|B]\leqslant\frac{\Pr[A]}{\Pr[B]}.$$

В самом деле, по определению условная вероятность $\Pr[A|B]$ есть $\Pr[A\wedge B]/\Pr[B]$, а $\Pr[A\wedge B]\leqslant\Pr[A]$. (Как обычно в логике, $\wedge$ означает «и».)

(б) Добавим всюду в предыдущем утверждение условие C («релятивизация»). Получим

$$\Pr[A|B\wedge C]\leqslant\frac{\Pr[A|C]}{\Pr[B|C]}.$$

В доказательстве леммы Ловаса это будет использоваться в ситуации, когда A и C независимы и потому числитель есть $\Pr[A]$, а знаменатель не слишком мал.

Теперь можно уже доказать лемму Ловаса по индукции. Как обычно, для индуктивного рассуждения полезно усилить утверждение. Будем доказывать такие факты ($\neg$ означает отрицание, то есть переход к дополнению):

(1) для любого i и для любых $p,q,\ldots$ (отличных от i и различных между собой) выполняется неравенство

$$\Pr[A_i|\neg A_p\wedge\neg A_q\wedge\ldots]\leqslant\varepsilon_i;$$

(2) для любых непересекающихся множеств индексов $i,j,\ldots$ и $p,q,\ldots$ выполняется неравенство

$$\Pr[\neg A_i\wedge\neg A_j\wedge\ldots|\neg A_p\wedge\neg A_q\wedge\ldots]\geqslant(1-\varepsilon_i)\cdot(1-\varepsilon_j)\cdot\ldots$$

Заметим, что из первого утверждения следует второе, если слева от черты (среди $i,j,\ldots$) только одно утверждение: если вероятность события при некотором условии не больше ε_i, то вероятность его отрицания при том же условии не меньше $(1-\varepsilon_i)$.

Более того, это рассуждение можно продолжить и для большего числа событий в левой части:

$$\begin{aligned}\Pr[\neg A_i\wedge\neg A_j|\neg A_p\wedge\neg A_q\wedge\ldots]&=\\ &=\Pr[\neg A_i|\neg A_j\wedge\neg A_p\wedge\neg A_q\wedge\ldots]\cdot\Pr[\neg A_j|\neg A_p\wedge\neg A_q\wedge\ldots];\end{aligned}$$

остаётся оценить каждый из сомножителей по (1).

С другой стороны, из второго утверждения можно вывести первое. Для этого разделим условия в (1) на две группы: входящие в $N(i)$ и не входящие в $N(i)$. (Здесь i — номер события, стоящего в левой части.) Пусть N и F — конъюнкции отрицаний соответствующих условий (соседних и остальных). Тогда, следуя ранее описанной схеме, можно оценить вероятность так:

$$\Pr[A_i|N \wedge F] = \frac{\Pr[A_i \wedge N|F]}{\Pr[N|F]} \leqslant \frac{\Pr[A_i|F]}{\Pr[N|F]} = \frac{\Pr[A_i]}{\Pr[N|F]}.$$

Знаменатель в предположении (2) не меньше, чем произведение скобок $(1-\varepsilon_t)$ по всем $t \in N$, и остаётся посмотреть на предположение леммы, где при ε_i стоит как раз этот множитель.

Остаётся лишь объяснить, почему взаимное сведение (1) к (2) и обратно не приводит к порочному кругу. Когда мы сводим (1) к (2) с помощью только что рассмотренной оценки, общее количество условий в N и F равно количеству условий в (1). Обратное сведение (2) к (1), с которого мы начали, уменьшает количество условий в (1) по крайней мере на единицу в сравнении с исходным. $\triangleright$

Вот пример комбинаторной задачи, где может быть полезна лемма Ловаса.

$\boxed{\textbf{185}}$ В каждой клетке конечной ленты мы хотим написать число от 1 до N. При этом для каждой границы между клетками некоторые пары чисел (l, r) запрещены, то есть нельзя, чтобы слева от границы стояло l, а справа r. Докажите. что если (для каждой границы) доля запрещённых пар среди всех пар (которых N^2) не больше $1/8$, то заполнение возможно. [Указание, Сопоставим каждой границе событие (запрещение пары для случайного набора чисел). Соседей у каждого события два, при $\varepsilon_i = \frac{1}{2}$ условие леммы выполняется]

$\boxed{\textbf{186}}$ Докажите аналогичный результат с несколько худшими параметрами без ссылки на лемму: если множество плохих пар имеет меру меньше $1/16$, то существует заполнение без плохих пар. [Указание. В каждой позиции имеются более $3/4$ чисел, для которых допустимы по крайней мере $3/4$ правых соседей, и более $3/4$ чисел, для которых допустим по крайней мере $3/4$ левых соседей, Значит, для большинства чисел допустимы большинство соседей слева и большинство соседей справа. Взяв число из этого большинства, можно дописывать к нему справа и слева числа из того же большинства (поскольку два множества, больших половины, обязательно пересекаются).]

8.5.3. Лемма Ловаса и запрещённые слова

Применим лемму Ловаса для доказательства теоремы 133. По соображениям компактности достаточно доказать существование сколь угодно длинных конечных последовательностей без запрещённых подслов.

Будем считать, что биты последовательности равновероятны и независимы. Появление запрещённой последовательности длины n в данной позиции (на данном интервале I) имеет вероятность $2^{(\alpha-1)n}$, где n — длина интервала. В качестве оценки в лемме Ловаса для этого события возьмём $2^{(\beta-1)n}$ для некоторого $\beta \in (\alpha, 1)$. Надо подобрать β так, чтобы выполнялось условие леммы Ловаса.

Соседями события на интервале I являются события на интервалах J, которые перекрываются с I. Поскольку вероятности событий зависят от длины, при подсчётах удобно группировать интервалы по длинам. Имеется $n+k-1$ интервалов J длины k, перекрывающихся с данным интервалом I длины n. Для каждого из них в правой части оценки леммы Ловаса появляется множитель $(1-2^{(\beta-1)k})$, и всего получается

$$(1-2^{(\beta-1)k})^{n+k-1}.$$

Теперь надо это перемножить по всем k, начиная с некоторого N (если мы доказываем существование последовательности без плохих подслов короче N). Таким образом, для применения леммы Ловаса нам нужно, чтобы

$$2^{(\alpha-1)n} \leqslant 2^{(\beta-1)n} \cdot \prod_{k\geqslant N} \left(1-2^{(\beta-1)k}\right)^{n+k-1}$$

(На самом деле в нашем подсчёте есть погрешность — при $n=k$ мы учитываем сам интервал в качестве соседа, но это в нашу пользу.) Нам достаточно очень грубой оценки: заменим $n+k-1$ на nk, извлечём корень n-степени и используем неравенство Бернулли в правой части. Останется доказать:

$$2^{\alpha-\beta} \leqslant 1-\sum_{k\geqslant N} k2^{(\beta-1)k}$$

Бесконечный ряд $\sum_k k2^{(\beta-1)k}$ сходится при $\beta<1$, а левая часть меньше 1 при $\alpha<\beta$, так что при достаточно большом N неравенство выполняется.

(Таким образом, порядок выбора параметров такой: берём любое $\beta\in(\alpha,1)$, а затем подбираем N. Затем применяем лемму Ловаса к словам длины не меньше N и получаем, что существуют сколь угодно длинные конечные последовательности без запрещённых слов длины N и более (оценки выполнены при любой длине последовательности). Наконец, по компактности получаем бесконечную последовательность.)

$\boxed{\textbf{187}}$ `[two-dimensional-forbidden]` Докажите двумерный аналог теоремы 133: можно заполнить бесконечную клетчатую бумагу нулями и единицами так, чтобы любой прямоугольник достаточно большой площади не был запрещённым. (Предполагаем, что для каждого прямоугольника площади k выбрано не более $2^{\alpha k}$ запрещённых заполнений, где $\alpha<1$ — некоторая константа.) [Указание: несложно провести аналогичные оценки.]

8.5.4. Запрещённые подпоследовательности

До сих пор мы говорили о запрещённых *подсловах*, то есть комбинациях подряд идущих битов. Это ограничение выглядит несколько искусственно. Более общая постановка вопроса может быть такой: есть счётное число битовых переменных и некоторые ограничения; каждое ограничение запрещает конечному числу переменных принимать (одновременно) некоторые значения. Нас интересуют результаты такого типа: если ограничений «не слишком много», то они заведомо совместны (независимо от того, что конкретно запрещается).

Два ограничения независимы, если не содержат общих переменных. Поэтому, имея в виду лемму Ловаса, имеет смысл ограничить число ограничений, включающих данную переменную.

Начнём с обозначений. Пусть $\omega = \omega_0\omega_1\omega_2\ldots$ — некоторая последовательность. Для произвольного конечного множества $F \subset \mathbb{N}$ индексов через $\omega(F)$ обозначим двоичное слово, составленное из членов последовательности с индексами в F (в порядке возрастания индексов). Рассмотрим пару (F, X), где F — конечное множество индексов, а X — двоичное слово, длина которого равна числу элементов в F. Последовательность ω *запрещается* парой (F, X), если $\omega(F) = X$. Пары такого вида будем называть *запрещениями*, а число элементов в F (длину слова X) — *размером* запрещения. Запрещение *покрывает* индексы, входящие в F.

Theorem 137. [lovasz-rumyantsev] *Пусть дано действительное число $\alpha \in (0, 1)$ и множество запрещений (F, X), при этом для любого индекса i и числа n имеется не более $2^{\alpha n}$ запрещений размера n, которые покрывают i. Тогда существует число N и последовательность, не запрещённая ни одним из запрещений размера больше N.*

$\lhd$ По соображениям компактности достаточно доказать утверждение для конечных последовательностей (для некоторого N, не зависящего от длины последовательности).

Событием будет нарушение запрещения. Вероятность такого события для запрещения размера n есть 2^{-n}. Применяя лемму Ловаса, выберем в качестве соответствующего ϵ_i для события размера n значение $2^{-\beta n}$, где β — некоторая константа, большая α (как мы увидим, годится любая).

Соседями запрещения будут запрещения, пересекающиеся с ним (покрывающие общее число). Для применения леммы Ловаса надо взять запрещение размера n и проверить, что что 2^{-n} не больше $2^{-\beta n}$, умноженного на произведение множителей $(1 - 2^{-\beta m})$ по всем запрещениям, пересекающимся с данным.

Произведение разделим на части, соответствующие возможным точкам пересечения. Всего таких возможных точек n. (Если в пересечении несколько точек, произвольно выберем одну из них.) Кроме того, в каждой части расклассифицируем сомножители по размерам. Тогда для данной точки и для данного размера m получим не более $2^{\alpha m}$ сомножителей, каждый из которых есть $(1 - 2^{\beta m})$. Дальше надо перемножить по всем m и возвести в степень n (поскольку частей n). Таким образом, нам надо проверить неравенство

$$2^{-n} \leqslant 2^{-\beta n} \prod_{m>N} (1 - 2^{-\beta m})^{2^{\alpha m} n},$$

или (если убрать возведение в степень n)

$$2^{\beta - 1} \leqslant \prod_{m>N} (1 - 2^{-\beta m})^{2^{\alpha m}}.$$

По неравенству Бернулли это заведомо выполнено, если

$$2^{\beta - 1} \leqslant 1 - \sum_{m>N} 2^{\alpha m} 2^{-\beta m}.$$

Поскольку левая часть меньше 1, а убывающая геометрическая прогрессия

$$\sum_m 2^{(\alpha - \beta)m}$$

сходится, то при подходящем N нужное неравенство выполнено. (Порядок выбора: сначала берём $\beta \in (\alpha, 1)$, затем выбираем N, глядя на сходящийся ряд, затем для любой длины последовательности применяем лемму Ловаса, и, наконец, ссылаемся на компактность.) $\triangleright$

Другое доказательство теоремы 137, не ссылающееся на лемму Ловаса (а проводящее аналогичное ей рассуждение непосредственно), сообщили авторам Ан. А. Мучник и А. Л. Семёнов. Пусть фиксировано множество запрещений, удовлетворяющее предположениям теоремы; пусть s — минимальный размер запрещений, входящих в это множество.

Для каждого конечного множества индексов $I \subset \mathbb{N}$ через $c(I)$ обозначим количество разрешённых раскрасок I, то есть отображений I в $\{0, 1\}$, которые не нарушают запрещений. (При этом рассматриваются только запрещения (F, X) с $F \subset I$, поскольку вне I отображение не определено и сравнивать его с запрещением нельзя.) Для пустого I естественно положить $c(I) = 1$.

Выберем $\beta \in (\alpha, 1)$ и будем доказывать, что при добавлении одной новой точки к множеству I величина $c(I)$ увеличивается в 2^β раз (в предположении, что s достаточно велико). Отсюда будет следовать, что число разрешённых раскрасок на множестве размера k не меньше $2^{\beta k}$ и уж заведомо положительно — нам нужно только это, но по индукции приходится доказывать более сильное утверждение.

Итак, пусть мы добавили к I новую точку i, получив $I' = I \cup \{i\}$. Каждое разрешённая раскраска I порождает две раскраски на I' (для двух возможных значений в новой точке). Получаем $2c(I)$ раскрасок, но часть из них может быть запрещёнными и их надо вычесть. Кто может их запрещать? Поскольку раскраска на I разрешена, то запрещение должно содержать i в дополнение к некоторым точкам в I; пусть K — множество этих точек, а k — их количество. Сколько раскрасок нужно вычесть из $2c(I)$ из-за одного такого запрещения? Их не больше, чем разрешённых раскрасок $I \setminus K$, а по предположению индукции таковых не больше $n(I)/2^{\beta k}$. (В самом деле, если добавление точки увеличивает число разрешённых раскрасок по крайней мере в 2^β раз, то удаление одной точки уменьшает это число в то же число раз, а удаление k точек уменьшает это число по крайней мере в $2^{\beta k}$ раз.) Теперь вычитаемое надо просуммировать по всем k от $s - 1$ до I, а для каждого k по не более чем $2^{\alpha(k+1)}$ запрещениям, содержащим i и k элементов в I.

В итоге получаем оценку:

$$c(I') \geqslant 2c(I) - \sum_{k=s-1}^{|I|} 2^{\alpha(k+1)} \frac{c(I)}{2^{\beta k}},$$

которую можно ослабить (заменяя 2^α на 2 и продолжая суммирование до бесконечности) до

$$c(I') \geqslant 2c(I) \left(1 - \sum_{k \geqslant s-1} \frac{2^{\alpha k}}{2^{\beta k}}\right).$$

Поскольку ряд в правой части сходится, то при достаточно большом s множитель в скобках не меньше 2^β, и индуктивное рассуждение завершается. (Заметим, что мы применяли предположение индукции лишь к множествам размера меньше $|I|$, так что порочного круга не получается.)

8.5.5. Сложные подпоследовательности

Мы хотим перевести доказанную теорему на сложностной язык и доказать, что существует последовательность со сложными подпоследовательностями (а не только подсловами).

Можно ли надеяться, что сложность любой подпоследовательности достаточно большой длины m не меньше αm для некоторого $\alpha < 1$ (как это было для подслов)? Очевидно, нет: ведь можно включить в подпоследовательности те индексы, где стоят нули (или единицы). Но при этом сложность перейдёт в сложность набора индексов.

И действительно, мы уже встречали подобное утверждение в задаче 113 (для равномерной меры — теорема 86, с. 140, (д)): если последовательность ω случайна в смысле Мартин-Лёфа по равномерной мере, то

$$\mathit{KP}(F, \omega(F)) \geqslant |F| - c$$

для некоторого c и для всех конечных множеств F.

Но это не то, к чему мы стремимся (это скорее аналог теоремы о том, что у случайной последовательности сложные начальные отрезки, а не существования последовательности со сложными подсловами). Нас будет интересовать другое утверждение:

Theorem 138. [rumyantsev] *Пусть $\alpha \in (0,1)$ — действительное число. Существует последовательность ω и константа N, для которых*

$$\max_{t \in F} \mathit{KS}(F, \omega(F) | t) \geqslant \alpha |F|$$

при всех F, содержащих не менее N элементов.

Заметим, что отсюда вытекает такое свойство последовательности ω: всякое конечное множество F размера не менее N имеет элемент t, для которого

$$\mathit{KS}(\omega(F) | F, t) \geqslant \alpha |F| - 2\,\mathit{KS}(F | t)$$

(константу 2 написана для простоты, её можно уменьшить). Опуская t в левой части, получаем, что для любого конечного F выполнено неравенство

$$\mathit{KS}(\omega(F) | F) \geqslant \alpha |F| - 2 \max_{t \in F} \mathit{KS}(F | t).$$

В частности, если индексы представить себе расположенными в плоскости, а в качестве F брать прямоугольники, то вычитаемое в правой части будет логарифмическим и потому его можно скомпенсировать изменением параметра α. Получаем утверждение задачи 187 (с. 219).

$\lhd$ Теорема 138 является сложностной переформулировкой теоремы 137. В самом деле, в ней рассматривается множество запрещений (F, Z), для которых $K(F, Z | t) < \alpha |F|$ для всех $t \in F$. Значит, для любого индекса t количество таких запрещений размера k, у которых F содержит t, не превосходит $2^{\alpha k}$, что и составляет предположение теоремы 137. $\rhd$

[далее следовало бы изложить многоуровневую конструкцию Андрея Румянцева и её применения, в частности, для дробных периодов]

8.6. Доказательство одного неравенства

Мы уже говорили (с. 19), что неравенства для колмогоровской сложности имеют разные любопытные следствия. Подробно этот вопрос разбирается в главе 10, но одно следствие подобного рода мы приведём уже сейчас.

Theorem 139. [triple-function-inequality] *Пусть X, Y и Z — конечные множества, а $f\colon X \times Y \to \mathbb{R}$, $g\colon Y \times Z \to \mathbb{R}$ и $h\colon X \times Z \to \mathbb{R}$ — функции с неотрицательными значениями. Тогда*

$$\left(\sum_{x,y,z} f(x,y)g(y,z)h(x,z)\right)^2 \leqslant \left(\sum_{x,y} f^2(x,y)\right)\cdot\left(\sum_{y,z} g^2(y,z)\right)\cdot\left(\sum_{x,z} h^2(x,z)\right)$$

◁ Как ни странно, но это неравенство можно доказать с использованием неравенства

$$2\,\mathit{KP}(x,y,z) \leqslant \mathit{KP}(x,y) + \mathit{KP}(y,z) + \mathit{KP}(x,z) + O(\log n)$$

(теорема 26, с. 50). Нам удобно записать последнее неравенство для префиксной сложности, а не для обычной, как раньше, но это несущественно, так как разница есть $O(\log n)$. (На самом деле для префиксной сложности это неравенство верно и с точностью $O(1)$ (задача 84, с. 110), но для нас бóльшая точность не нужна.)

Будем для удобства считать, что множества X, Y и Z, о которых идёт речь в теореме, состоят из двоичных слов. Достаточно доказать, что если суммы в правой части неравенства равны единице, то сумма в левой части не превосходит 1. (В самом деле, от умножения функции f на число и левая, и правая части возрастают в одинаковое число раз, так что можно «нормировать» f; аналогично для g и h.)

Итак, пусть $\sum_{x,y} f^2(x,y) = 1$ и аналогично для двух других сумм. Нам надо доказать, что $\sum_{x,y,z} f(x,y)g(y,z)h(x,z) \leqslant 1$.

Идея проста: функция f^2 задаёт распределение вероятностей на парах (x,y), поэтому $\mathit{KP}(x,y) \leqslant -\log f^2(x,y) = -2\log f(x,y)$ (если временно считать, что это распределение вероятностей меньше априорного не с точностью до константы, а прямо так). Аналогично $\mathit{KP}(y,z) \leqslant -2\log g(y,z)$ и $\mathit{KP}(x,z) \leqslant -2\log h(x,z)$. Поэтому (вспоминаем неравенство для $\mathit{KP}(x,y,z)$ и временно забываем о логарифмической добавке)

$$\mathit{KP}(x,y,z) \leqslant -\log f(x,y) - \log g(y,z) - \log h(x,z).$$

Поскольку сумма $2^{-\mathit{KP}(x,y,z)}$ по всем тройкам x,y,z не превосходит единицы (теорема 51, с. 90), получаем искомое неравенство.

На самом деле, конечно, все наши оценки носят асимптотический характер, и потому для настоящего доказательства нам надо перейти от отдельных слов к наборам слов. Вот как это делается.

Для начала заметим, что нам (по непрерывности) достаточно доказать неравенство для функций f, g и h с рациональными значениями.

Пусть N — произвольное натуральное число (которое потом будет стремиться к бесконечности). Рассмотрим множества X^N, Y^N и Z^N, элементы которых являются кортежами из N слов. Рассмотрим распределение вероятностей на $X^N \times Y^N = (X \times Y)^N$, соответствующее N независимым копиям распределения f^2 на $X \times Y$. (Формально говоря, вероятность точки $\langle\langle x_1, \ldots, x_N\rangle, \langle y_1, \ldots, y_N\rangle\rangle$ равна произведению $f^2(x_1, y_1) \ldots f^2(x_N, y_N)$.) Мы получаем семейство распределений, вычислимо зависящее от N, и потому найдётся такая константа c, что

$$\mathit{KP}(\langle x_1, \ldots, x_N\rangle, \langle y_1, \ldots, y_N\rangle | N) \leqslant 2\sum_i (-\log f(x_i, y_i)) + c$$

при всех N и при всех $x_1, \ldots, x_N, y_1, \ldots, y_N$ (сравниваем построенное распределение с априорной вероятностью). Можно убрать условие N в левой части неравенства, написав справа $c \log N$ вместо c. Далее, как и раньше, сложим три таких неравенства и оценим сложность тройки через сложность пар. При этом получится

$$\mathit{KP}(\langle x_1, \ldots, x_N\rangle, \langle y_1, \ldots, y_N\rangle, \langle z_1, \ldots, z_N\rangle) \leqslant \\ \leqslant \sum_i (-\log f(x_i, y_i)) + \sum_i (-\log g(y_i, z_i)) + \sum_i (-\log h(x_i, z_i)) + c \log N$$

для некоторой константы c и для всех N, $x_1, \ldots, x_N$, $y_1, \ldots, y_N$, $z_1, \ldots, z_N$ (заметим, что суммарная длина всех слов x_i, y_i, z_i при $i = 1, \ldots, N$ есть $O(N)$, поэтому все логарифмические добавки поглощаются $c \log N$). Подставляя эту оценку в неравенство $\sum_u 2^{-\mathit{KP}(u)} \leqslant 1$, получаем, что при любом N сумма

$$\sum \prod_i f(x_i, y_i) g(y_i, z_i) h(x_i, z_i)$$

по всем наборам $x_1, \ldots, x_N$, $y_1, \ldots, y_N$, $z_1, \ldots, z_N$ не превосходит $2^{O(\log N)}$, то есть некоторого полинома от N. Но эта сумма есть N-я степень суммы

$$\sum_{\langle x, y, z\rangle \in X \times Y \times Z} f(x, y) g(y, z) h(x, z),$$

поэтому если бы эта последняя сумма была больше 1, то получилось бы противоречие. Неравенство доказано. $\triangleright$

188 Покажите, что из доказанного неравенства следует оценка для объёма трёхмерного тела через площади его плоских проекций, упомянутая на с. 19. [Указание: в качестве f, g, h можно взять характеристические функции проекций тела; это даёт необходимую оценку в дискретном случае, а для непрерывного надо перейти к пределу либо разбивая тело на кубики, либо устремляя сумму к интегралу.]

Для сравнения приведём два других доказательства того же неравенства. Вот (простое) доказательство того же неравенства с помощью неравенства Коши (которое гласит, что

$(u,v)^2 \leqslant \|u\|^2 \cdot \|v\|^2$, в координатной записи $(\sum u_i v_i)^2 \leqslant (\sum u_i^2)(\sum v_i^2))$. В самом деле,

$$\left(\sum_{x,y,z} f(x,y)g(y,z)h(x,z)\right)^2 \leqslant$$

$$\leqslant \left(\sum_{x,y} f^2(x,y)\right)\left(\sum_{x,y}\left(\sum_z g(y,z)h(x,z)\right)^2\right) \leqslant$$

$$\leqslant \left(\sum_{x,y} f^2(x,y)\right)\sum_{x,y}\left(\left(\sum_z g^2(y,z)\right)\left(\sum_z h^2(x,z)\right)\right) =$$

$$= \left(\sum_{x,y} f^2(x,y)\right)\left(\sum_{y,z} g^2(y,z)\right)\left(\sum_{x,z} h^2(x,z)\right)$$

Ещё одно доказательство может быть получено с использованием шенноновской энтропии (и в каком-то смысле представляет собой «перевод» доказательства с колмогоровской сложностью). Итак, пусть $\sum f^2 = \sum g^2 = \sum h^2 = 1$. Мы хотим доказать, что $\sum_{x,y,z} p(x,y,z) \leqslant 1$, где $p(x,y,z) = f(x,y)g(y,z)h(x,z)$. Рассуждая от противного, предположим, что эта сумма равна $c > 1$. Тогда пропорционально уменьшим её члены, получив некоторое распределение вероятностей p' на $X \times Y \times Z$:

$$p'(x,y,z) = \frac{1}{c} f(x,y)g(y,z)h(x,z).$$

Соответствующую случайную величину со значениями в $X \times Y \times Z$ обозначим через ξ. Её можно рассматривать как тройку (вообще говоря, зависимых) случайных величин ξ_x, ξ_y и ξ_z; можно также рассматривать двумерные проекции $\xi_{xy} = \langle \xi_x, \xi_y \rangle$ и т. п. Например, величина ξ_{xy} принимает значение $\langle x, y\rangle$ с вероятностью $\sum_z p'(x,y,z)$. Напомним, что по определению шенноновская энтропия распределения $(p_1,\ldots,p_k)$ равна сумме $\sum p_i(-\log p_i)$ и не превосходит $\sum p_i(-\log q_i)$ для любого другого распределения $q_1+\ldots+q_k = 1$. Поэтому энтропию $H(\xi_{xy})$ можно оценить сверху, взяв в качестве другого распределения значения $f^2(x,y)$:

$$H(\xi_{xy}) \leqslant \sum_{x,y}\left(\sum_z p'(x,y,z)\right)(-2\log f(x,y)).$$

Записав аналогичные оценки для двух других направлений проекции и применив неравенство

$$H(\xi) = H(\xi_x,\xi_y,\xi_z) \leqslant \frac{1}{2}(H(\xi_{xy} + H(\xi_{yz}) + H(\xi_{xz})),$$

(задача 172, с. 196), получим, что

$$H(\xi) \leqslant \sum_{x,y,z} p'(x,y,z)(-\log f(x,y) - \log g(y,z) - \log h(x,z)) =$$

$$= \sum_{x,y,z} p'(x,y,z)(-\log p(x,y,z)).$$

Но по определению $H(\xi) = \sum_{x,y,z} p'(x,y,z)(-\log p'(x,y,z))$, и получается противоречие, так как p' меньше p в c раз (и потому $-\log p'$ больше $-\log p$ на $\log c$).

8.7. Нетранзитивность липшицевых преобразований

В этом разделе мы рассмотрим приложение колмогоровской сложности к анализу свойств бесконечных последовательностей. Начнём с такого определения, связанного с канторовским пространством Ω бесконечных двоичных последовательностей.

Отображение $f\colon \Omega \to \Omega$ называется *липшицевым*, если

$$d(f(\omega_1), f(\omega_2)) \leqslant cd(\omega_1, \omega_2)$$

для некоторой константы c и для всех $\omega_1, \omega_2 \in \Omega$. Здесь d — расстояние между последовательностями в канторовском пространстве, определяемое как 2^{-k}, где k — номер первого места, где последовательности различаются.

Неформально говоря, это свойство означает, что первые n знаков последовательности $f(\omega)$ определяются первыми $n{+}O(1)$ знаками последовательности ω. Поэтому, в частности, все отображения, задаваемые локальными правилами (каждый бит в $f(\omega)$ определяется некоторой окрестностью этого бита в ω), являются липшицевыми.

Нас будет интересовать такое свойство отображения f: для любых последовательностей ω_1 и ω_2 и любого $\varepsilon > 0$ существует число N и последовательности ω_1' и ω_2', для которых

$$\omega_2' = f(f(f(\ldots f(\omega_1')\ldots))) \quad (N \text{ раз})$$

и

$$d(\omega_1, \omega_1') < \varepsilon, \quad d(\omega_2, \omega_2') < \varepsilon.$$

(Другими словами, для любых двух открытых окрестностей найдётся орбита, начинающаяся в одной окрестности и попадающая в другую.) Это свойство мы будем для краткости называть "транзитивностью".

Легко понять, что отображение сдвига влево (отбрасывание первого члена последовательности) транзитивно: если мы хотим, чтобы последоватльность начиналась на некоторое слово x_1, а после нескольких сдвигов начиналась на другое слово x_2, достаточно начать её с x_1x_2.

Возникает вопрос, сохранится ли свойство транзитивности, если заменить d на так называемое *расстояние Безиковича*

$$\rho(\omega_1, \omega_2) = \limsup_{n\to\infty} d_n(\omega_1, \omega_2)/n,$$

где d_n — количество несовпадений среди первых n членов последовательностей, то есть количество тех $i < n$, при которых i-е члены ω_1 и ω_2 различны.

Оказывается, что в этом случае отображение сдвига уже не обладает аналогичным свойством (не является "B-транзитивным). Более того, имеет место следующая

Theorem 140. `[durand-cerville-bienvenu]` *Никакое липшицево отображение не является B-транзитивным.*

(Говоря о липшицевом отображении, мы имеем в виду исходное определение, с канторовским расстоянием.)

Причина этого проста: липшицево отображение почти не увеличивает сложности начальных отрезков последовательности, так как n битов выходной последовательности определяются $n + O(1)$ битами входной (и соответствующее правило можно считать вычислимым, выбрав подходящий оракул). С другой стороны, близкие в смысле Безиковича последовательности имеют близкие сложности начальных отрезков (поскольку изменение небольшой доли среди первых n битов кодируется малым количеством битов по сравнению с n).

$\triangleleft$ Формально удобно воспользоваться понятием эффективной хаусдорфовой размерности последовательности (которая равна $\lim inf\, KS(\omega_0 \ldots \omega_{n-1})/n$ для одноэлементного множества $\{\omega\}$), см. теорему 97 на с. 150 в разделе 5.8).

Лемма 1. Вычислимое липшицево отображение не увеличивает эффективную хаусдорфову размерность последовательности.

(Говоря о вычислимости липшицева отображения $f\colon \Omega \to \Omega$, мы имеем в виду, что n первых знаков $f(\omega)$ алгоритмически определяются по $n + c$ знакам ω для некоторого c.)

В самом деле, если $f(\omega_1) = \omega_2$, то сложность начального отрезка длины n последовательности ω_2 не превосходит константы (зависящей от сложности алгоритма для f) плюс сложность начального отрезка длины $n + c$ последовательности ω_1, и в пределе эти константы не играют роли.

Лемма 2. Если $\rho(\omega_1, \omega_2) < \varepsilon$, то эффективные хаусдорфовы размерности последовательностей ω_1 и ω_2 отличаются не более чем на $H(\varepsilon)$.

(Здесь $H(\varepsilon)$ — шенноновская энтропия случайной величины с двумя значениями, имеющими вероятности ε и $1 - \varepsilon$.)

В самом деле, если первые n членов последовательностей ω_1 и ω_2 отличаются в k позициях, то сложности соответствующих начальных отрезков отличаются не более чем на сложность их разности (или суммы) по модулю 2, а последовательность длины n с k единицами имеет сложность не более $nH(k/n) + O(\log n)$ (см. раздел 7.3.1, теорема 122, с. 197); отсюда легко следует утверждение леммы.

Таким образом, если мы берём последовательность нулевой размерности (скажем, вычислимую), то любая близкая к ней последовательность (в смысле Безиковича) имеет малую размерность, а вычислимое липшицево отображение не увеличивает размерность, и поэтому могут получиться только последовательности малой размерности. С другой стороны, любая последовательность, близкая по Безиковичу к случайной последовательности (которая имеет размерность 1), имеет (по той же лемме 2) размерность, близкую к единице.

Таким образом, для вычислимых липшицевых отображений теорема доказана. Остаётся заметить, что все эти рассуждения можно релятивизовать относительно любого оракула и что любое липшицево отображение вычислимо относительно некоторого оракула. $\triangleright$

[Как правильно называется это свойство? транзитивность? где поставлен вопрос и где дан ответ? - надо спросить у Лорана]

8.8. Эргодическая теорема

доказательство Вьюгина? сохранились ли записи?

9. Частотный подход к определению случайности

[mises]

9.1. Исходный замысел фон Мизеса

[misesorig]

Сейчас, когда все привыкли к построению теории вероятностей на основе теории меры, требуются специальные усилия, чтобы забыть об этом и попытаться оценить замысел Рихарда фон Мизеса, который в начале XX века предложил строить теорию вероятностей на основе понятия случайной последовательности, или, как он говорил, *коллектива*. Тем не менее попробуем.

В природе бывают явления, которые легко предсказывать (после того как изучены управляющие ими законы), например, движение планет. Но бывают и другие явления: как бы мы ни старались предсказать результат бросания монеты, обычно около половины предсказаний оказываются неверными. Именно такие явления и составляют предмет теории вероятностей.

Тем самым основным понятием теории вероятностей становится (более или менее неопределяемое) понятие *коллектива* — трудно предсказуемой последовательности символов (букв некоторого алфавита, который мы для простоты будем считать состоящим из нуля и единицы). А основным свойством коллектива $\omega = \omega_0\omega_1\ldots$ является следующее свойство *устойчивости частот*:

> существует предел
> $$p = \lim_{n\to\infty} \frac{\omega_0 + \omega_1 + \ldots + \omega_{n-1}}{n}$$
> и, более того, это свойство сохраняется (с тем же p), если из всей последовательности выбрать некоторую подпоследовательность (скажем, оставить каждый второй член, или члены с составными номерами, или члены, стоящие после единиц).

Указанное p и называют *вероятностью* появления единицы (в данном коллективе).

Существование коллективов обосновывается «экспериментально», ссылкой на успешную деятельность игорных домов, которые бы разорились, если бы существовала система игры, позволяющая выбирать подпоследовательности, не обладающие свойством устойчивости частот.

Так — или примерно так — говорил фон Мизес (см., например, его книгу “Wahrscheinlichkeit, Statistik und Wahrheit” [54]; эта книга была переведена на русский язык под названием «Вероятность и статистика» [55] — видимо, слово Wahrheit (истина) в названии показалось чрезмерно смелым посягательством на прерогативы единственно верного и подлинно научного учения). Но говорил он это не во времена Евклида или Спинозы, а в начале XX века, когда требования математической строгости были достаточно настойчивы. Допустим, существование последовательностей с некоторыми свойствами можно объявить аксиомой или экспериментальным фактом, это ещё куда ни шло (хотя, конечно, экспериментальный факт, говорящий нечто о *пределах* последовательностей, вызывает сомнения, если не насмешки). Но уж сформулировать математически точно свойство устойчивости частот следовало бы.

Проблема здесь в том, что не указано, какие подпоследовательности разрешено выбирать. Сам Мизес ограничивался лишь некоторыми примерами допустимых правил выбора подпоследовательностей (три таких правила мы привели), а также замечал, что решение о включении или невключении какого-либо члена последовательности в выбираемую подпоследовательность не должно зависеть от значения этого члена. В самом деле, иначе мы могли бы из любой последовательности выбрать подпоследовательность из одних единиц (или нулей), нарушив свойство устойчивости частот.

Поэтому возможны разные уточнения замысла Мизеса, и несколько таких уточнений мы рассмотрим. Для каждого из них прежде всего возникает вопрос о существовании коллективов (который после уточнения определения становится математическим); при этом вопрос о том, действительно ли бросание монеты даёт коллектив, позволительно считать естественно-научным или философским и оставить в стороне.

Для простоты мы в основном ограничиваемся случаем симметричной монеты ($p = 1/2$), если обратное не оговорено явно. Будем называть бесконечную последовательность нулей и единиц *сбалансированной*, если предел частоты единиц (или нулей) в ней существует и равен $1/2$.

9.2. Правила выбора как множества слов

[misessel]

Пожалуй, самое естественное толкование допустимого правила выбора состоит в следующем: мы принимаем решение о включении ω_n в подпоследовательность, глядя на все предыдущие члены, то есть в зависимости от $\omega_0\omega_1\ldots\omega_{n-1}$. Таким образом, допустимое правило выбора есть функция, отображающая все двоичные слова $\omega_0\ldots\omega_{n-1}$ в двухэлементное множество $\{\text{включать}, \text{не включать}\}$, или, что то же самое, некоторое множество R двоичных слов (соответствующих значению «включать»).

Формально говоря, для произвольного множества $R \subset \Xi$ мы определяем соответствующее ему правило выбора как отображение S_R, которое ставит в соответствие произвольной последовательности $\omega \in \Omega$ некоторую (конечную или бесконечную) подпоследовательность $S_R(\omega) \in \Sigma$. А именно, $S_R(\omega)$ состоит из тех ω_n, для которых $\omega_0\ldots\omega_{n-1} \in R$ (в том же порядке, в котором они шли в ω).

Пример: если принадлежность слова x к множеству R зависит лишь от длины слова x, то последовательность $S_R(\omega)$ получается из ω выбором членов ω_n, стоящих на заранее известных местах (при тех n, для которых слова длины n принадлежат R). Другой пример: правило «выбирать члены после единиц» соответствует множеству R, состоящему из всех слов, оканчивающихся на единицу.

Естественно ожидать, что если мы приходим в казино, где играют в орлянку, заранее наметив какое-то правило выбора такого рода, и делаем ставки в те моменты, когда правило выбора включает очередной член в подпоследовательность, то никакого преимущества мы не получим и предельная частота единиц в подпоследовательности будет равна $1/2$. Но это верно лишь при условии, что правило выбора фиксировано заранее: задним числом (зная последовательность) легко можно подобрать правило выбора, которое бы у этой последовательности выиграло. Другими словами, имеет место такой очевидный факт: для каждой последовательности ω существует множество R, при котором последовательность $S_R(\omega)$ состоит из одних нулей или одних единиц, и потому не сбалансирована. Поэтому нельзя определить коллектив как последовательность ω, для которой подпоследовательность $S_R(\omega)$ сбалансирована (или конечна) для всех множеств R: при таком определении коллективов вообще не будет.

Однако, как заметил Вальд [83], для любого *счётного* семейства правил выбора S_{R_i} (соответствующего счётному семейству множеств R_i) существует последовательность ω, которая обладает свойством устойчивости частот относительно всех правил этого семейства: при любом i последовательность $S_{R_i}(\omega)$ сбалансирована (или конечна).

Это вытекает из следующего утверждения:

Theorem 141. [wald-theorem] *Пусть R — произвольное множество слов. Тогда множество тех последовательностей $\omega \in \Omega$, для которых $S_R(\omega)$ бесконечна, но не сбалансирована, имеет меру нуль* (*относительно равномерной меры на пространстве бесконечных последовательностей нулей и единиц*).

Согласно этой теореме каждое правило выбора отбраковывает лишь нулевое множество. Счётное число правил выбора даёт счётное семейство нулевых множеств, объединение которых будет нулевым. Значит, неотбракованные последовательности останутся (и даже образуют множество меры 1).

$\triangleleft$ Утверждение теоремы вытекает из усиленного закона больших чисел (согласно которому все последовательности, за исключением множества меры нуль, сбалансированы, см. раздел 3.2), а также следующей леммы.

Лемма. Пусть $U \subset \Omega$ — множество меры нуль. Тогда его прообраз $S_R^{-1}(U)$ имеет меру нуль.

Доказательство леммы. В самом деле, каждый следующий бит последовательности $S_R(\omega)$ имеет равные шансы оказаться нулём и единицей (при известных предыдущих битах), и разница с равномерным распределением лишь в том, что он может вообще не появиться — но от этого вероятность только уменьшается.

(Наглядно это иллюстрируется старинной загадкой: изменится ли доля мужчин, если после рождения первого сына женщины больше не будут рожать детей, чтобы наследник был единственным? Ответ отрицательный по тем же причинам.)

Формально это рассуждение проводится так:

Рассмотрим множество Σ_x всех конечных и бесконечных продолжений слова x и два его подмножества Σ_{x0} и Σ_{x1}. Докажем, что прообразы этих подмножеств при отображении S_R имеют равную меру (другими словами, что появление в подпоследовательности нуля и единицы после данного слова x одинаково вероятно).

В самом деле, рассмотрим все слова z, из которых при выборе по правилу S_R получается x и для которых $z \in R$. Они соответствуют ситуациям, когда слово x уже выбрано и вот-вот будет выбран следующий бит. Из определения следует, что два таких слова не сравнимы друг с другом. Поэтому множества Ω_{z0} для всех таких z не пересекаются, и, как легко понять, вместе составляют прообраз множества Σ_{x0}. Аналогично прообраз множества Σ_{x1} есть объединение непересекающихся множеств вида Ω_{z1}. Мы разбили прообразы на равные части и, значит, меры прообразов равны.

Из доказанного по индукции легко следует, что мера прообраза множества Σ_x не превосходит $2^{-l(x)}$. Поэтому прообраз нулевого множества $U \subset \Omega$ является нулевым. В самом деле, возьмём покрытие множества U интервалами Ω_{x_i} с малой суммой мер; рассмотрим прообразы множеств Σ_{x_i}; каждый из них есть объединение счётного числа интервалов. Объединив все эти интервалы, получим покрытие множества $S_R^{-1}(U)$ с малой суммой мер. Лемма (а с ней и теорема 141) доказана. $\triangleright$

Отметим ещё, что из доказанного стандартным для теории меры способом следует, что

$$\mu(S_R^{-1}(U)) \leqslant \mu(U)$$

для любого измеримого множества $U \subset \Omega$. В случае, когда $S_R(\omega)$ бесконечно для всех (или почти всех) ω, верно и более сильное утверждение: $S_R(\omega)$ имеет равномерное распределение, то есть

$$\mu(S_R^{-1}(U)) = \mu(U)$$

для любого измеримого U.

189 Пусть фиксировано некоторое множество R. Покажите, что если ω имеет бернуллиево распределение (испытания независимы и имеют одну и ту же вероятность успеха), то подпоследовательность $S_R(\omega)$ имеет то же распределение (когда бесконечна).

Итак, понятие коллектива становится непустым (коллективы существуют), если из всех возможных правил выбора (соответствующих всем возможным подмножествам $R \subset \Xi$) выбрать некоторое счётное семейство. Но что это может быть за семейство?

9.3. Случайность по Мизесу – Чёрчу

[miseschurch]

С появлением теории вычислимых функций стало ясно, что естественно рассматривать правила выбора S_R, соответствующие всевозможным разрешимым множествам R. Это предложил А. Чёрч [11], поэтому такие правила выбора называют *допустимыми по Чёрчу*, а соответствующие последовательности (сохраняющие устойчивость частот при всех допустимых по Чёрчу правилах) называют *случайными по Мизесу – Чёрчу*.

Мы уже знаем, что таковые существуют и образуют множество меры 1. Верно и более сильное утверждение:

Theorem 142. [martinlof-church] *Всякая случайная в смысле Мартин-Лёфа последовательность (относительно равномерной меры) случайна по Мизесу – Чёрчу.*

$\triangleleft$ Эффективный вариант усиленного закона больших чисел (теорема 32, с. 67; см. также раздел 8.4) говорит, что множество U несбалансированных последовательностей (у которых предел частот не существует или не равен $1/2$) является эффективно нулевым.

Покажем, что если S_R — допустимое по Чёрчу правило выбора, то прообраз эффективно нулевого множества будет эффективно нулевым. В самом деле, при разрешимом R конструкция в доказательстве теоремы 141 становится эффективной (все интервалы, составляющие прообраз данного интервала, можно эффективно перечислять). Поэтому случайная по Мартин-Лёфу последовательность не принадлежит этому прообразу, то есть её образ сбалансирован (или конечен). $\triangleright$

Что можно доказать про случайные по Мизесу – Чёрчу последовательности, кроме усиленного закона больших чисел, которому они удовлетворяют по определению? Например, можно доказать, что частоты появления не только букв, но и подслов таковы, какими они должны быть:

Theorem 143. `[mises-frequencies]` *Пусть ω — случайная по Мизесу – Чёрчу последовательность, а U — некоторое двоичное слово. Рассмотрим те позиции k, начиная с которых U встречается в ω (то есть $U_0U_1\ldots = \omega_k\omega_{k+1}\ldots$). Тогда доля таких позиций среди первых N позиций стремится к $1/2^{l(U)}$ при $N \to \infty$.*

$\lhd$ По условию нули встречаются примерно в половине позиций. Рассмотрим правило «выбирать после нулей». По предположению в выбранной подпоследовательности будет примерно половина нулей, что означает, что после нулей одинаково часто бывают нули и единицы (то есть что частота групп 00 и 01 стремится к $1/4$). Теперь рассмотрим правило «выбирать после 00» (или после 01) и так далее. $\rhd$

$\boxed{\textbf{190}}$ Докажите, что если случайную по Мизесу – Чёрчу последовательность разрезать на блоки длиной k, то каждый из 2^k возможных блоков будет встречаться с частотой, стремящейся к $1/2^k$. [Это утверждение отличается от предыдущей теоремы тем, что здесь учитываются блоки не на всех местах, а лишь на кратных k; его доказательство аналогично.]

Последовательности из теоремы 143, в которых каждая комбинация символов встречается с одинаковой частотой (в пределе), рассматривались и до Мизеса и назывались «нормальными» (тот же термин применяют и к действительным числам, двоичными записями которых являются нормальные последовательности).

$\boxed{\textbf{191}}$ Покажите, что при умножение на рациональное число нормальность действительного числа сохраняется. [Указание. При умножении столбиком «в уме» достаточно хранить конечное количество информации.]

[Вроде бы это доказал J. Wall в своей PhD диссертации в 1949 году (University of California, Berkeley), а также это написано в книге L. Kuipers, H. Niederreiter, Uniform Distribution of Sequences, Wiley-Interscience, 1949. Но хорошо бы проверить, а ещё лучше изложить доказательство понятно. Ещё хорошо бы узнать, зависит ли нормальность действительного числа от системы счисления. Видимо, зависит — надо посмотреть в Кнуте. Что известно про нормальность, скажем, e, π или $\sqrt{2}$?]

Среди нормальных последовательностей есть и вычислимые: например, если выписать подряд числа $1, 2, 3, \ldots$ в двоичной записи и объединить цифры в одну последовательность, то получится нормальная последовательность

$$\underline{1}\,\underline{10}\,\underline{11}\,\underline{100}\,\underline{101}\,\underline{110}\,\underline{111}\,\underline{1000}\,\underline{1001}\,\underline{1010}\,\underline{1011}\ldots$$

(пример Бореля).

$\boxed{\textbf{192}}$ Докажите это. [Указание. Каков бы ни был размер блока k, начиная с некоторого места границы между записями натуральных чисел не сильно меняют частоты, а в среднем по всем натуральным числам заданной длины частоты правильны.]

[это построение годится только для одной системы счисления. Построение абсолютно нормальных чисел: для каждой системе есть нулевое по Шнорру множество, объединение тоже, есть вычислимая функция вне него; по существу это было в неопубликованной работе Тьюринга (Becher)]

Но случайная по Мизесу – Чёрчу последовательность, разумеется, не может быть вычислимой, иначе было бы вычислимо правило выбора, отбирающее члены на тех местах, на которых в последовательности нули (единицы). Более того, легко видеть, что верна такая

Theorem 144. [mises-guess] *Для всякого всюду определённого алгоритма, пытающегося предсказывать следующий член последовательности по её предыдущим членам, доля успешных предсказаний на случайной по Мизесу – Чёрчу последовательности стремится к* $1/2$.

$\triangleleft$ В самом деле, каждому предсказывающему алгоритму соответствуют два правила выбора: одно отбирает те члены, где предсказан нуль, другое — где единица. Тем самым последовательность разбивается в «смесь» двух своих подпоследовательностей. Каждая из них сбалансирована (или конечна, но тогда всё очевидно), и потому на ней доля успешных предсказаний стремится к $1/2$. Значит, и общая доля успешных предсказаний стремится к $1/2$. $\triangleright$

Это утверждение можно ещё несколько обобщить. Представим себе такую игру: перед появлением следующего члена последовательности мы можем поставить некоторую сумму (неотрицательное рациональное число, не превосходящее единицы) на один из исходов (нуль или единицу). Если мы угадали, то ставка удваивается, если не угадали — пропадает. Наша стратегия в такой игре задаётся функцией S, определённой на двоичных словах и принимающей рациональные значения в отрезке $[-1, 1]$. (Положительные значения соответствуют ставке на 0, отрицательные — на 1.) Суммарный выигрыш стратегии S на начальном отрезке $\omega_0 \ldots \omega_{n-1}$ можно записать как

$$\sum_{i=0}^{n-1} S(\omega_0 \ldots \omega_{i-1}) \cdot (-1)^{\omega_i}$$

[отрицательные значения соответствуют проигрышу — мы сейчас разрешаем игру в долг]

Theorem 145. [continuous-bets] *Пусть S — всюду определённая вычислимая стратегия описанного вида, а ω — случайная по Мизесу – Чёрчу последовательность. Тогда выигрыш S на начальном отрезке длины n последовательности ω есть $o(n)$.*

$\triangleleft$ Пусть сначала стратегия принимает только значения 1 и -1. Тогда её деятельность сводится к уже рассмотренному угадыванию следующего члена; доля успехов стремится к $1/2$, что в точности и означает, что средний выигрыш (в расчёте на одну игру) стремится к нулю.

Если стратегия S принимает лишь значения, кратные $1/k$ для некоторого целого k (от -1 до 1 с шагом $1/k$), то её можно представить как среднее арифметическое $2k$ стратегий, принимающих значения -1 и 1; при этом выигрыш S будет средним арифметическим выигрышей всех этих стратегий, и если для каждой из них выигрыш есть $o(n)$, то и для среднего арифметического он будет равен $o(n)$. (Мы предполагаем, что число k фиксировано и не зависит от n.)

Осталось перейти к случаю вычислимой стратегии с произвольными рациональными значениями. Для произвольного $\varepsilon > 0$ мы должны показать, что выигрыш S на начальном отрезке длины n не превосходит (по модулю) εn для всех достаточно бо́льших n.

Выберем k так, чтобы $1/k$ было меньше ε, и приблизим нашу стратегию S стратегией S' со значениями, кратными $1/k$ (заменив каждое значение на ближайшее кратное; ошибка будет не больше $\varepsilon/2$). Для стратегии S' выигрыш есть $o(n)$ и потому меньше $(\varepsilon/2)n$ для достаточно больших n, а разница между выигрышем для S и S' не превосходит $(\varepsilon/2)n$. $\triangleright$

Ещё одно естественное свойство (упоминавшееся фон Мизесом как одно из основных свойств коллективов):

Theorem 146. *Применение допустимого по Чёрчу правила выбора к случайной по Мизесу – Чёрчу последовательности даёт либо конечную, либо случайную по Мизесу – Чёрчу последовательность.*

$\triangleleft$ В самом деле, последовательное применение двух правил выбора сводится к однократному: если мы сначала предварительно отбираем члены в подпоследовательность, а потом просматриваем члены подпоследовательности и выбираем из них только часть, то в конечном счёте решение определяется предыдущими членами. (А композиция разрешимых правил будет, очевидно, разрешимой.) $\triangleright$

(В дальнейшем (раздел 9.12, с. 255) мы рассмотрим более общий класс правил (немотонные правила выбора) и соответственно изменённое определение случайности по Мизесу. При этом окажется, что новый класс правил не замкнут относительно композиции, и, более того, соответствующий класс последовательностей не замкнут относительно правил выбора, см. теорему 177, с. 273.)

Естественный вопрос: как соотносится случайность по Мизесу – Чёрчу и Мартин-Лёфу? Как мы впоследствии увидим, случайность по Мартин-Лёфу — более сильное требование.

Но сначала ещё несколько замечаний по поводу определения Мизеса.

9.4. Пример Вилля

[misesville]

Мы уже знаем, что для любого счётного семейства множеств R_i существует последовательность, которая обладает свойством устойчивости частот относительно всех правил выбора S_{R_i} — поскольку множество таких последовательностей имеет меру 1. Но можно дать и более прямую конструкцию такой последовательности, следуя работам Вилля [79] и Лавлэнда [40].

[Неплохо бы всё-таки прочесть работу Вилля и выяснить, та ли там конструкция на самом деле! Лавлэнд ссылается тоже не на Вилля, а на работу Levin, Minsky, Silver]

Для начала рассмотрим случай единственного правила S_R, соответствующего множеству R. В этом случае легко построить последовательность ω, для которой $S_R(\omega) = 01010101\ldots$ (нули и единицы чередуются) и потому последовательность $S_R(\omega)$ сбалансирована. В самом деле, будем строить ω постепенно (слева направо). Когда правило S_R предлагает выбрать очередной член подпоследовательности, мы смотрим, каким по счёту будет этот член (чётным или нечётным) и в зависимости от этого ставим в ω нуль или единицу. (Те члены ω, которые не выбраны правилом S_R, можно сделать любыми.)

Пусть теперь имеется конечное число множеств $R_1, \ldots, R_m$. Мы хотим построить последовательность ω, в которой выполнено свойство устойчивости частот относительно любого из правил S_{R_i}. Для каждого члена ω_n последовательности ω посмотрим, какие из правил

S_{R_i} (при $i = 1, \ldots, m$) включают его в подпоследовательность, и составим m-битовый вектор, содержащий эту информацию. Тем самым любая последовательность ω расслаивается на 2^m подпоследовательностей, соответствующих 2^m значениям этого битового вектора. (Некоторые из этих подпоследовательностей могут быть конечны.)

Построим последовательность ω, для которой все эти 2^m подпоследовательностей будут иметь вид $010101\ldots$ (конечная или бесконечная последовательность, в которой нули и единицы чередуются, начиная с нуля). В самом деле, пусть $\omega_0 \ldots \omega_{n-1}$ уже известны. Как определить ω_n? К этому моменту уже ясно, какой битовый вектор соответствует ω_n (какие правила его выберут) и тем самым известно, в какую из 2^m подпоследовательностей попадёт ω_n. Остаётся посмотреть, каким по счёту (чётным или нечётным) членом этой подпоследовательности будет ω_n.

Заметим, что $S_{R_i}(\omega)$ будет смесью 2^{m-1} подпоследовательностей, а именно, тех, у которых в битовом векторе на i-м месте стоит единица, и потому обладает свойством устойчивости частот (любой начальный отрезок последовательности $S_{R_i}(\omega)$ содержит не меньше нулей, чем единиц, а разница в количестве нулей и единиц ограничена числом 2^{m-1}, поскольку в каждой подпоследовательности разница не больше единицы).

Осталось рассмотреть общий случай счётного семейства правил $R_1, R_2, \ldots$. Мы будем постепенно подключать эти правила, в каждый момент имея дело лишь с конечным числом правил. Вот как это делается.

Пусть уже построен некоторый начальный отрезок $\omega_0 \ldots \omega_{n-1}$ последовательности ω. Тем самым уже известно, каким из множеств R_i он принадлежит (и какие правила S_{R_i} выберут следующий, пока ещё не построенный, член ω_n последовательности). Теперь эта информация является не m-битовым вектором, а бесконечной последовательностью битов $u_1u_2\ldots$ (бит u_i равен единице, если правило S_{R_i} выбирает следующий член). Другими словами, у нас имеется путь $u_1u_2\ldots$ в бесконечном двоичном дереве.

Зафиксируем достаточно быстро растущую последовательность $k_0 < k_1 < k_2 < \ldots$ натуральных чисел. Например, пусть $k_i = 2^{2i}$. На каждом шаге конструкции (то есть для каждого члена последовательности) одна из вершин двоичного дерева будет считаться *активной*. А именно, двигаясь вдоль пути $u_1u_2\ldots$, найдём первую вершину, которая была активной менее k_i раз, где i — высота вершины в дереве, и объявим активной её. Другими словами, активной вершиной будет кратчайшее слово x, для которого

- i-ый бит x равен единице тогда и только тогда, когда правило S_{R_i} выбирает ω_n;
- до сих пор (в процессе построения начального отрезка $\omega_0 \ldots \omega_{n-1}$) вершина x была активна менее $k_{l(x)}$ раз.

Таким образом строимая последовательность $\omega_0\omega_1\ldots$ разлагается в смесь счётного числа (конечных) подпоследовательностей, соответствующих счётному числу возможных активных вершин. Подпоследовательность, соответствующая вершине x, имеет длину не более $k_{l(x)}$, но может быть короче.

Как и раньше, мы строим последовательность ω так, чтобы каждая из этих подпоследовательностей имела вид $010101\ldots$. Теперь, правда, последовательность ω будет смесью *бесконечного* числа (конечных) последовательностей, и надо аккуратно проверить, что правило выбора S_{R_i} выберет из неё сбалансированную последовательность.

Выбранная этим правилом подпоследовательность состоит из членов двух типов. Во-первых, она содержит члены, для которых активные вершины короче i (и правило S_{R_i} вообще не учитывалось при определении этих членов). Во-вторых, она содержит члены, которым соответствуют активные вершины, в которых i-й бит равен единице. Количество членов первого типа ограничено (не больше $2^0 k_0 + \ldots + 2^{i-1} k_{i-1}$) и в дальнейшем мы ими пренебрегаем.

Что касается членов второго типа, то для каждой использованной активной вершины числа нулей и единиц, ей соответствующих, отличаются не более чем на 1. Поэтому если самая длинная активная вершина, встретившаяся при построении некоторого начального отрезка последовательности S_{R_i}, имеет длину $N \geqslant i$, то разница между числом нулей и единиц (второго рода) в этом начальном отрезке не больше числа активных вершин, использованных при его построении, то есть $O(2^N)$, а длина отрезка не меньше k_{N-1}, поскольку предшествующие активные вершины должны быть использованы полностью, прежде чем мы перейдём к более длинной. Остаётся заметить, что $2^N = o(k_{N-1})$.

Итак, мы описали явную конструкцию последовательности, обладающей свойством устойчивости частот для любого счётного семейства правил выбора. Что даёт нам эта конструкция (по сравнению с уже известным нам доказательством существования по соображениям меры)? Например, можно заметить, что поскольку каждая из смешиваемых последовательностей $010101\ldots$ начинается с нуля, то любой начальный отрезок их смеси содержит не меньше нулей, чем единиц. Отсюда получаем такое следствие:

Theorem 147. [mises-church-ville] *Существует случайная по Мизесу – Чёрчу последовательность, в любом начальном отрезке которой не меньше нулей, чем единиц.*

Из этого можно было бы уже вывести, что существует последовательность, случайная по Мизесу – Чёрчу, но не по Мартин-Лёфу, если бы мы знали, что множество всех последовательностей, в которых все начальные отрезки содержат не меньше нулей, чем единиц, является эффективно нулевым. Это действительно так и следует из эффективного варианта закона повторного логарифма — но, к сожалению, не той его части, которую мы доказали в разделе 8.4 (теорема 132).

193 [ville-1] Покажите, что в данном случае достаточно знать, что указанное множество нулевое, и что из этого можно вывести, что оно эффективно нулевое. [Указание. Пусть p_n — вероятность того, что вплоть до длины n начальные отрезки содержат не меньше нулей, чем единиц. Очевидно, последовательность p_n убывает и вычислима, и её предел есть мера указанного множества, поэтому для любого рационального $\varepsilon > 0$ можно дождаться момента, когда p_n станет меньше ε.]

194 [ville-2] Докажите, что указанное множество является эффективно нулевым, не ссылаясь на закон повторного логарифма. [Указание. Для каждого n вероятность того, что в начальном отрезке длины n нулей не меньше, чем единиц, примерно равна $1/2$. Если взять быстро растущие значения n, то эти события близки к независимым (отклонение на меньшем отрезке мало по сравнению с ожидаемым отклонением на большем отрезке).]

Мы не будем проводить эти рассуждения подробно. Вместо этого мы покажем, что существуют случайные по Мизесу – Чёрчу последовательности с логарифмической сложностью начальных отрезков (что невозможно для случайных по Мартин-Лёфу последовательностей) с помощью той же конструкции.

Theorem 148. [mises-church-logarithm] *Существует случайная по Мизесу – Чёрчу последовательность* $\omega = \omega_0\omega_1\ldots$, *для которой*

$$KS(\omega_0\ldots\omega_{n-1}) = O(\log n).$$

$\lhd$ Мы хотим применить описанное построение ко всем правилам выбора, используемым при определении случайности по Мизесу – Чёрчу, то есть взять в качестве R_i все разрешимые множества слов. При этом построение не удаётся сделать алгоритмическим, поскольку мы не можем эффективно перечислять разрешающие алгоритмы для всех разрешимых множеств. (Что не удивительно — иначе бы получилась вычислимая случайная по Мизесу – Чёрчу последовательность.)

Мы можем подряд перебирать все программы, но надо знать, какие из этих программ задают разрешимые множества (остальные программы можно пропустить, заменив каким-либо фиксированным правилом выбора). Такая информация для первых m программ занимает m битов (по одному для каждой программы) и позволяет проводить нашу конструкцию, пока мы не дойдём до активных слов длины m. А к этому моменту построенная последовательность будет иметь длину не меньше $k_{m-1} = 2^{2m-2}$. Тем самым объём дополнительной информации — логарифмический (по отношению к длине начального отрезка), что и требовалось. $\rhd$

Отметим ещё раз уже упоминавшееся следствие:

Theorem 149. *Существует случайная по Мизесу – Чёрчу последовательность, не являющаяся случайной по Мартин-Лёфу.*

Возникает естественный вопрос: не следует ли усилить требования к последовательности, предъявляемые определением случайности по Мизесу – Чёрчу? Например, можно разрешить переменные ставки, а также немонотонные правила выбора. В следующих разделах этой главы мы рассмотрим получающиеся при этом определения.

9.5. Мартингалы

[misesmart]

Говоря о существовании коллективов, мы апеллировали к азартным играм. Но с этой точки зрения, надо признать, рассматриваемая ситуация выглядит довольно странно: клиент приходит в казино, где бросают честную монету, выбирает некоторые бросания и затем «выигрывает» (точнее, опровергает гипотезу о честности монеты), если результаты выбранных бросаний не имеют предельной частоты $1/2$.

Приближаясь к игровой терминологии, мы можем сказать, что клиент делает ставку фиксированного размера (скажем, рубль) в избранных играх (при этом казино при необходимости верит ему в долг); требуется, чтобы его средний выигрыш (в расчёте на одну ставку) стремился к нулю. А если это не так, то клиент объявляет монету нечестной.

Более естественным представляется другой вариант игры. Пусть мы приходим в казино, имея рубль. Перед каждым бросанием мы делим наш капитал на две части, которые ставим на нуль и на единицу. Та часть, где мы не угадали, пропадает, а где угадали — удваивается, и мы продолжаем игру. (Например, если мы поделили капитал поровну между 0 и 1, то в любом случае останемся «при своих». Отсюда ясно, что отдельно предусматривать возможность вообще не ставить часть капитала не нужно.)

Наша стратегия в такой игре представляет собой функцию, которая (по начальному отрезку последовательности) говорит, сколько нужно ставить на нуль и сколько на единицу. Технически удобнее говорить о несколько другой функции. А именно, пусть $m(x)$ — капитал, который у нас будет после начального отрезка x при данной стратегии. Такая функция однозначно определяет стратегию: после появления слова x мы ставим $m(x0)/2$ на нуль и $m(x1)/2$ на единицу. При этом

- $m(\Lambda) = 1$ (в начале игры, при пустом слове, наш капитал равен единице);
- $m(x) = (m(x0) + m(x1))/2$ (мы ставим то, что у нас есть к данному моменту).

Функцию m, обладающую этими двумя свойствами, будем называть *мартингалом* относительно равномерной меры на пространстве нулей и единиц. (Впоследствии мы будем рассматривать мартингалы относительно других мер на Ω. В теории вероятностей рассматриваются и более общие виды мартингалов, но для наших целей большая общность не потребуется.) Таким образом, вместо стратегий мы будем говорить о мартингалах, им соответствующим.

Пусть ν — произвольная мера на пространстве нулей и единиц. Легко проверить, что отношение $\nu(\Omega_x)/\mu(\Omega_x)$ (где μ — равномерная мера на Ω, а Ω_x есть множество всех продолжений слова x) есть мартингал, и что всякий мартингал соответствует некоторой мере ν.

$\boxed{\textbf{195}}$ Убедитесь в этом.

Имеет место следующий почти очевидный факт (называемый иногда неравенством Колмогорова):

Theorem 150. [martingale-inequality] *Пусть фиксирован некоторый мартингал m и число k. Рассмотрим те слова, на которых значение мартингала не меньше k, и множество всех последовательностей, у которых есть начальный отрезок с таким свойством. Тогда (равномерная) мера этого множества не превосходит $1/k$.*

$\triangleleft$ Рассмотрим стратегию, соответствующую мартингалу m, и решим, что как только наш капитал достигнет (или превысит) k, мы заканчиваем игру и уходим из казино. Ясно, что в силу «честности» игры средний выигрыш при любой стратегии не больше 1, поэтому доля тех случаев, когда он больше или равен k, не превосходит $1/k$.

Формально это проще сказать на языке мер. Пусть $m(x)$ есть отношение $\nu(\Omega_x)/\mu(\Omega_x)$ для некоторой меры ν (а μ — равномерная мера). Мы рассматриваем те x, у которых ν-мера конуса Ω_x превосходит μ-меру этого же конуса в k или более раз. Из них можно оставить только минимальные (не имеющие начал с тем же свойством), которым соответствуют непересекающиеся конусы. Суммарная μ-мера этих конусов по крайней мере в k раз меньше их суммарной ν-меры, которая не превосходит единицы. $\triangleright$

Верно и обратное утверждение: если есть множество малой меры, то можно построить стратегию (мартингал), которая много выигрывает у любой последовательности из этого множества.

Theorem 151. [winning-martingale] *Пусть имеется открытое подмножество* $U \subset \Omega$ *меры* $\varepsilon > 0$*. Тогда существует мартингал* m *с таким свойством: у каждой последовательности* $\omega \in U$ *существует начальный отрезок, на котором* m *не меньше* $1/\varepsilon$*.*

$\lhd$ Рассмотрим меру ν, для которой $\nu(X) = \mu(X \cap U)/\varepsilon$. (Вне множества U — нуль, внутри U — в $(1/\varepsilon)$ раз больше равномерной.) Отношение $m(x) = \nu(\Omega_x)/\mu(\Omega_x)$ и будет искомым мартингалом. В самом деле, если $\omega \in U$, то существует начало x последовательности ω, при котором $\Omega_x \subset U$ и $m(x) = 1/\varepsilon$. $\rhd$

Эту теорему можно объяснить следующим образом. Пусть сотрудники казино — жулики и торгуют, как теперь говорят, «инсайдерской информацией». А именно, они готовы заранее указать открытое множество U, в котором окажется будущая последовательность бросаний. Какова «рыночная цена» такой информации вместе с правом после этого начать игру с капиталом 1? Ответ: $1/\mu(U)$. Например, зная результаты первых N бросаний (что соответствует множеству меры $1/2^N$), мы можем N раз подряд выигрывать и в итоге получить 2^N. Доказанная теорема говорит, что это верно и для более сложно устроенных множеств U. Например, если известно, что некоторая последовательность невозможна («в нашем казино никогда не бывает N нулей подряд с начала игры»), то и это позволит гарантированно выиграть (правда, совсем немного, получив $2^N/(2^N - 1)$ из начального капитала 1).

Легко понять, какова соответствующая стратегия игрока (из доказательства теоремы). Если гарантированное множество U на первом ходу делится в пропорции $a_0 : a_1$ (между последовательностями, начинающимися на нуль и на единицу), то мы должны ставить наш капитал в этой самой пропорции. (Например, если все элементы U начинаются на 0, то весь капитал нужно ставить на нуль.) Тогда отношение

$$\frac{\text{текущий капитал}}{\text{доля } U \text{ среди продолжений текущей позиции}}$$

не меняется в процессе игры. В начале числитель равен 1, а знаменатель равен ε, а при попадании в открытое множество U (которое обязательно произойдёт, если только нам не продали ложную информацию) знаменатель равен 1, а потому числитель равен $1/\varepsilon$.

Аналогичные утверждения о связи мартингалов и мер можно сделать и в «предельном» случае. Будем говорить, что мартингал m *выигрывает* на последовательности ω, если значения m на начальных отрезках последовательности ω не ограничены.

Theorem 152. [martingale-null]

(а) *Пусть* m *— произвольный мартингал. Тогда множество тех последовательностей* $\omega \in \Omega$*, на которых он выигрывает, имеет меру нуль.*

(б) *Пусть* X *— произвольное множество меры нуль. Тогда существует мартингал* m*, который выигрывает на всех последовательностях из* X*.*

$\lhd$ (а) Множество U_k тех последовательностей, на начальных отрезках которых мартингал достигает значений k или больше, имеет меру не более $1/k$ (и является открытым); все последовательности, на которых m выигрывает, принадлежат этому множеству.

(б) Для каждого k рассмотрим открытое множество U_k, имеющее меру не больше $1/k$, содержащее X, и соответствующий мартингал m_k, который гарантирует выигрыш не меньше k на всех элементах U_k (и тем самым на всех элементах X). Теперь из всех этих мартингалов нужно собрать один. Заметим, что взвешенная сумма мартингалов есть мартингал (мы можем разделить наш капитал на конечное или счётное число частей, и с каждой частью играть отдельно, применяя свой мартингал). Будем использовать мартингалы m_{4^n} с начальным капиталом 2^{-n} параллельно при всех n (разложив начальный капитал 1 в сумму ряда $\sum 2^{-n}$); тогда для последовательностей из U_{4^n} (и потому для всех последовательностей из X) гарантирован выигрыш $4^n \cdot 2^{-n} = 2^n$. Значит, общий выигрыш на любой последовательности из X не ограничен. $\triangleright$

Доказательство этой теоремы сильно напоминает доказательство критерия случайности (теорема 82, с. 136), который можно рассматривать как эффективный вариант только что доказанной теоремы.

По существу мы доказали более сильное утверждение. Будем говорить, что мартингал m *сильно выигрывает* на последовательности ω, если его значения на начальных отрезках последовательности ω стремятся к бесконечности. Построенный в теореме 152 мартингал, как легко видеть, *сильно* выигрывает на всех элементах X. (В самом деле, мартингал из доказательства теоремы 151 равен $1/\varepsilon$ на *всех* достаточно длинных начальных отрезках.)

Заметим, что теорему 85, с. 138 можно рассматривать как конструктивный аналог утверждения о существовании мартингалов, сильно выигрывающих у последовательностей из множества меры нуль. (О различных конструктивизация понятий и результатов, связанных с мартингалами, мы ещё будем много говорить.)

Отметим ещё, что переход от выигрыша к сильному выигрышу можно провести и непосредственно:

Theorem 153. *Для всякого мартингала m существует мартингал m', который сильно выигрывает у всех последовательностей, у которых выигрывает мартингал m.*

$\triangleleft$ Мартингал m' должен действовать как запасливый игрок: достигнув выигрыша 2 (по m-стратегии), он откладывает половину «на чёрный день» (это означает, что эту часть выигрыша игрок ставит поровну на нуль и единицу), а с другой половиной поступает как m (но только все суммы вдвое меньше). Когда будет достигнут выигрыш 4 (что соответствует выигрышу 8 для m), откладываем половину (то есть 2) на чёрный день, а остаток снова пускаем в игру в соответствии с m, и так далее. $\triangleright$

До сих пор мы предполагали, что вероятности выпадения нуля и единицы (декларируемые казино) равны, и потому ставки и на нуль, и на единицу удваиваются. Но это не обязательно. Пусть, например, вероятность появления нуля считается равной $1/3$, а вероятность единицы, соответственно, $2/3$. Тогда ставка на нуль должна утраиваться, а на единицу — увеличиваться всего лишь в полтора раза. Соответственно и определение мартингала изменится: $m(x)$ (капитал после x) должен быть равен сумме его части, поставленной на нуль (то есть $m(x0)/3$) и его части, поставленной на единицу (то есть $2m(x1)/3$):

$$m(x) = \frac{1}{3}m(x0) + \frac{2}{3}m(x1).$$

Это равенство можно прочесть и так: капитал до очередной игры равен математическому ожиданию капитала после этой игры.

Дадим формальное определение. Пусть π — произвольное распределение вероятностей на Ω (декларируемое казино распределение вероятностей для монеты). Соответствующую функцию на словах будем также обозначать буквой π, положив $\pi(x) = \pi(\Omega_x)$.

Функция $m(x)$, определённая на двоичных словах и принимающая неотрицательные действительные значения, называется *мартингалом относительно* π (с единичным начальным капиталом — в дальнейшем это особо не оговаривается), если $m(\Lambda) = 1$ и

$$m(x)\pi(x) = m(x0)\pi(x0) + m(x1)\pi(x1)$$

при всех x. (Это определение соответствует предыдущему: поделив на $\pi(x)$, мы получим в правой части условные вероятности $\pi(x0)/\pi(x)$ и $\pi(x1)/\pi(x)$ появления нуля и единицы после x.)

Видно, что это определение означает, что функция $m(x)\pi(x)$ задаёт меру, так что мартингал относительно π (π-мартингал) — это отношение некоторой другой меры к π.

Далее всё аналогично случаю равномерной меры. А именно:

(1) вероятность того, что π-мартингал достигнет k на начальном отрезке последовательности ω, распределённой по мере π, не превосходит $1/k$;

(2) для всякого открытого множества U существует мартингал, который достигает значения $1/\pi(U)$ на всех последовательностях из U;

(3) множество X имеет π-меру нуль тогда и только тогда, когда существует π-мартингал, который выигрывает (вариант: сильно выигрывает) у любой последовательности из X.

В заключение отметим следующее почти очевидное обстоятельство:

Theorem 154. *Для всякого мартингала существует последовательность, у которой он не выигрывает (и даже не превосходит единицы на всех её начальных отрезках).*

$\triangleleft$ В самом деле, из определения следует, что одно из чисел $m(x0)$ и $m(x1)$ не больше $m(x)$, поэтому любое слово можно продолжить на один бит, не увеличивая мартингал. $\triangleright$

(Если крупье может манипулировать монетой, то он может всегда добиться, чтобы клиент ничего не выиграл.)

[Теорема Дуба: для почти всех последовательностей мартингал име- ет предел на начальных отрезков. Конструктивный вариант: это так для случайных по Мартин-Лёфу последовательностей. Определение условной вероятности.]

9.6. Отступление: мартингалы в теории вероятностей

`[misesmart1]`

Теорему 152 можно интерпретировать так:

(а) чтобы доказать, что какое-то множество нулевое, достаточно построить мартингал, который выигрывает на всех его элементах;

(б) этот метод можно применить к любому нулевому множеству (найдя соответствующий мартингал).

Эта интерпретация имеет глубокий смысл. Во-первых, с чисто технической точки зрения мы получаем удобный способ доказательства различных теорем теории вероятностей. Чтобы доказать, что некоторое множество имеет меру нуль, достаточно предъявить мартингал, который выигрывает на всех последовательностях интересующего нас множества.

В качестве иллюстрации изложим в таком стиле доказательство усиленного закона больших чисел. Рассмотрим бернуллиево распределение вероятностей B_p, в котором испытания независимы и вероятность появления единицы равна p.

Для данного $q > p$ докажем, что вероятность события «частота единиц в последовательности бесконечно много раз превосходит q» равна нулю. (Для $q < p$ и для события «частота меньше q» рассуждение симметрично.) Для этого рассмотрим мартингал, равный отношению мер B_q/B_p. Другими словами, рассмотрим мартингал, который на последовательности z длины n с частотой единиц r (из nr единиц и $n(1-r)$ нулей) принимает значение

$$\frac{q^{nr}(1-q)^{n(1-r)}}{p^{nr}(1-p)^{n(1-r)}},$$

а его логарифм равен

$$n[(r\log q + (1-r)\log(1-q)) - (r\log p + (1-r)\log(1-p))].$$

Поскольку $q > p$, то это выражение является возрастающей (линейной) функцией от r: коэффициент при r равен $\log[q/p] + \log[(1-p)/(1-q)]$, и оба слагаемых положительны. Значит, при $r > q$ (интересующий нас случай) замена r на q лишь уменьшит это выражение и логарифм мартингала не меньше

$$n[(q\log q + (1-q)\log(1-q)) - (q\log p + (1-q)\log(1-p)].$$

Согласно неравенству Гиббса (с. 185) коэффициент в квадратных скобках, то есть расстояние Кульбака – Лейблера между распределениями $(q, 1-q)$ и $(p, 1-p)$, положителен. Таким образом, построенный мартингал неограничен на интересующих нас последовательностях (где частота бесконечно много раз превосходит q).

Это доказательство усиленного закона больших чисел не совсем соответствует описанной схеме, так как для разных q получаются свои мартингалы. Каждый из них используется для оценки меры своего множества, а лишь затем мы замечаем, что объединение счётного числа счётных множеств нулевое.

Можно действительно и в другом порядке: взять счётное семейство q_i, соответствующие мартингалы, а затем сложить эти мартингалы с положительными весами. Если хотя бы один из мартингалов неограничен, то и сумма будет неограниченной.

Примерно то же самое мы уже делали в разделе 3.2 (задача 47, с. 59), только говорили о конечных последовательностях и не употребляли слова «мартингал», ограничиваясь упоминаниями двух мер. (По существу те же рассуждения встретятся и в дальнейшем, в разделе 9.13.)

196 Постройте мартингал (относительно равномерной меры), который выигрывает на всех последовательностях, у которых все начальные отрезки содержат не меньше нулей, чем единицы. [Указание. Рассмотрите смесь мартингалов: μ_i выигрывает, если нижний предел превышения нулей над единицами равен i; если мы заранее знаем, что с некоторого места эта разница не меньше i и бесконечно много раз она равна i, то можно предсказывать нуль, когда разница достигла i, и ошибиться только конечное число раз. Отдельно надо построить мартингал на случай, когда превышение стремится к бесконечности.]

Во-вторых, можно смотреть на мартингалы с более философской точки зрения. Что значит доказать некоторую теорему теории вероятностей с помощью мартингала? Это значит предъявить некоторое свойство последовательностей L (скажем, закон больших чисел) и некоторый мартингал m и доказать, что *для любой* последовательности ω верно (хотя бы) одно из двух:

- последовательность ω обладает свойством L;
- мартингал m выигрывает на ω.

Двигаясь в этом направлении, можно предложить следующий «рыночный» (или «игровой») подход к понятию случайности и сказать, что

> *случайность последовательности нулей и единиц — это не свойство последовательности как таковой, а тип гарантии, с которой она продаётся.*

Что это значит? Представим себе продавца, который за рубль продаёт случайную последовательность на карточке. Покупатель может стирать краску, открывая бит за битом. Продавец даёт гарантию, что мартингал покупателя (копию которого покупатель отдаёт продавцу в запечатанном конверте в момент покупки) много не выиграет. Точнее, если этот мартингал в какой-то момент примет значение k на уже открытых битах, то покупатель может вернуть неиспользованные биты, стребовав с продавца k рублей неустойки. (Здесь существенно, что дальнейшие биты не известны покупателю, иначе он мог бы сжульничать: подглядеть вперёд и отказаться от использования невыгодных его мартингалу битов.)

Такой договор (как, наверно, сказали бы финансисты) позволяет «хеджировать» риски покупателя: если покупатель заправит купленную последовательность в свою машину, и машина потерпит ущерб из-за того, что последовательность не удовлетворяет закону больших чисел, то продавец вынужден будет возместить ущерб согласно мартингалу, предусмотрительно заявленному покупателем при покупке. (Конечно, на «практике» речь скорее будет идти о конечных последовательностях, но сама схема остаётся без изменений.)

Заметим, что при подписании гарантийного договора стороны оговаривают меру (поскольку от неё зависит класс мартингалов). Таким образом, при таком подходе мера из чего-то существующего в природе превращается в тип договора, заключаемого покупателем и продавцом! (Но, конечно, продавец должен учитывать даваемую гарантию при производстве последовательностей!)

Теорема 152, если следовать этой метафоре, показывает, что любой закон теории вероятности можно (более или менее естественным образом) уложить в эту схему.

Подробно такой подход к теории вероятностей обсуждается в книге [78].

9.7. Перечислимые мартингалы

[misesmart1a]

Доказанные нами результаты о мартингалах имеют естественные эффективные аналоги. Нулевые множества, как мы видели, соответветствуют мартингалам. Поэтому можно ожидать, что эффективно нулевые множества соответствуют некоторому классу «эффективных» мартингалов, и так оно и есть.

Мы будем рассматривать перечислимые снизу мартингалы и будем разрешать начальному капиталу (значению мартингала на пустом слове) быть произвольным числом, не превосходящим единицы. (Определение перечислимости снизу было дано в разделе 4.1; функция m перечислима снизу, если множество пар $\langle r, x\rangle$, для которых рациональное число r меньше $m(x)$, перечислимо.)

$\boxed{\textbf{197}}$ Покажите, что перечислимый снизу мартингал m, для которого $m(\Lambda) = 1$, вычислим.

Вычислимых мартингалов нам недостаточно, поэтому мы и отказались от требования $m(\Lambda) = 1$.

Теперь можно сформулировать эффективный вариант теоремы 152:

Theorem 155. [martingale-null-effective]

(а) *Пусть m — произвольный перечислимый снизу мартингал. Тогда множество тех последовательностей $\omega \in \Omega$, на которых он выигрывает, является эффективно нулевым.*

(б) *Пусть X — произвольное эффективно нулевое множество. Тогда существует перечислимый снизу мартингал m, который выигрывает на всех последовательностях из X.*

$\lhd$ Если мартингал перечислим снизу, то множество тех последовательностей, на которых он больше (целого) k, является эффективно открытым и имеет меру не больше $1/k$. (При этом мы имеем в виду именно меру множества, а не сумму мер возникающих в процессе перечисления интервалов, поскольку интервалы могут пересекаться: сначала появляется меньший интервал, а потом содержащий его больший. Но это не страшно, так как можно представить разность большего и меньшего как объединение конечного числа интервалов и взять их.)

Наоборот, если множество эффективно открыто и имеет меру меньше $1/k$, то строимый по нему в доказательстве теоремы 151 мартингал будет перечислим снизу (появление нового интервала в множестве увеличивает мартингал). Надо только делить не на меру множества (она не обязательно вычислима, и помещение в знаменатель перечислимого снизу числа может сделать частное неперечислимым снизу), а на её верхнюю оценку $1/k$ (то есть умножать на k). При этом значение мартингала в корне может оказаться меньше единицы, но теперь это разрешено.

Остаётся (как в доказательстве теоремы 152) сложить постронные для разных чисел k мартингалы с надлежащими (вычислимыми) коэффициентами; это не нарушает перечислимость снизу. $\rhd$

Утверждение этой теоремы можно усилить, причём даже в обе стороны. Во-первых, как мы уже говорили, построенный мартингал будет *сильно* выигрывать на всех последовательностях из множества X. (Отметим в скобках, что для перечислимых снизу мартингалов идея «откладывания части выигрыша на чёрный день» так просто не проходит.)

Во вторых, можно рассматривать не только перечислимые снизу мартингалы, но и перечислимые снизу *супермартингалы*, или, как иногда говорят, *полумартингалы*. Они соответствуют играм, в которых казино может отбирать часть выигрыша игрока:

$$m(x) \geqslant (m(x0) + m(x1))/2$$

(для равномерной меры) или

$$m(x)\pi(x) \geqslant m(x0)\pi(x0) + m(x1)\pi(x1)$$

(для произвольной меры). Поскольку это лишь ухудшает положение игрока, верхняя оценка вероятности выигрыша остаётся в силе.

Из определения ясно, что супермартингалы — это отношения *полумер* к мере π, которая предполагается вычислимой. Поэтому, как мы знаем, среди перечислимых снизу полумартингалов существует максимальный

$$m(x) = a(x)/\pi(x),$$

где a — максимальная перечислимая полумера на дереве (см. раздел 5.2), а $\pi(x)$ — вероятность множества Ω_x по мере π. Максимальность понимается с точностью до мультипликативной константы.

Отсюда получается новое доказательство теоремы Левина – Шнорра для априорной вероятности (теорема 83, с. 138): последовательность ω случайна по вычислимой мере π тогда и только тогда, когда на её начальных отрезках отношение $a(x)/\pi(x)$ ограничено.

[Кажется, это было ещё в Звонкине и Левине? Хорошо бы проверить!]

9.8. Вычислимые мартингалы

[misesmart2]

С точки зрения стратегий рассматривать перечислимые мартингалы странно: величина ставки в этом случае выражается через отношения перечислимых снизу чисел, и этому сложно придать естественный игровой смысл.

Поэтому интересно посмотреть, что получится, если ограничиться лишь вычислимыми мартингалами. Пусть фиксирована некоторая вычислимая мера π. Для простоты будем предполагать, что все значения $\pi(x) = \pi(\Omega_x)$ положительны (это нужно, поскольку эти значения стоят в знаменателях).

Будем называть последовательность ω *случайной по мере* π *относительно вычислимых мартингалов*, если никакой вычислимый π-мартингал на ней не выигрывает — другими словами, если любой вычислимый π-мартингал ограничен на её начальных отрезках.

Первые два утверждения следующей теоремы относятся к общему случаю случайности по мере π; два других — к случайности по бернуллиевой мере (все испытания независимы и имеют равную вычислимую вероятность успеха p).

Theorem 156. [martingale-random]

(а) *Всякая случайная в смысле Мартин-Лёфа последовательность случайна относительно вычислимых мартингалов.*

(б) *Существует случайная относительно вычислимых мартингалов последовательность, начальные отрезки которой имеют логарифмическую сложность. (Отсюда следует, что предыдущее утверждение теоремы нельзя обратить.)*

(в) *Всякая случайная относительно вычислимых мартингалов последовательность случайна по Мизесу – Чёрчу.*

(г) *Не всякая случайная по Мизесу – Чёрчу последовательность случайна относительно вычислимых мартингалов.*

◁ (а) Случайность по Мартин-Лёфу гарантирует, что даже и перечислимые снизу мартингалы не выигрывают (теорема 155).

(б) Мы уже отмечали, что для любого мартингала можно найти последовательность, на которой он ограничен (достаточно из двух исходов игры выбирать тот, где капитал игрока не увеличивается).

Если мартингал вычислим, то можно найти вычислимую последовательность, на которой он ограничен. Это немного сложнее: мы не можем сравнить значения мартингала на двух продолжениях $x0$ и $x1$ слова x, так как знаем эти значения лишь с некоторой точностью. Но это и не нужно, достаточно на n-м шаге выбрать такое продолжение, где мартингал растёт менее чем на $1/2^n$ (а это можно сделать, находя всё более точные приближения и ожидая приближения, которое это гарантирует).

(Кстати, отсюда немедленно следует, что среди вычислимых мартингалов не существует максимального. Именно поэтому переход к перечислимым снизу мартингалам и супермартингалам существен.)

Следующий шаг: пусть даны два вычислимых мартингала; как найти вычислимую последовательность, на которой они оба ограничены? Для этого достаточно взять взвешенную сумму этих мартингалов, в которую они оба входят с положительным коэффициентом (например, полусумму). Получится снова вычислимый мартингал. Если хотя бы один из двух мартингалов не ограничен на некоторой последовательности ω, то и их полусумма не ограничена на ω, поэтому можно применить предыдущее утверждение.

То же рассуждение годится, если есть вычислимая последовательность вычислимых мартингалов (то есть последовательность программ, эти мартингалы задающих). Тогда их можно сложить, взяв i-й мартингал с весом 2^{-i}.

Все вычислимые мартингалы нельзя расположить в вычислимую последовательность (это, кстати, следует из только что нами доказанного: ведь тогда можно было бы найти вычислимую последовательность, на которой все вычислимые мартингалы ограничены!). Поэтому построение последовательности ω, на которой все вычислимые мартингалы ограничены, не может быть алгоритмическим. Дополнительная информация, которая нам нужна — это сведения о том, какие из программ задают мартингалы, а какие нет, то есть по биту на программу. Чтобы получить последовательность с логарифмической сложностью, мы должны эти биты использовать очень понемногу, включая в рассмотрение i-ю программу, когда последовательность будет уже достаточно длинной (экспоненциальной от i длины или ещё длиннее).

Опишем этот процесс подробнее. В каждый момент мы имеем некоторое слово x, а также некоторую линейную комбинацию $m_1(x)+\varepsilon_2 m_2(x)+\ldots+\varepsilon_k m_k(x)$, где m_i — мартингал, задаваемый i-й программой (или его заменитель — скажем, всюду равный 1 мартингал, — если дополнительная информация об i-й программе говорит, что она не задаёт мартингал). При этом мы следим, чтобы это выражение (для текущего слова x) было строго меньше 2. (Вначале, когда у нас только m_1 и слово x пусто, это выражение равно 1.)

Как мы видели, слово x всегда можно продолжить на один бит так, чтобы это выражение оставалось меньше двух (и такое продолжение можно алгоритмически найти, если мы знаем программы для мартингалов). Также время от времени мы будем добавлять новый член $\varepsilon_k m_k(x)$ к сумме, подбирая ε_k настолько малым, чтобы сумма осталась меньше 2 (чем ближе текущая сумма к 2 и чем больше значение $m_k(x)$ для текущего слова x, тем меньше берётся ε_k).

На полученной последовательности все m_i будут ограничены, так как каждый из них входит пусть с малым, но с постоянным ненулевым коэффициентом.

Сложность разрешения начального отрезка такой последовательности не превышает числа использованных битов дополнительной информации, и может быть очень мала, если мы добавляем новые мартингалы редко. А обычная колмогоровская сложность отрезков длины n будет $O(\log n)$, что и требовалось.

(в) Вспомним, что усиленный закон больших чисел говорит, что некоторое множество нулевое, и потому существует соответствующий мартингал (см. раздел 9.6, где мы строили мартингалы для каждой границы, а потом их смешивали). Ясно, что получающийся при этом мартингал можно сделать вычислимым (напомним, что p вычислимо).

Более того, если R — некоторое множество слов, а S_R — соответствующее правило выбора, то легко построить мартингал, который выигрывает на любой последовательности ω, для которой подпоследовательность $S_R(\omega)$ несбалансирована. В самом деле, нужно играть лишь с теми членами последовательности, которые отбираются правилом S_R (а в остальных случаях сохранять значение мартингала неизменным, пропуская ход).

При этом, если множество R разрешимо, то мы получаем вычислимый мартингал, который выигрывает на любой последовательности ω, для которой подпоследовательность $S_R(\omega)$ бесконечна, но не сбалансирована.

Поэтому для всякой неслучайной по Мизесу – Чёрчу последовательности существует вычислимый мартингал, который на ней не ограничен.

(г) Рассмотрим (для случая равномерной меры) случайную по Мизесу – Чёрчу последовательность, у которой любой начальный отрезок содержит не меньше нулей, чем единиц (теорема 147). Пусть p_n — вероятность того, что все начальные отрезки длины не более n содержат не меньше нулей, чем единиц. Как мы уже обсуждали (задачи 193, 194), последовательность p_n вычислима, убывает и стремится к нулю. Для каждого n можно эффективно указать мартингал M_n, который выигрывает $1/p_n$ у любой последовательности, начальные отрезки (вплоть до длины n) которой содержат не меньше нулей, чем единиц. Взяв теперь взвешенную сумму некоторых M_n (выбранных так, чтобы выигрыши росли быстрее, чем убывали коэффициенты), получим вычислимый мартингал, который не ограничен на любой последовательности, у которой во всех начальных отрезках не меньше нулей, чем единиц.

Это же рассуждение можно после некоторых модификаций использовать для произвольной бернуллиевой меры (одновременно модифицировав конструкцию теоремы 147). Можно также применить конструкцию с двумя мерами (см. ниже раздел 9.13). $\triangleright$

Заметим, что из утверждений (б) и (в) вытекает, что существуют случайные по Мизесу – Чёрчу последовательности с логарифмической сложностью начальных отрезков, тем самым мы получили другое доказательство теоремы 148 (следуя работе [48]).

В заключение раздела отметим, что всё сказанное можно перевести на язык стратегий — алгоритмов, которые смотрят на уже происшедшие испытания и определяют, в какой пропорции нужно делать ставку на следующем шаге. При этом можно ограничиться лишь стратегиями с рациональными значениями, которые вычислимы как функции $\Xi \to \mathbb{Q}$.

198 Докажите это. [Указание. Приближая стратегию с достаточной точностью, можно гарантировать, что от замены точной стратегии на приближённую выигрыш изменится не более чем вдвое (скажем).]

9.9. Мартингалы и случайность по Шнорру

[misesmart3]

Понятие случайности относительно вычислимых мартингалов связно с понятием случайности по Шнорру (см. раздел 3.4). Оба этих понятия были введены и изучены в книге Шнорра [66]. Там же доказана и следующая теорема.

Theorem 157. [schnorr-martingale-randomness] *Пусть дана некоторая вычислимая мера π, при которой все интервалы Ω_x имеют положительную меру. Последовательность ω не случайна по Шнорру относительно π тогда и только тогда, когда существует вычислимый мартингал m относительно меры π и вычислимая неубывающая неограниченная функция $g\colon \mathbb{N} \to \mathbb{N}$ с таким свойством:*

$$m(\omega_0\omega_1 \ldots \omega_{n-1}) \geqslant g(n)$$

для бесконечно многих n.

Эта теорема говорит, что неслучайность по Шнорру влечёт не только существование вычислимого мартингала, выигрывающего на последовательности, но и некоторую нижнюю оценку скорости выигрыша (справедливую для бесконечно многих начальных отрезков).

$\lhd$ Пусть последовательность ω не случайна по Шнорру. Тогда, как мы видели в разделе 3.4 (задача 65, с. 71), существует последовательность слов $x_0, x_1, x_2 \ldots$, для которых ряд $\sum \pi(x_i)$ вычислимо сходится, причём бесконечно многие среди x_i являются началами ω.

Сгруппируем члены ряда

$$\pi(x_0) + \pi(x_1) + \pi(x_2) + \ldots + \pi(x_i) + \ldots$$

так, чтобы сумма членов k-й группы была не больше 4^{-k}. Поскольку ряд вычислимо сходится, это можно сделать алгоритмически. Можно считать, что слова разных групп разделены по длинам: есть вычислимая последовательность чисел $n_0 < n_1 < n_2 < \ldots$, и слова k-группы имеют длину в интервале $[n_k, n_{k+1})$. В самом деле, любое слово можно заменить на набор всех его продолжений большей длины, и мы можем по очереди обрабатывать слова разных групп, используя длины больше длин предыдущей группы. Суммы по группам от этого не изменятся.

Рассмотрим отдельно слова k-группы. Они покрывают множество меры меньше 4^{-k} и можно построить мартингал m_k, который достигает на них значения 4^k. Теперь смешаем их и получим мартингал $m = \sum 2^{-k} m_k$. Этот мартингал достигает значения 2^k на словах k-ой группы. Остаётся положить $g(n) = 2^k$ для всех n в интервале $[n_k, n_{k+1})$ и заметить, что в бесконечно многих группах встречаются начала последовательности ω.

Напротив, пусть есть вычислимый мартингал m и вычислимая функция g. Пусть про последовательность ω известно, что для бесконечно многих n значение $m(\omega_0 \ldots \omega_{n-1})$ не меньше $g(n)$. В доказательстве теоремы 155 мы покрывали ω перечислимым семейством интервалов Ω_x, рассматривая слова x с $m(x) > k$, и сумма соответствующих мер была не больше $1/k$. Теперь мы знаем дополнительно, что

- мартингал m не только перечислим снизу, но и вычислим;

• мартингал m достигает значения n или больше на начальном отрезке (последовательности ω) длины не более $g(n+1)$.

Как это можно использовать? Во-первых, можно заменить мартингал его вычислимым рациональным приближением. Пусть, скажем, m' есть такое приближение сверху с ошибкой меньше 1. Тогда слова x, для которых $m'(x) > 2^k$, образуют разрешимое множество с суммарной мерой не больше $1/(2^k-1)$. Во-вторых, можно отбросить длинные слова (для которых значение g на их длине больше, скажем, 2^{k+2}), и от этого разрешимого множества остается конечное, которое можно построить алгоритмически по k. Правда, это конечное множество уже не обязано покрывать ω при всех k. Но такое происходит при бесконечно многих k. В самом деле, при бесконечно многих n случается так, что $m(\omega_0 \ldots \omega_{n-1}) \geqslant g(n)$. Взяв одно из таких n, выберем k, при котором $2^k < g(n) < 2^{k+2}$. Тогда для $x = \omega_0 \ldots \omega_{n-1}$ имеем $m(x) > 2^k$ и $m'(x) > 2^k$, так что слово x попадёт в разрешимое множество. С другой стороны, $g(n) < 2^{k+2}$ гарантирует, что слово x не будет выброшено и попадёт в построенное конечное множество.

Теперь, соединив все слова из построенных конечных множеств при всех k, мы получим вычислимый вычислимо сходящийся ряд, покрывающий последовательность ω бесконечно много раз, и потому она не случайна по Шнорру (задача 65, с. 71). $\triangleright$

[Не забыть про законы невозрастания информации — В]

9.10. Мартингалы и эффективная размерность

[misesmart4]

В предыдущих разделах мы видели, как можно перевести понятия нулевого и эффективно нулевого (а также нулевого по Шнорру) множества на язык мартингалов. Аналогичный перевод возможен и для понятия размерности по Хаусдорфу. В одной фразе этот перевод можно описать так: чем меньше размерность множества, тем быстрее может расти мартингал на его элементах. Начнём с классического варианта (без алгоритмов):

Theorem 158. [hausdorf-martingale] *Множество $A \subset \Omega$ является α-нулевым тогда и только тогда, когда существует мартингал m с таким свойством: для любого $\omega \in A$ отношение*

$$\frac{m(x)}{2^{(1-\alpha)l(x)}}$$

не ограничено на начальных отрезках ω.

(При $\alpha = 1$ получается теорема 152.)

Немного более наглядная формулировка теоремы 158 такова: пусть игрок облагается налогом, в результате которого его капитал на каждом шаге (после игры и перед следующей) уменьшается в $2^{1-\alpha}$ раз, то есть умножается на $2^{\alpha-1}$. Тогда капитал игрока после x будет не мартингалом, а функцией m, для которой

$$2^{\alpha-1}m(x) = \frac{m(x0)}{2} + \frac{m(x1)}{2},$$

то есть

$$2^{\alpha}m(x) = m(x0) + m(x1)$$

Такие функции, (следуя работе [42]) называют "α-gales", что мы по-русски передаём как *α-мартингалы*. Они тоже соответствуют мерам: α-мартингал есть функция вида

$$p(x)2^{\alpha l(x)},$$

где $p(x)$ — мера интервала Ω_x относительно некоторого распределения. Аналогично можно определять и *α-супермартингалы*, где допускаются дополнительные потери игрока (помимо налога):

$$2^\alpha m(x) \geqslant m(x0) + m(x1)$$

Теперь можно пересказать утверждение теоремы 158 так: множество тех последовательностей, на начальных отрезках которых данный α-мартингал неограничен, является α-нулевым, и всякое α-нулевое множество содержится в множестве такого вида.

$\triangleleft$ Доказательство по существу повторяет рассуждение из теоремы 152. Пусть m — произвольный α-мартингал. Докажем, что множество тех последовательностей, для которых он неограничен (на начальных отрезках), имеет α-меру нуль. Для этого достаточно показать, что слова, где m достигает значения k впервые, имеют сумму α-степеней мер не больше $1/k$. Если записать мартингал $m(x)$ как $p(x)2^{\alpha l(x)}$, то для выбранных слов x имеем $p(x) \geqslant k2^{-\alpha l(x)}$. Все выбранные слова несравнимы (одно не является началом другого), поэтому сумма p-мер не больше единицы, поэтому сумма величин $2^{-\alpha l(x)}$ по всем выбранным словам x, то есть сумма α-мер соответствующих интервалов, не превосходит $1/k$.

Обратно, пусть дано некоторое множество α-меры нуль. Нам надо построить α-мартингал, который неограничен на всех его элементах. Возьмём покрытие интервалами с суммой α-мер меньше $1/k$, и построим α-мартингал m_k, который достигает значения k на этих элементах. (Затем можно сложить m_{4^k} с коэффициентами 2^{-k}, сумма α-мартингалов есть также α-мартингал.) Как построить m_k? Для данного слова x можно построить α-мартингал, который равен 1 на x и 0 на всех словах той же длины $l(x)$, продолжив его на более короткие слова (однозначно) и на более длинные (любым способом). На пустом слове значение этого α-мартингала будет $2^{-\alpha l(x)}$, то есть как раз α-мера соответствующего интервала. Значит, сумма этих α-мартингалов для всех x будет не больше $1/k$, и умножив её на k, получим искомый α-мартингал. $\triangleright$

Теперь попытаемся сформулировать эффективный вариант теоремы. Пусть α — вычислимое действительное число из $(0,1]$. Естественно определяются понятие *перечислимого снизу α-мартингала* и *α-супермартингала*. Как и в случае $\alpha = 1$, значение перечислимого снизу мартингала на пустом слове не обязательно равно единице — это произвольное (перечислимое снизу) число, не превосходящее единицы.

Можно предположить, что α-мартингалам (и супермартингалам) соответствуют введённые в разделе 5.8 эффективно α-нулевые множества и что доказательство этого факта повторяет только что приведённое доказательство теоремы 158. В одну сторону это действительно так:

Theorem 159. [hausdorf-martingale-effective-a] *Пусть $\alpha \in (0,1]$ — вычислимое действительное число и множество $A \subset \Omega$ является эффективно α-нулевым. Тогда существует перечислимый снизу α-мартингал, который неограничен на начальных отрезках любой последовательности $\omega \in A$.*

$\lhd$ В самом деле, приведённая выше конструкция даёт перечислимые снизу α-мартингалы m_k, и смешивание сохраняет перечислимость снизу. $\rhd$

В другую сторону мы встречаемся с неожиданной трудностью. Пусть m — перечислимый снизу α-мартингал. Рассмотрим множество тех слов, на которых $m(x) > k$. Это множество перечислимо. Кроме того, сумма α-мер его минимальных элементов не превосходит $1/k$. Но множество минимальных элементов может быть неперечислимым, а если оставить все элементы (не только минимальные), не получается оценка для суммы мер. Поэтому доказать, что множество тех последовательностей, на которых перечислимый снизу α-мартингал неограничен, является *эффективно* α-нулевым, этим способом не удаётся.

[А верно ли это вообще?]

Однако можно доказать следующее ослабленное утверждение

Theorem 160. [hausdorf-martingale-effective-b] *Пусть m — перечислимый снизу α — мартингал, а $\beta > \alpha$. Тогда множество тех последовательностей, на (начальных отрезках) которых m неограничен, является β-нулевым.*

(Напомним, что в данных нами определениях предполагается, что α и β вычислимы.)

$\lhd$ Пусть k — положительное натуральное число. Рассмотрим те слова x, для которых $m(x) > k$, и соответствующие им интервалы. Получится покрытие интересующего нас множества. Найдём сумму β-мер этих интервалов. Как мы видели, если выбрать несравнимое множество из этих интервалов, то сумма α-мер будет не больше $1/k$. В частности, для каждой длины N сумма α-мер интервалов этой длины (из покрытия) не превосходит $1/k$. А сумма β-длин будет меньше в $2^{N(\beta-\alpha)}$ раз. Поэтому суммирование по длинам приведёт лишь к умножению на сумму бесконечно убывающей геометрической прогрессии, так что можно с увеличением k получить сколь угодно малую сумму β-мер. $\rhd$

Из двух последних теорем вытекает такое следствие:

Theorem 161. [effective-martingale-dimension] *Эффективная хаусдорфова размерность произвольного множества $A \subset \Omega$ равна точной нижней грани тех α, при которых существует α-мартингал, неограниченный на всех элементах множества A.*

В этом утверждении можно заменить мартингалы на супермартингалы.

Заметим, что отсюда легко вывести утверждение теоремы 97. В самом деле, α-супермартингалы получаются из полумер умножением на $2^{\alpha l(x)}$. Поэтому среди перечислимых снизу α-супермартингалов есть максимальный (получающийся из максимальной перечислимой снизу полумеры a), и в последней теореме можно рассматривать только его. Значит, эффективная размерность множества $\{\omega\}$ равна точной нижней грани тех α, при которых

$$a(\omega_0\omega_1\ldots\omega_{n-1})2^{\alpha n}$$

неограничено. Логарифм этого выражения есть

$$\alpha n - \mathit{KA}\,(\omega_0\omega_1\ldots\omega_{n-1}),$$

и потому точная нижняя грань таких α есть

$$\liminf \frac{\mathit{KA}\,(\omega_0\ldots\omega_{n-1})}{n}.$$

(В теореме 97 использовалась простая колмогоровская сложность, но с интересующей нас точностью это не играет роли.)

9.11. Частичные правила выбора

[misesdaley]

Вернёмся теперь от мартингалов к правилам выбора и подпоследовательностям. Определяя случайность по Мизесу – Чёрчу, мы рассматривали правила выбора, соответствующие *разрешимым* множествам R. Такое правило никогда не «зависает», решая вопрос о том, нужно или нет сделать ставку на очередной член последовательности.

Можно рассмотреть и более широкий класс правил выбора. А именно, пусть r — частичная вычислимая функция на двоичных словах, принимающая значения 0 и 1. Имея начальный отрезок $\omega_0 \ldots \omega_{n-1}$ и решая вопрос о том, включать ли следующий член ω_n в подпоследовательность, мы вычисляем значение $r(\omega_0 \ldots \omega_{n-1})$. Если оно равно единице, то ω_n включается, если оно равно нулю — то нет, а если не определено — процесс выбора вообще на этом обрывается и получается конечная подпоследовательность. Формально говоря, это эквивалентно рассмотрению множества R, определённого следующим образом: слово x принадлежит R, если функция r определена на x и на всех его началах, причём $r(x) = 1$. (Это множество может быть неразрешимым для вычислимых частичных r.) Таким образом, каждой вычислимой функции r описанного вида соответствует правило выбора, которое мы будем обозначать S_r.

Получаем более широкий класс правил выбора, чем допустимые по Чёрчу. Этот класс рассматривал Дэли [12] и мы будем называть такие правила *допустимыми по Чёрчу – Дэли*. Последовательность называется *случайной по Мизесу – Чёрчу – Дэли*, если любое допустимое по Чёрчу – Дэли правило выбора даёт сбалансированную или конечную последовательность.

$\boxed{199}$ Докажите, что применение допустимого по Чёрчу – Дэли правила к случайной по Мизесу – Чёрчу – Дэли последовательности даёт случайную по Мизесу – Чёрчу – Дэли последовательность.

При таком расширении класса допустимых правил класс случайных последовательностей уменьшается. Это следует из теоремы 148 (с. 237) и такого утверждения (доказанного Меркле):

Theorem 162. *Не существует случайной по Мизесу – Чёрчу – Дэли последовательности ω, для которой*

$$KS(\omega_0 \ldots \omega_{n-1}) = O(\log n)$$

$\lhd$ Пусть

$$KS(\omega_0 \ldots \omega_{n-1}) < c \log n$$

для некоторого c и для всех достаточно больших n. Мы хотим показать, что такая последовательность не может быть случайной по Мизесу – Чёрчу – Дэли, то есть построить правило выбора, нарушающее устойчивость частот.

Для начала рассмотрим случай, когда $c < 1$. Множество всех слов, имеющих сложность не больше $c \log n$, представляет собой перечислимое множество из не более чем n^c элементов, и при бóльших n число элементов в этом множестве (обозначим его C_n) не больше $n/10$. Зафиксируем одно такое n. Читая слева направо начальный отрезок длины n, будем пытаться предсказывать следующий бит по предыдущим. Оказывается, что из первых n битов можно гарантированно отгадать 90%. Это делается так. Начнём перечислять C_n и дождёмся появления первого элемента. Этот элемент мы объявляем «текущим кандидатом» и предсказываем те биты, которые в нём стоят, пока не ошибёмся. Такая ошибка означает, что начальный отрезок последовательности ω, хотя и принадлежит C_n, но отличается от найденного нами элемента. Поэтому продолжим перечисление C_n, пока в нём не обнаружится другой элемент, и будем использовать этот элемент для предсказаний — снова до первой ошибки. Ясно, что число ошибок не больше числа различных элементов в C_n, и поэтому как минимум $0{,}9n$ членов мы предскажем правильно.

Будем применять такой метод предсказания для быстро возрастающей последовательности значений $n_0 < n_1 < n_2 \ldots$ и достаточно большого n_0 (начиная с которого сложность меньше $c \log n$). Используя C_{n_i} на участке $[n_{i-1}, n_i]$, мы допускаем не более $0{,}1n_i$ ошибок. Если n_{i-1}/n_i мало (скажем, меньше $0{,}1$), то в целом среди первых n_i предсказаний будет не более $0{,}2n_i$ ошибок.

Осталось заметить, что (как и в теореме 144) алгоритм предсказания соответствует двум правилам выбора: одно выбирает те члены, где мы предсказывали единицу, а второе — где предсказывали нуль, и одна из подпоследовательностей будет сильно несбалансированной.

Что же делать при $c > 1$? Пусть, скажем, $c = 1{,}5$. Тогда оценка для числа элементов в C_n будет $n^{1{,}5}$, что много больше n, так что наше рассуждение не проходит. Но можно сделать так. Разделим слово $\omega_0 \ldots \omega_{n-1}$ на две половины u и v (по $n/2$ битов каждая). Пара $\langle u, v\rangle$ имеет сложность не более $1{,}5 \log n$. Но сложность пары равна сумме $\mathit{KS}(u) + \mathit{KS}(v|u)$ с точностью до членов порядка $O(\log \mathit{KS}(u, v)) = O(\log\log n)$, поэтому либо $\mathit{KS}(u) < 0{,}8 \log n$, либо $\mathit{KS}(v|u) < 0{,}8 \log n$. В каждом случае мы можем применить уже разобранный метод угадывания, поскольку оценка сверху для числа кандидатов, $n^{0{,}8}$, менее одной десятой от числа угадываемых битов ($n/2$). (Заметим, что при предсказании битов второй половины мы уже знаем первую половину, так что можем перечислять слова, у которых условная сложность меньше $0{,}8 \log n$)

Таким образом, один из двух алгоритмов предсказания будет успешным на своей половине. А если алгоритм предсказания на данной половине успешен, то одно из двух правил выбора, пропускающих все члены из другой половины и выбирающее предсказанные нули/единицы в этой, будет успешным.

Тут, однако, возникает существенная трудность. Всё сказанное относилось к одному значению n; далее нам нужно соединить наши алгоритмы предсказания для разных n_i в один. Проблема в том, что если мы пытались предсказывать биты (скажем, в левой половине), считая, что $\mathit{KS}(u) < 0{,}8n$, в то время как на самом деле это было не так, то наш алгоритм не только не будет успешным (это было бы не страшно), но вообще «зависнет» и тем самым никаких предсказаний не даст не только для этого n, но и для всех следующих n из последовательности n_i.

Чтобы преодолеть эту трудность, надо вспомнить, как доказывалась формула для сложности пары (теорема 21, с. 42) и встроить её доказательство в нашу конструкцию. Вот что при этом получится.

Мы по-прежнему будем отдельно предсказывать левую и правую половины (u и v) начального отрезка длины n, но делать это несколько иначе. Предсказывая v слева направо (при известном u), мы перечисляем множество C_n возможных значений начального отрезка (состоящее из всех слов длины n и сложности меньше $1{,}5 \log n$), ожидая появления кандидата, согласованного с u и уже известными битами в v. Когда такой кандидат найден, мы используем его для предсказания, пока не ошибёмся (после чего ищем следующего кандидата). Успешность таких предсказаний (при данном u) зависит от того, сколько существует различных v, при которых $uv \in C_n$. Если их много, то мы можем ошибаться каждый раз (и каждый раз заменять кандидата). А если таких v мало, скажем, не больше $n^{0,8}$, то наши предсказания будут успешными. Важно, что в любом случае этот способ предсказания никогда не «зависнет», если только uv действительно содержится в C_n. (А это можно гарантировать, взяв n_1 достаточно большим.)

Теперь о предсказании левой половины. Здесь мы в качестве кандидатов будем рассматривать такие слова u, для которых существует не менее $n^{0,8}$ значений v, при которых $uv \in C_n$. Эти предсказания будут удачными, если левая половина попадёт в число кандидатов (что обязательно случится, если предсказания в правой половине будут неудачными). Но если левая половина не окажется среди кандидатов, то этот способ предсказания «зависнет» (в некоторый момент мы будем ждать появления кандидата, согласованного с уже известными членами, и никогда не дождёмся).

Что произойдёт при объединении этих алгоритмов для разных n_i? Сначала рассмотрим алгоритмы предсказания правых половин. Эти алгоритмы никогда не зависают (считаем, что n_1 достаточно велико, так что все начальные отрезки длин n_i имеют сложность меньше $1{,}5\log n_i$). Если бесконечное число среди этих алгоритмов успешно (соответствующие множества малы), то соединённый алгоритм будет успешен (доля успешных предсказаний не стремится к $1/2$). Значит, достаточно рассмотреть случай, когда лишь конечное число среди этих алгоритмов успешно. Тогда при всех достаточно больших i число кандидатов v (при данном u) будет больше $n_i^{0,8}$, и потому, отбросив некоторый начальный отрезок, мы можем считать левый алгоритм предсказания всегда успешным. Таким образом, и в этом случае последовательность не будет случайной по Мизесу – Чёрчу – Дэли.

Что же делать, если наша константа c ещё больше? Надо действовать аналогично, только разбивать последовательность на большее число частей. Алгоритм предсказания для самой правой части никогда не зависает. Если он успешен в бесконечном числе случаев, всё доказано. Если с некоторого момента он неуспешен, то с этого момента следующий алгоритм (для предпоследней части) никогда не зависает. Если он успешен в бесконечном числе случаев, всё доказано. Если нет, то рассмотрим предыдущую часть и так далее.

(Аккуратное изложение со всеми деталями см. в статье [48].) $\triangleright$

Таким образом, сложность начальных отрезков случайных по Мизесу – Чёрчу – Дэли последовательностей не может быть логарифмической. Но она может расти лишь немного быстрее логарифма (скажем, быть $O(\log n \log\log n)$.

Theorem 163. [mises-daley-example] *Пусть $f\colon \mathbb{N}\to\mathbb{N}$ — всюду определённая неубывающая вычислимая неограниченная функция. Тогда существует случайная по Мизесу – Чёрчу – Дэли последовательность, у которой сложность начального отрезка длины n не больше $f(n)\log n + O(1)$ (при всех n).*

$\triangleleft$ Вспомним, как строились случайные по Мизесу – Чёрчу последовательности малой сложности (теорема 148). Тогда дополнительная информация состояла в том, какие из правил действительно являются правилами (какие программы всюду определены) — по одному биту на правило. Теперь этого уже мало. Для каждой программы надо знать, в какой момент она перестала быть определённой (то есть первый момент, когда она не определена на текущем начальном отрезке), чтобы с этого момента её заменить на что-нибудь безобидное. Таким образом, если при построении начального отрезка длины n в игру вводятся $f(n)$ программ, то для его описания при известном n достаточно $f(n)\log n$ битов информации (для каждой программы нужно $\log n$ битов, чтобы указать момент её неопределённости).

Заметим, что наличие n в качестве условия не существенно, так как меняет сложность на $O(\log n)$, что соответствует изменению функции f на $O(1)$. $\triangleright$

Аналогичное утверждение можно сформулировать и для мартингалов. Напомним, что мы рассматривали последовательности, случайные относительно вычислимых мартингалов, и при этом в качестве мартингалов (для равномерной меры) можно было рассматривать вычислимые всюду определённые функции с рациональными значениями. Теперь разрешим брать и частичные функции с рациональными значениями, при этом неравенство из определения мартингала должно выполняться в тех случаях, когда все три входящих в него величины определены. Будем говорить, что такой частичный мартингал *выигрывает* на некоторой последовательности, если он определён на всех её начальных отрезках и не ограничен на них. Будем говорить, что последовательность случайна (по равномерной мере) *относительно частичных вычислимых мартингалов*, если не существует вычислимого частичного мартингала, выигрывающего на ней.

Theorem 164. [partial-martingale-merkle] (а) *Всякая последовательность, случайная относительно частичных вычислимых мартингалов, случайна по Мизесу – Чёрчу – Дэли.*

(б) *Пусть $f\colon \mathbb{N}\to\mathbb{N}$ — неубывающая неограниченная вычислимая функция. Тогда существует случайная относительно частичных вычислимых мартингалов последовательность, у которой сложность начального отрезка длины n не больше $f(n)\log n+O(1)$ (при всех n).*

Отметим, что из этих утверждений вытекает предыдущая теорема 163.

$\triangleleft$ (а) Применим ту же самую конструкцию, преобразующую правило выбора в мартингал, что и в теореме 156, пункт (в). Если правило R не всюду определено, то и мартингал будет частичным. Но если правило в применении к некоторой последовательности ω даёт бесконечную подпоследовательность, то соответствующий ему мартингал будет определён на всех начальных отрезках последовательности ω.

(б) И здесь практически без изменений проходит рассуждение из теоремы 156, пункт (б). Для каждого вводимого в игру мартингала нам надо знать, в какой момент он становится неопределённым (чтобы с этого момента заменить его чем-нибудь безобидным, например, его последним значением). Эта информация занимает не более $\log n$ битов для каждого мартингала, использованного при построении начального отрезка длины n, и если к этому моменту ввести в дело не больше $f(n)$ мартингалов, получится как раз требуемая оценка. $\triangleright$

9.12. Немонотонные правила выбора

[miseskl]

До сих пор наши правила выбора были монотонны (сохраняли порядок членов в исходной последовательности). Немонотонные правила были предложены Колмогоровым [23] и независимо Лавлэндом [40,41].

Наглядно идею Колмогорова и Лавлэнда можно объяснить следующим образом. Пусть в казино бросают монету не при нас, а заранее, и результаты бросаний (нули и единицы) записывают на карточках. Эти карточки кладут «лицом вниз», так что мы не видим, что на них написано.

Глядя на последовательность карточек, мы имеем право попросить перевернуть любую из них (не делая ставки), а также сделать ставку на любую ещё не перевёрнутую карточку.

Наша стратегия в такой игре задаётся двумя функциями. Первая функция F отображает двоичные слова в натуральные числа и говорит, какую карточку надо переворачивать следующей (в зависимости от того, что мы увидели на уже перевёрнутых карточках). Мы предполагаем, что значения функции F на любых двух словах, из которых одно является началом другого, различны (никакая карточка не переворачивается дважды).

Вторая функция G определена также на двоичных словах и принимает значения 0 и 1; договоримся, что значение 0 означает, что очередная карточка переворачивается «для информации», а 1 означает, что на неё «делается ставка» (соответствующий член включается в подпоследовательность).

Формально говоря, для любой частичной функции F (с указанным свойством) и для любой частичной функции G мы определяем отображение (правило выбора) $S_{F,G} : \Omega \to \Sigma$ следующим образом: сначала строим конечную или бесконечную последовательность натуральных чисел $n_0, n_1, \ldots$ по формулам

$$n_0 = F(\Lambda),\ n_1 = F(\omega_{n_0}),\ n_2 = F(\omega_{n_0}\omega_{n_1}), \ldots$$

(построение прерывается, как только очередное значение F не определено). Наше условие на функцию F гарантирует, что все номера n_i различны.

Затем отбираются те члены ω_{n_i}, при которых значение G на слове $\omega_{n_0}\omega_{n_1}\ldots\omega_{n_{i-1}}$ определено и равно 1, а значения G на всех началах этого слова определены (и могут быть любыми). Соответствующие ω_{n_i} (в порядке возрастания i) и образуют последовательность $S_{F,G}(\omega)$.

Правила выбора $S_{F,G}$, соответствующие частичным вычислимым функциям F и G, называются *допустимыми по Колмогорову – Лавлэнду*. Последовательность $\omega \in \Omega$ называется *случайной по Мизесу – Колмогорову – Лавлэнду* относительно равномерной меры, если любое допустимое по Колмогорову – Лавлэнду правило выбора даёт сбалансированную (или конечную) последовательность.

Аналогично определяется случайность и относительно бернуллиевой меры (независимые испытания с вероятностью успеха p) при любом p.

Следующее простое, но неожиданное наблюдение сделал Меркле [49]:

Theorem 165. [kolmogorov-loveland-total] *В определении случайности по Мизесу – Колмогорову – Лавлэнду можно ограничиться лишь всюду определёнными вычислимыми функциями F и G; получится тот же самый класс последовательностей.*

$\triangleleft$ Пусть имеются частичные функции F и G, выбирающие из некоторой последовательности ω бесконечную несбалансированную подпоследовательность. Разобьём эту последовательность на две, в зависимости от чётности порядкового номера члена в исходной последовательности. Одна из них будет бесконечной, но не сбалансированной. Поэтому без ограничения общности можно считать, что наше правило включает в подпоследовательность лишь (скажем) чётные члены последовательности. Тогда нечётные члены можно читать заранее, это не повредит, так как в последовательность включены они всё равно не будут. Поэтому, если частичное правило долго работает, можно параллельно про запас читать нечётные члены последовательности. Такое модифицированное правило задаётся всюду определёнными функциями F и G. (Если исходный алгоритм «завис», то новый будет читать подряд нечётные члены последовательности, но ничего так и не выберет.) $\triangleright$

(В этом доказательстве мы свели одно правило с частичными функциями к двум правилам со всюду определёнными функциями.)

Как соотносится новое определение с уже известными нам? Частичный ответ на этот вопрос даёт следующая теорема.

Theorem 166. [mises-kl-relations] (а) *Всякая случайная по Мизесу – Колмогорову – Лавлэнду последовательность случайна по Мизесу – Чёрчу – Дэли* (*и, следовательно, по Мизесу – Чёрчу*).

(**б**) *Всякая случайная по Мартин-Лёфу последовательность случайна по Мизесу – Колмогорову – Лавлэнду.*

Точнее, пункт (а) имеет место для любого действительного $p \in (0, 1)$; пункт (б) дополнительно предполагает, что p вычислимо и случайность по Мартин-Лёфу понимается относительно бернуллиевой меры с параметром p.

$\triangleleft$ (а) Допустимые по Мизесу – Чёрчу – Дэли (и тем более по Мизесу – Чёрчу) правила выбора являются частными случаями допустимых по Колмогорову – Лавлэнду правил.

(б) По существу это рассуждение повторяет аналогичное рассуждение для случайности по Мизесу – Чёрчу. Мы предполагаем, что фиксировано некоторое вычислимое p, для которого изучается случайность по Мизесу – Колмогорову – Лавлэнду (предел частот в подпоследовательностях должен быть равен p) и по Мартин-Лёфу (рассматривается вычислимая мера μ_p на Ω, соответствующая независимым испытаниям, в каждом из которых вероятность успеха равна p).

Пусть фиксированы вычислимые функции F и G, задающее допустимое по Мизесу – Чёрчу правило выбора.

Для каждого целого n и рационального q рассмотрим множество $D_{n,q}$ всех двоичных слов длины n, у которых доля единиц больше q. Мы знаем, что при фиксированном $q > p$ и при $n \to \infty$ мера множества $D_{n,q}$ (точнее, μ_p-мера множества $\mathbf{D}_{n,q}$ всех последовательностей, имеющих начало в $D_{n,q}$) экспоненциально убывает.

Рассмотрим теперь прообраз этого множества относительно $S_{F,G}$, то есть множество всех последовательностей ω, для которых описанный процесс выбора даёт подпоследовательность длины не меньше n и частота единиц среди первых n выбранных членов больше q. Легко заметить, что *μ_p-мера этого прообраза не больше $\mu_p(\mathbf{D}_{n,q})$.*

В самом деле, пусть t — некоторое двоичное слово длины $k - 1$. Рассмотрим условную вероятность того, что k-ая по счёту выбранная (включённая в подпоследовательность) карточка содержит единицу, при условии что предыдущие члены подпоследовательности образуют слово t и до выбора k-й карточки вообще дело дойдёт. Эта условная вероятность равна p, поскольку к моменту переворачивания k-й карточки её содержимое ещё никак не использовалось и мы можем «отложить» бросание монеты до этого момента. Более формально, событие, которое фигурирует в качестве условия, является счётным объединением непересекающихся вариантов, каждый из которых соответствует конкретному ходу процесса выбора вплоть до момента, когда выбрана (но ещё не перевёрнута) k-я карточка подпоследовательности (какие карточки перевёрнуты и что в них оказалось). В каждом из вариантов уже определено, какая карточка будет следующей выбрана, и этот вариант разбивается на две части в пропорции $p : (1-p)$ в зависимости от содержания этой карточки. Отсюда по индукции легко следует выделенное курсивом утверждение.

Остаётся заметить, что упомянутый прообраз не только имеет малую меру (быстро убывающую с ростом n), но ещё и перечислим (любой из вариантов зависит от конечного числа значений функций F и G и потому рано или поздно обнаружится), поэтому, как и в усиленном законе больших чисел, возникает эффективно нулевое множество (для каждого q — своё). Все эти эффективно нулевые множества (а также аналогичные множества для последовательностей с частотами меньше p) содержатся в максимальном эффективно нулевом множестве. $\triangleright$

В следующем разделе мы докажем такое усиление утверждения (б): если вычислимая последовательность p_n действительных чисел в интервале $(0,1)$ вычислимо сходится к пределу $p \in (0,1)$, то всякая последовательность, случайная по Мартин-Лёфу относительно произведения мер (испытания независимы, вероятность удачи в i-м испытании равна p_i), случайна по Мизесу – Колмогорову – Лавлэнду с пределом частот p.

Это позволит нам строить примеры случайных по Мизесу – Колмогорову – Лавлэнду последовательностей с плохими свойствами (неслучайные по Мартин-Лёфу, содержащие больше единиц, чем нулей и другие).

Пока же мы хотим действовать в другом направлении и показать — для случая равномерной меры, — что случайная по Мизесу – Колмогорову – Лавлэнду последовательность (в отличие от случайных по Мизесу – Чёрчу и Мизесу – Чёрчу – Дэли) имеет достаточно большую сложность начальных отрезков (теорема Мучника).

Theorem 167. [muchnik-complexity] *Пусть $\omega \in \Omega$ такова, что $KS(\omega_0 \ldots \omega_{n-1}) < \alpha n$ для некоторого $\alpha < 1$ и всех достаточно больших n. Тогда последовательность ω не случайна по Мизесу – Колмогорову – Лавлэнду.*

Для доказательства нам понадобится некоторое вспомогательное утверждение о цене «инсайдерской информации» в играх с ограниченными ставками. Сейчас мы его сформулируем и докажем, а потом перейдём к доказательству теоремы 167.

Представим себе, что крупье бросает монету раз за разом, и перед каждым бросанием мы можем сделать ограниченную ставку $u \in [-1,1]$ (положительные u соответствуют ставкам на единицу, отрицательные — на нуль). После этого мы получаем u рублей, если выпадает единица, и $(-u)$ рублей, если выпадает нуль.

В отличие от игр с мартингалами максимальная ставка всегда есть 1, независимо от того, сколько мы выиграли или проиграли в предыдущих партиях (и тем самым наш проигрыш может быть сколь угодно большим — а раньше он был ограничен начальным капиталом).

Во избежание путаницы подчеркнём ещё, что мы делаем ставки в том же порядке, в котором монету бросают (никаких немонотонных правил выбора пока нет).

Лемма. Пусть заранее известно множество A слов длины n, содержащее не более 2^s элементов для некоторого $s < n$. Тогда существует стратегия, гарантирующая выигрыш не менее $n-s$ на любом элементе множества A (для любой последовательности бросаний, соответствующей слову из множества A).

Например, если A содержит единственный элемент (то есть мы заранее знаем результаты всех бросаний), то можно выиграть n рублей (что не удивительно: в каждой партии мы выигрываем рубль). (В рассмотренном выше «мартингальном» варианте игры выигрыш был гораздо больше: 2^n.) Если мы знаем результаты каких-то фиксированных k бросаний, то по лемме можем выиграть k рублей (что тоже неудивительно). Чуть более сложный пример: пусть мы знаем, что число орлов будет чётно (что соответствует $s = n - 1$). Лемма говорит, что на этом можно выиграть рубль. Легко понять, как: нужно делать нулевые ставки до последнего момента, а результат последнего бросания нам фактически известен, и на него надо сделать максимальную ставку.

После этих примеров общее доказательство выглядит довольно естественно. В каждый момент нам известен некоторый начальный отрезок w последовательности (пусть его длина j). Имеется 2^{n-j} продолжений слова w, имеющих длину n, однако не все они принадлежат A. Рассмотрим условную вероятность попадания в A после w (долю продолжений, попадающих в A); минус логарифм этой вероятности назовём *информационным капиталом игрока*.

Вначале этот капитал равен $n - s$. Мы докажем, что можно так выбрать стратегию (так определять размер ставок в каждый момент), чтобы сумма информационного и реального капитала не убывала. Тогда в конце игры, когда информационный капитал равен нулю (мы предполагаем, что последовательность исходов на самом деле оказалась внутри множества A), реальный капитал будет не меньше $n - s$, что и требовалось.

Почему это возможно? Пусть в данный момент информационный капитал равен $(-\log p)$ (где p — текущая доля элементов из A среди продолжений). Зная A, мы можем найти этот капитал, а также узнать, как он изменится после появления нуля и единицы: после появления нуля он будет равен $(-\log p_0)$, а после единицы будет равен $(-\log p_1)$, где p_0 и p_1 — соответствующие доли. Ясно, что $p = (p_0 + p_1)/2$. Нам нужно найти такой размер очередной ставки d, чтобы в обоих случаях сумма информационного и реального капитала не уменьшилась:

$$\begin{gathered} -\log p_0 - d \geqslant -\log p; \\ -\log p_1 + d \geqslant -\log p; \end{gathered}$$

или, разрешая эти неравенства относительно d,

$$\begin{gathered} -\log p_0 + \log p \geqslant d \geqslant -\log p + \log p_1 \\ -\log(p_0/p) \geqslant d \geqslant \log(p_1/p) \end{gathered}$$

Ясно, что такое d можно выбрать, если и только если $p/p_0 \geqslant p_1/p$, то есть $p^2 \geqslant p_0p_1$ (а это так согласно неравенству о среднем арифметическом и геометрическом). Заметим также, p_0 и p_1 не больше $2p$, поэтому границы для d (и тем самым само d) не выходят за пределы промежутка $[-1, 1]$. Лемма доказана.

Эта лемма позволяет доказать такое утверждение:

Theorem 168. [muchnik-preliminary] *Пусть K — произвольная вычислимая всюду определённая верхняя оценка для функции KS, а последовательность $\omega \in \Omega$ такова, что*

$$K(\omega_0 \ldots \omega_{n-1}) \leqslant \alpha n$$

для некоторого $\alpha < 1$ и для всех достаточно больших n. Тогда последовательность ω не случайна по Мизесу – Чёрчу.

$\lhd$ Для каждого n можно алгоритмически построить список A_n всех слов длины n, для которых значение функции K не превосходит αn. Этот список будет содержать не более $2^{\alpha n + O(1)}$ слов. Мы знаем, что при достаточно больших n начальный отрезок последовательности ω попадает в A_n. Для этих n построенная по лемме (для множества A_n) стратегия будет выигрывать не меньше $(1-\alpha)n$ рублей при игре с начальным отрезком последовательности ω длины n.

Рассмотрим последовательность быстро растущих n_i (насколько быстро, что $n_{i-1}/n_i \to 0$) и соединим стратегии для A_{n_i} в одну. Фактически стратегия для A_{n_i} будет применяться не на всём начальном отрезке длины n_i, а лишь после окончания предыдущего отрезка n_{i-1}, что, по нашему предположению, составляет бесконечно малую долю от n_i. Поэтому соединённая стратегия будет успешна в том смысле, что выигрыш её на начальном отрезке длины n будет превосходить εn для фиксированного $\varepsilon > 0$ и для бесконечно многих n. (Достаточно взять $\varepsilon < 1 - \alpha$ и $n = n_i$ при бóльших i.)

А это противоречит случайности по Мизесу – Чёрчу (теорема 145, с. 233). $\rhd$

Теперь мы уже готовы к доказательству теоремы 167.

$\lhd$ По аналогии с только что проведённым доказательством, мы можем рассмотреть множество A_n всех слов длины n, имеющих сложность не более αn. Оно содержит примерно $2^{\alpha n}$ слов, среди которых заведомо имеется начальный отрезок последовательности ω. Но только теперь мы не можем алгоритмически построить список всех слов множества A_n, а можем лишь перечислять эти слова (не зная ни в какой момент, все ли слова уже появились или ещё нет). Преодолеть эту трудность можно, вспомнив, что нам разрешено использовать немонотонные правила выбора. Вот как это делается.

Снова рассмотрим быстро растущую вычислимую последовательность n_i, например, $n_i = i!$ Разобьём последовательность на отрезки, проведя границы в точках n_i; длина i-го отрезка будет $n_i - n_{i-1}$. Немного увеличив α, мы можем считать, что сложность i-го отрезка последовательности не больше α, умноженного на его длину. Другими словами, мы будем считать, что удельная (в расчёте на букву) сложность i-го отрезка последовательности не превосходит α. Пусть A_i — множество всех слов длины $n_i - n_{i-1}$, для которых удельная сложность не больше α. Мы знаем, что i-ый отрезок последовательности ω, который мы обозначим через ω^i, лежит в A_i (по крайней мере при достаточно бóльших i), и можем перечислять множество A_i, зная i.

Пусть t_i — число шагов этого перечисления, которое нужно сделать, прежде чем ω_i появится в A_i. Будем отдельно рассматривать чётные и нечётные номера, и сравним соседние значения t_{2m} и t_{2m+1}. Тривиальное замечание: либо $t_{2m} \leqslant t_{2m+1}$, либо $t_{2m+1} \leqslant t_{2m}$ (а может, и то, и другое). Что это нам даёт? А вот что: одна стратегия может посмотреть (не делая ставок, в порядке информации) участок ω^{2m}, после чего перечислять A_{2m} до появления ω^{2m}, и тем самым найти число t_{2m}. Затем эта стратегия делает столько же (то есть t_{2m}) шагов перечисления множества A_{2m+1}, надеется на то, что неизвестное ей ω_{2m+1} принадлежит перечисленной части множества A_{2m+1} и делает ставки как в предыдущей теореме. Это приведёт к успеху, если $t_{2m} \geqslant t_{2m+1}$; в противном случае мы можем всё проиграть. Но тогда парная стратегия, которая просматривает ω_{2m+1}, находит t_{2m+1}, а затем делает столько же шагов перечисления A_{2m}, будет успешной.

Таким образом, при каждом m у нас есть пара стратегий (со ставками в промежутке $[-1,1]$), одна из которых является успешной (выигрывает не менее $1-\alpha$ в расчёте на каждую сделанную ставку). (Формально говоря, такие стратегии имеются лишь при достаточно больши́х m, но мы можем все меньшие пропустить.) Одна из этих стратегий монотонна и (как в теореме 145) приближается средним арифметическим конечного числа правил выбора, допустимых по Чёрчу. Число правил зависит от допустимой погрешности; нам нужно, чтобы погрешность была мала по сравнению с $1-\alpha$, и этого можно добиться для фиксированного (не зависящего от m) числа правил, которое мы обозначим через N. Другая стратегия немонотонна, и потому получаются N правил выбора, допустимых по Колмогорову – Лавлэнду. Всего получается $2N$ правил выбора, имеющих дело с двумя соседними участками последовательности.

Соединяя правила для разных отрезков, получим $2N$ допустимых по Колмогорову – Лавлэнду правил выбора. Вспоминая, что n_{i-1}/n_i мало и результат выбора на предыдущем отрезке почти не меняет частоту на следующем, получаем, что для каждого m хотя бы одно из построенных $2N$ правил даёт значительное отклонение частоты. Значит, одно из них даёт такое отклонение для бесконечно многих m, и последовательность ω не случайна по Мизесу – Колмогорову – Лавлэнду.

Теорема 167 доказана. $\triangleright$

Из неё (и теоремы 163) вытекает такое следствие:

Theorem 169. [daley-not-kolmogorov] *Существуют случайные по Мизесу – Чёрчу – Дэли последовательности, не являющиеся случайными по Мизесу – Колмогорову – Лавлэнду.*

9.13. Случайность по изменённой мере

[miseslamb]

В этом разделе мы изложим метод (предложенный Ламбальгеном [28]), с помощью которого можно строить случайные по Мизесу – Колмогорову – Лавлэнду последовательности с разными «патологическими» свойствами (не случайные по Мартин-Лёфу, с превышением нулей над единицами в начальных отрезках и другие).

9.13.1. Случайность по двум мерам

[kakutani-section]

Для начала обсудим вопрос, имеющий и более общее значение: насколько зависит случайность (относительно данной меры) от выбора меры?

Вот два примера противоположного характера.

Пример 1. Рассмотрим равномерную меру μ, а также меру μ', в которой испытания независимы, причём во всех, кроме первого, вероятность успеха равна $1/2$, а в первом она равна, скажем, $2/3$. Интуитивно ясно, что случайными должны быть одни и те же последовательности: в самом деле, в первом испытании возможны оба исхода (и по той, и по другой мерам), а дальше меры одинаковы. Это действительно так для всех разумных определений случайности. Можно заметить, что эффективно нулевые множества относительно мер μ и μ' одни и те же (поскольку меры любого множества отличаются не более чем в два раза). Отсюда ясно, что случайными в смысле Мартин-Лёфа по этим мерам будут одни и те же последовательности.

$\boxed{\textbf{200}}$ Как доказать это, исходя из критерия случайности в терминах сложности?

$\boxed{\textbf{201}}$ Докажите, что случайными относительно вычислимых мартингалов (по мерам μ и μ') будут одни и те же последовательности.

Пример 2. Рассмотрим равномерную меру μ, а также меру μ', в которой испытания независимы и вероятность успеха в каждом из них равна $2/3$. Может ли одна и та же последовательность быть случайной (скажем, по Мартин-Лёфу) относительно μ и μ'? В полном согласии с нашей интуицией, нет — в самом деле, для случайной последовательности предел частот равен вероятности успеха, а $1/2 \neq 2/3$.

Возникает следующий естественный вопрос. Пусть имеются две последовательности чисел $p_i, p_i' \in (0,1)$. Рассмотрим меры μ и μ' на Ω, соответствующие независимым испытаниям, вероятность успеха в i-м равна p_i для меры μ и p_i' для меры μ'. Что можно сказать о классах случайных последовательностей относительно этих двух мер? Естественно ожидать, что при близких p_i и p_i' случайными должны быть одни и те же последовательности, а при сильно отличающихся p_i и p_i' классы случайных последовательностей не должны пересекаться.

Мы докажем, что это так и есть для случая отделённых от 0 и 1 вероятностей, когда все p_i и p_i' лежат в интервале $(\varepsilon, 1-\varepsilon)$ для некоторого положительного ε. Кроме того, мы предполагаем, что последовательности p_i и p_i' являются вычислимыми последовательностями вычислимых действительных чисел (иначе не имеет смысла говорить о случайности по Мартин-Лёфу).

Имеет место такой критерий (конструктивный аналог классической теоремы Какутани):

Theorem 170. `[kakutani-constructive]` *(а) Если сумма ряда $\sum(p_i - p_i')^2$ конечна, то классы случайных по Мартин-Лёфу последовательностей относительно мер μ и μ' совпадают.*

(б) Если сумма ряда $\sum(p_i - p_i')^2$ бесконечна, то эти классы не пересекаются.

$\triangleleft$ Кажется естественным доказывать пункт (а) следующим образом. Пусть последовательность $\omega \in \Omega$ случайна по мере μ, соответствующей вероятностям p_i. Тогда (монотонная) сложность её начального отрезка длины n близка к логарифму его вероятности, которая равна произведению

$$\prod_{i=0}^{n-1} r_i$$

где $r_i = p_i$ при $\omega_i = 1$ и $r_i = 1 - p_i$ при $\omega_i = 0$. Если p_i близко к p_i', то r_i близко к r_i' (где r_i' определяется аналогичным образом для другой меры). Поэтому произведение всех r_i близко к произведению всех r_i', так что случайность по одной мере влечёт за собой случайность и по другой.

Это рассуждение несложно формализовать, но при этом нам придётся предположить, что разность

$$\sum_{i=0}^{n-1}(-\log r_i) - \sum_{i=0}^{n-1}(-\log r_i')$$

ограничена; легко проверить, что это так, когда $\sum |p_i - p'_i| < \infty$ (напомним, что мы предполагаем p_i и p'_i отделёнными от нуля и единицы). А это больше, чем нам дано по условию; нам известно лишь, что сумма *квадратов* ограничена.

Как же быть? Заметим, что на самом деле нам достаточно, чтобы упомянутая разность была ограничена для любой *случайной* по мере μ последовательности ω. Мы знаем, что тогда

$$\mathit{KM}(\omega_0 \ldots \omega_{n-1}) = \sum_{i=0}^{n-1} (-\log r_i) + O(1)$$

(сложность отличается от логарифма вероятности не более чем на константу), поэтому верхняя оценка для сложности (которая имеет место для любой вычислимой меры, в частности, для μ') даёт нам

$$\sum_{i=0}^{n-1} (-\log r_i) \leqslant \sum_{i=0}^{n-1} (-\log r'_i) + O(1)$$

Обозначим разницу между $r'_i - r_i$ через δ_i. Используя это обозначение и переходя к произведениям, получим (верное с точностью до мультипликативной константы) неравенство

$$\prod_{i=0}^{n-1} r_i \geqslant \prod_{i=0}^{n-1} (r_i + \delta_i)$$

Мы знаем, что $\sum_i \delta_i^2 < \infty$ и, в частности, $\delta_i \to 0$ при $i \to \infty$. Поэтому почти все δ_i меньше числа ε, отделяющего наши вероятности от нуля и единицы. Изменив конечное число членов p'_i (что не влияет на случайность), мы можем считать, что это верно для всех i. Тогда можно рассмотреть меру μ'', «симметричную» мере μ' относительно μ (это означает, что p_i есть середина отрезка $[p'_i, p''_i]$). Для этой меры можно записать аналогичное неравенство, которое будет отличаться лишь знаком при δ_i:

$$\prod_{i=0}^{n-1} r_i \geqslant \prod_{i=0}^{n-1} (r_i - \delta_i)$$

(оно также верно с точностью до константы). Перемножим эти два неравенства:

$$\prod_{i=0}^{n-1} r_i^2 \geqslant \prod_{i=0}^{n-1} (r_i^2 - \delta_i^2)$$

А это неравенство (само по себе очевидное, и даже безо всякой константы), в силу наших предположений о δ_i является равенством (с точностью до ограниченного и отделённого от нуля множителя): как известно из курса математического анализа, бесконечное произведение $\prod(1 - h_i)$ больше нуля (при $0 < h_i < 1$) тогда и только тогда, когда $\sum h_i < \infty$.

Следовательно, и оба перемножаемых неравенства являются равенствами с точностью до константы. После этого критерий случайности (теорема 82) говорит, что последовательность ω случайна и по мере μ' (а также μ'', что для нас неважно). Утверждение (а) доказано.

Утверждение (б) является частным случаем более общей теоремы, в которой рассматриваются зависимые испытания, и мы докажем его сразу в общем случае (см. ниже теорему 171). $\triangleright$

$\boxed{202}$ Пусть $p_0, p_1, p_2, \ldots$ и $p'_0, p'_1, p'_2, \ldots$ находятся в интервале $(\varepsilon, 1-\varepsilon)$ для некоторого $\varepsilon > 0$, а меры μ и μ' соответствуют независимым испытаниям с вероятностями p_i и p'_i. Докажите, что если $\sum(p_i - p'_i)^2 < \infty$, то нулевыми по мерам μ и μ' являются одни и те же множества. [Указание. Воспользуйтесь теоремой 170 в релятивизованном варианте и заметьте, что для всякого нулевого множества можно найти оракул, относительно которого меры вычислимы, а множество эффективно нулевое.]

[Во что превратится доказательство, если развернуть все ссылки на алгоритмическую теорию информации и получить чисто классическое доказательство???]

Пусть μ — произвольная мера на пространстве Ω, а $p(x)$ — соответствующая функция на двоичных словах: $p(x) = \mu(\Omega_x)$. Для каждого слова x, которое имеет положительную вероятность появления в качестве начального отрезка последовательности (что означает $p(x) > 0$), рассмотрим условную вероятность появления единицы после x, то есть отношение $p(1|x) = p(x1)/p(x)$. (Если бросания независимы, то эта величина зависит лишь от длины слова x и обозначалась нами раньше через p_i.)

Если теперь, помимо меры, задана ещё и последовательность $\omega \in \Omega$, то можно рассмотреть условные вероятности появления единицы на каждом шаге для этой последовательности, то есть последовательность чисел

$$p_i = p(1|\omega_0 \ldots \omega_{i-1}) = p(\omega_0 \ldots \omega_{i-1}1)/p(\omega_0 \ldots \omega_{i-1})$$

(которые заведомо определены, если p не обращается в нуль на начальных отрезках последовательности).

Theorem 171. [kakutani-vovk] *Если последовательность ω случайна в смысле Мартин-Лёфа по двум мерам μ и μ', и числа p_i и p'_i, построенные описанным образом для этих мер, все находятся в интервале $(\varepsilon, 1-\varepsilon)$ для некоторого $\varepsilon > 0$, то*

$$\sum_i (p_i - p'_i)^2 < \infty$$

Заметим, что отсюда прямо следует утверждение теоремы 170. Отметим также, что для случайной последовательности числа p_i определены, так как знаменатель дроби $p(x1)/p(x)$ не может обращаться в нуль (слово, имеющее нулевую вероятность, не может быть началом случайной последовательности).

$\triangleleft$ Рассмотрим третью меру $\tilde{\mu}$, которая усредняет вероятности появления единицы для мер μ и μ'. А именно, вероятность появления единицы после любого слова x с точки зрения меры $\tilde{\mu}$ равна среднему арифметическому вероятностей появления единицы после того же x с точки зрения мер μ и μ'. (Заметим, что это *не* соответствует усреднению мер: если положить $\tilde{\mu}(X) = (\mu(X) + \mu'(X))/2$, тоже получится мера, но совсем другая!)

Если последовательность случайна, то для сложности её начальных отрезков можно записать равенство

$$KM(\omega_0 \ldots \omega_{n-1}) = \sum_{i=0}^{n-1}(-\log r_i) + O(1),$$

где, как и раньше, $r_i = p_i$ при $\omega_i = 1$ и $r_i = (1 - p_i)$ при $\omega_i = 0$. Аналогичное равенство можно записать и для r'_i (построенных для меры μ'), а для $\tilde{r}_i$ (построенных по мере $\tilde{\mu}$) можно записать лишь неравенство (сложность не больше минус логарифма меры).

Исключая из этих равенств и неравенств колмогоровскую сложность, получаем неравенства для условных вероятностей:

$$\sum_{i=0}^{n-1}(-\log r_i) \leqslant \sum_{i=0}^{n-1}(-\log \tilde{r}_i) + O(1);$$
$$\sum_{i=0}^{n-1}(-\log r'_i) \leqslant \sum_{i=0}^{n-1}(-\log \tilde{r}_i) + O(1).$$

После сложения и деления пополам получается неравенство

$$\sum_{i=0}^{n-1}\frac{(-\log r_i) + (-\log r'_i)}{2} \leqslant \sum_{i=0}^{n-1}(-\log \tilde{r}_i) + O(1);$$

Вспоминая, что $\tilde{r}_i = (r_i + r'_i)/2$ и перенося всё в одну часть, получаем

$$\sum_{i=0}^{n-1}\left(\frac{(-\log r_i) + (-\log r'_i)}{2} - \left(-\log\frac{r_i + r'_i}{2}\right)\right) \leqslant O(1)$$

Каждое слагаемое в левой части неотрицательно (выпуклость логарифма) и по порядку величины равно $(r_i - r'_i)^2$, поэтому $\sum(r_i - r'_i)^2 = \sum(p_i - p'_i)^2 < \infty$, что и требовалось доказать. $\triangleright$

Отметим, что в этой теореме говорится лишь о близости мер вдоль последовательности ω, случайной по обеим мерам; на других последовательностях они могут и сильно различаться.

203 Приведите соответствующий пример. [Указание. Рассмотрим две меры, которые на левой половине отрезка равномерны, а на правой сильно различаются.]

204 Покажите, что фактически мы использовали не случайность по Мартин-Лёфу, а более слабое свойство случайности относительно вычислимых мартингалов.

Приведённое доказательство имеет естественную игровую интерпретацию. Представим себе, что имеются два тотализатора, принимающих ставки на одну и ту же последовательность событий (каждое событие имеет два исхода: нуль и единицу). Однако их организаторы по-разному оценивают шансы и поэтому принимают ставки на разных условиях. В одном случае вероятности появления единицы и нуля считаются равными p и q (и потому поставленная на 1 [0] сумма возвращается с коэффициентом $1/p$ $[1/q]$), а в другом случае вероятности считаются равными p' и q' (а коэффициенты равны $1/p'$ и $1/q'$). (Вероятности p, p', q, q' могут меняться от события к событию).

В этом случае можно одновременно участвовать в двух играх, причём делать это так, чтобы по крайней мере в одной выиграть. Более того, в случае, когда все p, q, p', q' отделены от нуля и сумма $\sum(p - p')^2$ (по всей последовательности событий) бесконечна, то хотя бы в одной из игр наш выигрыш будет неограничен.

Как этого добиться? Пусть в первой игре у нас имеется капитал u, оценки вероятностей для следующего события равна p, p', q, q'. Поделим u между двумя ставками так, чтобы после игры наш капитал был равен $(p + p')u/2p$ (если выпадет единица) и $(q + q')u/2q$ (если выпадет нуль). (Легко проверить, что это допускается правилами игры, то есть что математическое ожидание капитала после партии равно u.) Во второй игре капитал v превращается после игры в $(p + p')v/2p'$ (если выпадет единица) и $(q + q')v/2q'$ (если выпадет нуль). Будем следить за произведением капиталов в обеих играх. Оно умножается либо на $(p + p')^2/4pp'$, либо на $(q + q')^2/4qq'$. Неравенство о среднем арифметическом и среднем геометрическом показывает, что в обоих случаях произведение uv возрастает. Переходя к логарифмам и оценивая этот рост, легко заметить, что (для отделённых от нуля p, p', q, q') при $\sum(p - p')^2 = \infty$ произведение капиталов стремится к бесконечности, и, следовательно, хотя бы в одной игре капитал не ограничен.

Другими словами, мы построили для каждой из мер свой мартингал с таким свойством: на любой последовательности, для которой условные вероятности по обеим мерам отделены от 0 и 1 и различия в условных вероятностях имеют бесконечную сумму квадратов, хотя бы один из мартингалов неограничен.

205 Пусть $p_0, p_1, p_2, \ldots$ и $p'_0, p'_1, p'_2, \ldots$ находятся в интервале $(\varepsilon, 1 - \varepsilon)$ для некоторого $\varepsilon > 0$, а меры μ и μ' соответствуют независимым испытаниям с вероятностями p_i и p'_i. Докажите, что если $\sum(p_i - p'_i)^2 = \infty$, то существует множество, имеющее меру 0 относительно μ и меру 1 по мере μ'. [Указание. Используйте теорему 171 в релятивизованном варианте, выбрав оракул, относительно которого всё вычислимо. Можно вместо этого повторить рассуждение с мартингалами.]

9.13.2. Закон больших чисел для переменных вероятностей

[lln-variable]

Усиленный закон больших чисел говорит, что для любого $p \in (0, 1)$ и соответствующей ему бернуллиевой меры μ_p (испытания независимы, вероятность успеха в каждом равна p) множество последовательностей, у которых предел частот равен p, имеет меру 1 (а его дополнение, то есть множество последовательностей, не имеющих предела частот или имеющих другой предел частот, является нулевым).

Пусть теперь вероятности не постоянны, и в i-м испытании вероятность успеха равна p_i. Естественно ожидать, что если $p_i \to p$, то предел частот для почти всех последовательностей будет равен p. Не вполне строго это можно объяснить так: возьмём какое-то $\varepsilon > 0$. Все частоты p_i, кроме конечного числа (а им можно пренебречь), меньше $p + \varepsilon/2$. Мы знаем, что если заменить вероятности p_i на $p+\varepsilon/2$, то почти наверное начиная с некоторого места частота будет меньше $p + \varepsilon$. Значит, это тем более так для $p_i < p + \varepsilon/2$ по монотонности.

Это рассуждение можно (с некоторым трудом, впрочем) сделать строгим, но мы предпочитаем доказать более общую оценку, которая полезна во многих случаях. Она относится к произвольной вычислимой мере μ на пространстве Ω; пусть p — соответствующая ей функция на словах. Для любого конечного слова $x = x_0 \ldots x_{n-1}$ длины n рассмотрим число единиц m в этом слове, а также сумму $p_0+\ldots+p_{n-1}$ условных вероятностей появления единицы в каждой его позиции:

$$p_i = p(x_0 \ldots x_{i-1}1)/p(x_0 \ldots x_{i-1}).$$

Обе эти величины зависят от слова x; наша оценка утверждает, что с большой вероятностью (по мере μ) для данной длины n эти величины будут близки. Вот точная формулировка (смысл m и p_i объяснён только что, вероятность берётся по мере μ):

Theorem 172. [chernov-variable]

$$\Pr[|m - (p_0 + \ldots + p_{n-1})| > d] \leqslant 2e^{-d^2/4n}$$

(Заметим, что бесконечные последовательности в этой формулировке по существу не нужны; фактически речь идёт о неравенстве, верном для любого распределения вероятностей на последовательностях нулей и единиц длины n.)

$\lhd$ Отдельно оцениваем вероятность избытка и недостачи; оба варианта симметричны, поэтому достаточно доказать, что

$$\Pr[m - (p_0 + \ldots + p_{n-1}) > d] \leqslant e^{-d^2/4n}$$

Мы используем обычный приём: строим другую меру μ', для которой отношение μ'/μ велико на всех последовательностях, попадающих в событие из левой части. (Отношение μ'/μ будет мартингалом, который велик, а именно, не меньше $e^{d^2/4n}$ на всех таких последовательностях.)

Поскольку мы хотим увеличения меры для последовательностей, в которых много единиц, то естественно, что в новой мере нужно условную вероятность единицы увеличить (по сравнению с μ). А именно, если раньше после некоторого слова x условные вероятности единицы и нуля были p и q, то теперь мы положим их равными

$$p' = p + \varepsilon pq; \quad q' = q - \varepsilon pq,$$

где ε — некоторое положительное число, меньшее $1/2$ (его мы выберем позже). Легко проверить, что p' и q' не выйдут за пределы отрезка $[0,1]$.

Получаем новое распределение вероятностей на последовательностях нулей и единиц длины n. Во сколько раз оно отличается от исходного на некотором слове x длины n? Каждый новый символ слова x даёт множитель p'/p (если это была единица) или q'/q (если это был нуль), где p, q, p', q' — условные вероятности единицы и нуля в исходной и изменённой мерах. Переходя к логарифмам, видим, что логарифм интересующего нас отношения равен сумме величин

$$\ln(p'/p) = \ln(1 + \varepsilon q) \geqslant \varepsilon q - (\varepsilon q)^2 \geqslant \varepsilon(1-p) - \varepsilon^2$$

или

$$\ln(q'/q) = \ln(1 - \varepsilon p) \geqslant -\varepsilon p - (\varepsilon p)^2 \geqslant -\varepsilon p - \varepsilon^2$$

для всех букв слова x (с соответствующими вероятностями p и q); первый вариант для единицы и второй вариант для нуля. (Мы использовали неравенство $\ln(1+h) \geqslant h - h^2$, выполненное при всех $|h| \leqslant 1/2$.)

Что получится, если сложить все величины? Помимо общего множителя ε, в правой части будет число единиц m (от каждого слагаемого $(1-p)$ по единице) минус сумма всех

p_i, где p_i — условная вероятность единицы на очередном месте в слове x, и ещё минус $n\varepsilon^2$:

$$\ln \frac{p'(x)}{p(x)} \geqslant \varepsilon(m - \sum p_i) - n\varepsilon^2.$$

Для тех x, которые нас интересуют (у которых превышение больше d) имеем

$$\ln \frac{p'(x)}{p(x)} > \varepsilon d - n\varepsilon^2 = \varepsilon(d - n\varepsilon).$$

Это верно при всех $\varepsilon \in (0, 1/2)$, поэтому выберем то из них, при котором правая часть наибольшая, а именно $\varepsilon = d/2n$ (ясно, что случай $d > n$ можно не рассматривать, так как число единиц никогда не превосходит n, поэтому $\varepsilon = d/2n < 1/2$).

Получаем, что

$$\ln \frac{p'(x)}{p(x)} > d^2/4n$$

и

$$\frac{p'(x)}{p(x)} > e^{d^2/4n}$$

что нам и требовалось. $\triangleright$

(Поучительно посмотреть, какая оценка получается для простейшего случая, когда нули и единицы равновероятны. Вероятность того, что число единиц превосходит ожидаемое $n/2$ более чем на $2\sqrt{n}$, не больше $1/e$, более чем на $2k\sqrt{n}$ — не больше $1/e^k$. Видно, что наша оценка не оптимальная — теорема Муавра – Лапласа даёт немного лучше — но близка к ней, отличие лишь в коэффициенте в показателе экспоненты.)

Теперь можно повторить доказательство усиленного закона больших чисел с этой новой оценкой и получить такую теорему, верную для произвольной меры μ на пространстве Ω:

Theorem 173. *Для последовательности $\omega = \omega_0\omega_1\ldots$, распределённой по мере μ, почти наверное выполнено такое свойство:*

$$\lim_{n\to\infty} \left(\frac{m}{n} - \frac{p_0 + \ldots + p_{n-1}}{n}\right) = 0,$$

где m — число единиц среди $\omega_0, \omega_1, \ldots, \omega_{n-1}$, а p_i — условная вероятность появления единицы после $\omega_0\omega_1\ldots\omega_{i-1}$.

В частности, из этой теоремы следует, что если испытания независимы, а вероятности успеха p_i стремятся к некоторому пределу p, то с вероятностью единица предел частот равен p (поскольку предел средних арифметических $(p_0 + \ldots + p_{n-1})/n$ равен p).

Мы, однако, хотим большего — чтобы с вероятностью единица последовательность ω была случайной по Мизесу (в том или ином варианте), то есть чтобы не только для самой последовательности, но и для её «законно выбранных» подпоследовательностей предел частот равнялся p.

Для начала рассмотрим случай монотонных правил выбора (задаваемых множествами слов, после которых делается ставка).

9.13.3. Закон больших чисел для подпоследовательностей

Пусть имеется некоторая мера μ на пространстве Ω, а также некоторое правило выбора, задаваемое множеством R. Напомним, что такое правило задаёт отображение $S_R\colon \Omega \to \Sigma$: от последовательности ω остаются только те члены ω_i, для которых $\omega_0 \ldots \omega_{i-1} \in R$.

Параллельно с выбранной подпоследовательностью можно записывать и условные вероятности единицы перед появлением каждого из членов подпоследовательности. Другими словами, для данной последовательности ω мы не только отбираем те ω_i, при которых $\omega_0 \ldots \omega_{i-1} \in R$, но также записываем условные вероятности p_i появления единицы:

$$p_i = p(\omega_0 \ldots \omega_{i-1}1)/p(\omega_0 \ldots \omega_{i-1}),$$

и наряду с подпоследовательностью

$$S_R(\omega) = \omega_{i_0}\omega_{i_1}\omega_{i_2}\ldots$$

(здесь $i_0 < i_1 < i_2 < \ldots$ — номера выбранных членов) получаем ещё и последовательность вероятностей

$$p_{i_0}, p_{i_1}, p_{i_2}, \ldots$$

Theorem 174. [chernov-variable-subsequence] *Пусть фиксирована мера μ на пространстве Ω и правило выбора, то есть множество слов R. Тогда для любого $n > 0$ выполнена оценка теоремы 172:*

$$\Pr[|m - (p_{i_0} + \ldots + p_{i_{n-1}})| > d] \leqslant 2e^{-d^2/4n},$$

если m — число единиц среди первых n выбранных членов подпоследовательности ($m = \omega_{i_0} + \ldots + \omega_{i_{n-1}}$), а $p_{i_0}, \ldots, p_{i_{n-1}}$ — условные вероятности появления единицы на соответствующих местах.

Вообще говоря, для некоторых последовательностей ω выбранная подпоследовательность $S_R(\omega)$ содержит менее n членов; такие ω не входят в событие, вероятность которого оценивается в этой теореме.

Пример. Рассмотрим правило выбора, которое отбирает те члены последовательности, для которых условная вероятность появления единицы меньше $0{,}5$. Тогда утверждение теоремы гарантирует, что доля единиц среди первых n выбранных членов превышает 51% лишь с малой (экспоненциально убывающей с ростом n) вероятностью.

$\lhd$ Будем действовать так же, как при доказательстве теоремы 172; при построении новой меры мы меняем условные вероятности только в тех позициях, где очередной член включается в подпоследовательность (а в остальных случаях условные вероятности остаются неизменными). Тогда в оценку для отношения вероятностей будет входить длина подпоследовательности (вместо длины всей последовательности), число выбранных единиц (вместо числа всех единиц) и сумма вероятностей в моменты выбора (вместо суммы всех вероятностей), то есть получится ровно то, что требуется. $\rhd$

Доказанная теорема позволяет строить случайные по Мизесу – Чёрчу – Дэли последовательности. Вот как это делается.

Пусть $p_0, p_1, \ldots$ — вычислимая последовательность вычислимых действительных чисел, вычислимо сходящаяся к числу $p \in (0,1)$. (Вычислимая сходимость означает, что по данному $\varepsilon > 0$ можно алгоритмически указать то N, начиная с которого члены последовательности отстоят от предела меньше чем на ε; ясно, что тогда p вычислимо.) Рассмотрим вычислимую меру μ, соответствующую независимым испытаниям с вероятностями p_i.

Theorem 175. `[lambalgen-original]` *Всякая случайная в смысле Мартин-Лёфа по мере μ последовательность является случайной по Мизесу – Чёрчу – Дэли с пределом частот p.*

◁ Пусть фиксированы некоторое (рациональное) $\varepsilon > 0$ и некоторое допустимое в смысле Чёрча – Дэли (вычислимое, возможно не всюду определённое) правило выбора R. Мы должны показать, что множество U тех последовательностей, для которых применение этого правила даёт бесконечную последовательность и частота единиц в ней бесконечно много раз превышает $p + \varepsilon$, является эффективно нулевым. (Аналогично для частот, меньших $p - \varepsilon$.)

Фиксируем некоторое n и рассмотрим множество U_n тех ω, для которых выбранная подпоследовательность имеет длину не менее n и доля единиц среди первых n членов больше $p + \varepsilon$. Это множество эффективно открыто (применяя правило выбора вдоль всех ветвей, мы можем перечислять все последовательности, продолжения которых попадают в U_n).

Теорема 174 даёт верхнюю оценку для множества U_n; для достаточно больших n среднее арифметическое вероятностей не больше $p + \varepsilon/2$ (и границу можно найти, пользуясь вычислимой сходимостью), и даваемая этой теоремой оценка экспоненциально убывает с ростом n. Поэтому можно покрыть U перечислимым семейством интервалов сколь угодно малой меры (взяв все интервалы, образующие множества $U_N, U_{N+1}, U_{N+2}, \ldots$ для достаточно большого N; напомним, что всякий элемент U по определению входит в бесконечно много разных U_n). Следовательно, U является эффективно нулевым и не может содержать случайной в смысле Мартин-Лёфа последовательности. ▷

На самом деле в тех же предположениях верно и более сильное утверждение:

Theorem 176. `[lambalgen-modified]` *Всякая случайная в смысле Мартин-Лёфа по мере μ последовательность является случайной по Мизесу – Колмогорову – Лавлэнду с тем же пределом частот p.*

◁ Применение допустимого по Колмогорову – Лавлэнду правила $S_{F,G}$ к последовательности можно разбить на два этапа. Сначала с помощью функции F отбирается подпоследовательность просмотренных членов (как выбранных, так и не выбранных). Затем с помощью функции G производится выбор, и эта вторая часть соответствует применению допустимого по Чёрчу-Дэли правила.

Посмотрим на результат первого этапа. Как распределена получающаяся последовательность ω_F, если исходная последовательность ω получается в результате независимых испытаний с вероятностью успеха p_i в i-м испытании?

Первый член ω_F имеет в ω заранее известный номер $n_0 = F(\Lambda)$ (наше правило начинает с просмотра этого члена). Вероятность появления единицы в начале ω_F есть, таким образом, p_{n_0}. Какой член будет вторым в ω_F? Это уже зависит от результата первого просмотра.

Соответственно для вероятности появления единицы на втором месте есть две возможности, соответствующие двум разным позициям в ω. Общее правило: вероятность появления 1 после некоторого слова x в ω_F равна $p_{F(x)}$, поскольку в этом случае следующим будет просмотрен член с номером $F(x)$. Заметим, что в силу ограничений на F (никакой член не просматривается дважды) условные вероятности вдоль любой ветви образуют подпоследовательность последовательности p_i без повторений, и потому среди них вне интервала $(p-\varepsilon/2, p+\varepsilon/2)$ окажется не больше, чем в исходной последовательности $p_0, p_1, \ldots$.

Это позволяет применить оценку теоремы 174 к последовательности ω_F и правилу выбора, задаваемому функцией G, как и в доказательстве теоремы 175.

Поскольку функции F и G вычислимы, множество тех последовательностей ω, у которых правило выбора $S_{F,G}$ даёт подпоследовательность, начинающееся на данное слово x, может быть эффективно перечислено по x, так что после получения оценки на вероятность больших уклонений частоты от p мы легко строим нужное нам эффективно нулевое множество. $\triangleright$

Замечание. Можно было бы дать и более прямое доказательство теоремы 176. Вот одно из возможных рассуждений (заимствованное из [71]).

Пусть фиксированы некоторое (рациональное) $\varepsilon > 0$ и функции F, G, задающее допустимое в смысле Колмогорова – Лавлэнда правило выбора R. Мы должны показать, что множество U тех последовательностей, для которых применение этого правила даёт бесконечную последовательность и частота единиц в ней бесконечно много раз превышает $p+\varepsilon$, является эффективно нулевым относительно меры μ. (Аналогично для частоты, меньшей $p-\varepsilon$.)

Обозначим через U_n множество тех последовательностей ω, для которых $S_{F,G}(\omega)$ содержит не менее n членов и частота единиц среди первых n членов больше $p+\varepsilon$. Достаточно доказать, что ряд $\sum \mu(U_n)$ вычислимо сходится (где μ — рассматриваемая нами мера, соответствующая независимым испытаниям с вероятностями успеха p_i).

Обозначим через $\alpha_{n,k}(r_1, \ldots, r_n)$ вероятность того, что в n независимых испытаниях с вероятностями успеха $r_1, \ldots, r_n$ случилось не менее k успехов. Функция $\alpha_{n,k}$ неубывает по всем своим аргументам (и, кстати, является полилинейной, то есть многочленом от r_i степени не более 1 по каждому аргументу). Кроме того, $\alpha_{n,k} \leqslant \alpha_{n,l}$ при $k \geqslant l$.

Мы утверждаем, что $\mu(U_n) \leqslant \alpha_{n,k}(r_1, \ldots, r_n)$, где $k = n(p+\varepsilon)$, а r_i — i-ый по величине член последовательности $p_0, p_1, \ldots$ (r_1 — максимальный, r_2 — максимальный из оставшихся и т. д.; точнее, r_1 — точная верхняя грань всех p_i, r_1 — точная верхняя грань минимумов пар $\min(p_i, p_j)$ при $i \neq j$ и так далее).

Покажем, как из этой оценки можно получить сходимость ряда $\mu(U_n)$. Очевидно, что $r_1 \geqslant r_2 \geqslant \ldots$ и $\lim r_i = p$. Заменим те r_i, которые больше $p+\varepsilon/2$, на единицу (пусть их количество равно s), а остальные — на $p+\varepsilon/2$. Получим, что

$$\mu(U_n) \leqslant \alpha_{n-s,k-s}(p+\varepsilon/2, \ldots, p+\varepsilon/2),$$

так что мы свели дело к оценке уклонений для случая постоянных вероятностей, который мы уже многократно разбирали. (Важно отметить, что при больших n отношение $(k-s)/(n-s)$ примерно равно $p+\varepsilon$ и заметно больше $p+\varepsilon/2$.)

Осталось доказать указанную оценку для $\mu(U_n)$. Будем представлять себе (см. выше), что члены последовательности написаны на карточках, лежащих лицом вниз, и что правило выбора говорит, какие карточки нужно перевернуть для просмотра и какие — для включения в выбранную подпоследовательность. При этом сведения о том, какие карточки переворачивались, и что на них оказалось, записываются в протокол применения правила выбора к последовательности. Пусть π — начальный отрезок такого протокола. Через $n(\pi)$ обозначим количество членов, включённых в подпоследовательность на отрезке π, а через $k(\pi)$ — количество единиц среди этих членов. Через $r_i(\pi)$ обозначим i-ый по величине член последовательности, получаемой из $p_0, p_1, \ldots$ выкидыванием членов, соответствующим перевёрнутым в ходе π карточкам (независимо от того, были ли они включены в подпоследовательность или только просмотрены).

Через $\mu(U_n|\pi)$ обозначим условную вероятность события U_n при условии того, что протокол применения правила выбора начинается на π. Мы докажем следующее обобщение интересующего нас неравенства: при $n(\pi) \leqslant n$

$$\mu(U_n|\pi) \leqslant \alpha_{n-n(\pi),k-k(\pi)}(r_1(\pi), r_2(\pi), \ldots) \tag{$*$}$$

(при пустом π получаем интересующую нас оценку). Если $n(\pi) = n$, то это неравенство превращается в равенство (левая и правая части одновременно равны либо нулю, либо единице). Пусть $n(\pi) < n$ и m — номер карточки, переворачиваемой сразу после π (если такой нет, то $\mu(U_n|\pi) = 0$). Тогда

$$\mu(U_n|\pi) = p_m \mu(U_n|\pi_1) + (1 - p_m)\mu(U_n|\pi_0),$$

где π_0 и π_1 — протоколы, получаемые добавлением к π информации о нуле (единице) на m-й карточке. Покажем, что если доказываемое нами неравенство $(*)$ верно для π_0 и π_1, то оно верно и для π. В самом деле, тогда $\mu(U_n|\pi)$ не превосходит

$$p_m \alpha_{n-n(\pi_1),k-k(\pi_1)}(r_1(\pi_1), \ldots) + (1 - p_m)\alpha_{n-n(\pi_0),k-k(\pi_0)}(r_1(\pi_0), \ldots) \tag{$**$}$$

Если m-я карточка была выбрана только для просмотра, то $n(\pi_0) = n(\pi_1) = n(\pi)$ и $k(\pi_0) = k(\pi_1) = k(\pi)$, и остаётся воспользоваться монотонностью $\alpha_{n,k}$ и тем, что $r_i(\pi_0) = r_i(\pi_1) \leqslant r_i(\pi)$. Если же m-я карточка включена в подпоследовательность, то $n(\pi_0) = n(\pi_1) = n(\pi) + 1$, $k(\pi_0) = k(\pi)$ и $k(\pi_1) = k(\pi) + 1$, а выражение $(**)$ равно

$$\alpha_{n-n(\pi),k-k(\pi)}(p_m, r_1(\pi_1), r_2(\pi_1), \ldots)$$

(заметим, что $r_i(\pi_1) = r_i(\pi_0)$) и тем самым не превосходит

$$\alpha_{n-n(\pi),k-k(\pi)}(r_1(\pi), r_2(\pi), \ldots)$$

Это завершает доказательство неравенства $(*)$ в том случае, когда все начальные отрезки протоколов π с $n(\pi) = n$ имеют ограниченную длину. Если же это не так, то приведённые рассуждения позволяют получить оценку для $\mu(U_{n,t}|\pi)$, где $U_{n,t}$ — множество тех последовательностей, для которых после не более чем t переворачиваний карточек будет выбрана подпоследовательность длины не меньше n с числом единиц (в начальном отрезке длины n) не менее k. Остаётся лишь перейти к пределу при $t \to \infty$.

9.13.4. Примеры

Теперь уже легко построить примеры случайных по Мизесу – Колмогорову – Лавлэнду последовательностей с различными «патологическими» свойствами.

Theorem 177. [miseskl-pathology] (**а**) *Существует случайная по Мизесу – Колмогорову – Лавлэнду последовательность* (*с частотой* $1/2$), *не случайная по Мартин-Лёфу* (*относительно равномерной меры*).

(**б**) *Существует случайная по Мизесу – Колмогорову – Лавлэнду последовательность* (*с частотой* $1/2$), *любой начальный отрезок которой содержит не меньше нулей, чем единиц.*

(**в**) *Существует случайная по Мизесу – Колмогорову – Лавлэнду* (*с частотой* $1/2$) *последовательность, из которой с помощью допустимого по Мизесу – Колмогорову – Лавлэнду* (*и даже по Чёрчу*) *правила выбора получается не случайная по Мизесу – Колмогорову – Лавлэнду последовательность.*

(**г**) *Существует случайная по Мизесу – Чёрчу – Дэли последовательность, становящаяся неслучайной по Мизесу – Чёрчу после вычислимой перестановки.*

$\lhd$ (**а**) Рассмотрим вычислимую последовательность рациональных чисел из $(0,1)$, которая сходится к $1/2$, но медленно, например,

$$p_i = \frac{1}{2} - \frac{1}{\sqrt{i+5}}$$

(число 5 добавлено, чтобы p_i были положительны). Рассмотрим (вычислимую) меру μ, соответствующую независимым испытаниям с вероятностями успеха p_i.

Поскольку ряд $\sum(p_i - 1/2)^2$ расходится, никакая случайная по Мартин-Лёфу (относительно меры μ) последовательность ω не будет случайной по Мартин-Лёфу относительно равномерной меры (теорема 171). С другой стороны, любая такая последовательность будет случайной по Мизесу – Колмогорову – Лавлэнду с предельной частотой $1/2$ (теорема 176).

(**б**) Это утверждение можно доказать аналогично предыдущему, только нужно взять последовательность p_i, ещё медленнее стремящуюся к пределу $1/2$. Пусть, например,

$$p_i = \frac{1}{2} - \frac{1}{\log(i+5)}$$

Какова вероятность (по мере μ) события «начальный отрезок длины n содержит меньше нулей, чем единиц»? Другими словам, насколько вероятно, что частота единиц в этом начальном отрезке больше $1/2$? По теореме 172 эта вероятность (обозначим её δ_n) не больше $e^{-n/O(\log^2 n)}$ (граница d в этой теореме есть примерно $n/O(\log n)$, и $d^2/4n = n/O(\log^2 n)$). Ряд $\sum_n \delta_n$ сходится, и поэтому для некоторого N остаточный член ряда меньше единицы, и тем самым вероятность события «все начальные отрезки, имеющие длину N и более, содержат не меньше нулей, чем единиц» положительна. Множество положительной μ-меры обязано содержать хотя бы одну случайную в смысле Мартин-Лёфа (относительно меры μ) последовательность, поэтому существует случайная в смысле Мартин-Лёфа относительно меры μ последовательность, у которой все начальные отрезки, начиная с длины N, содержат не меньше нулей, чем единиц. Легко понять, что заменив первые N членов на нули, мы получим случайную по Мартин-Лёфу относительно меры μ последовательность, у которой *все* начальные отрезки содержат не меньше нулей, чем единиц. Остаётся сослаться на теорему 176.

(в) И здесь нужный пример доставляет последовательность, случайная относительно меры μ, соответствующей независимым испытаниям с вероятностью успеха p_i. Однако p_i будут устроены некоторым более сложным образом. Разобьём всю последовательность на последовательные участки; k-ый по счёту участок состоит из одного члена, для которого вероятность равна $1/2$, и двух кусков одинаковой (и достаточно большой) длины n_k. В первом из них вероятности одинаковы и чуть больше $1/2$, а во втором — одинаковы и чуть меньше $1/2$ (рис. 23).

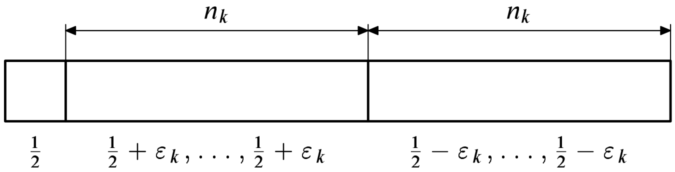

Рис. 23. Участок номер k: вероятности.

[miseslam-1]

При этом ε_k положительно, стремится к нулю с ростом k, но достаточно медленно. Точнее говоря, нам нужно такое соотношение между n_k и ε_k: вероятность того, что среди n_k независимых испытаний, в каждом из которых вероятность единицы равна $\frac{1}{2}+\varepsilon_k$, будет больше единиц, чем нулей, не меньше $1-2^{-(k+3)}$. (Этого несложно добиться для любых ε_k, если взять n_k достаточно большими.)

Если так, то с положительной вероятностью мы получим последовательность, в которой при всех k в каждом из кусков (для каждого k имеется два куска длиной n_k) имеется дисбаланс между нулями и единицами в ожидаемую сторону (в левом куске длиной n_k больше единиц, чем нулей, а в правом — наоборот).

Следовательно, существует случайная по Мартин-Лёфу (относительно построенной нами меры μ) последовательность ω с таким дисбалансом. По теореме 176 она будет случайной по Мизесу – Колмогорову – Лавлэнду.

Покажем теперь, как использовать дисбаланс, чтобы построить (допустимое по Чёрчу) правило выбора, выбирающее из ω неслучайную подпоследовательность ω'. Правило это совсем простое: первый член каждого участка («ключ») выбирается всегда, а в зависимости от его значения (0 или 1) выбирается целиком левый кусок (без правого) или целиком правый (без левого).

Почему ω' не будет случайна по Мизесу – Колмогорову – Лавлэнду? Это совсем просто: условие дисбаланса позволяет восстановить значение ключа по числу единиц в следующем за нём куске длиной n_k. (Мы сначала читаем «в порядке информации» n_k членов, идущих за ключом, а потом отгадываем значение ключа.)

Это доказывает утверждение (в). Попутно можно заметить, что последовательность ω' случайна по Мизесу – Чёрчу – Дэли (поскольку двукратное применение допустимого по Чёрчу – Дэли правила сводится к однократному). С другой стороны, если мы подвергнем ω' вычислимой перестановке членов, переместив ключ в позицию после управляемого им участка, то получится последовательность, не случайная по Мизесу – Чёрчу. Утверждение (г) доказано. ▷

Заметим, что эта теорема в значительной степени дискредитирует наши уточнения случайности в стиле Мизеса. Дело в том (такое свойство коллективов отмечал ещё сам Мизес), что применение допустимого правила выбора к коллективу должно давать коллектив. Утверждение (в) показывает, что определение Мизеса – Колмогорова – Лавлэнда этим свойством, увы, не обладает. Определения с монотонными правилами (Мизеса – Чёрча и Мизеса – Чёрча – Дэли) этим свойством обладают (двукратное применение монотонных правил сводится к однократному), но зато неустойчивы относительно вычислимых перестановок, что тоже нехорошо.

206 Докажите, что применение допустимого по Мизесу – Колмогорову – Лавлэнду правила к случайной по Мизесу – Колмогорову – Лавлэнду последовательности даёт случайную по Мизесу – Чёрчу – Дэли последовательность.

207 Докажите, что (неожиданным образом) определение случайности по Мизесу – Колмогорову – Лавлэнду не изменится, если требовать, чтобы функции F и G были всюду определёнными вычислимыми функциями. [Указание: про это говорил Меркле, надо запрашивать какие-то фиктивные значения, но как? вспомнить!]

[Может быть, здесь нарисовать схему всех определённых к этому моменту классов и включений? Можно ещё добавить случайность по Шнорру.]

[Конечно! — В]

10. Неравенства для энтропии, сложности и размера

[combin]

10.1. Постановка задачи и результаты

[ineq-introduction]

Первой публикацией Колмогорова, где давалось определение сложности конечного объекта, была статья «Три подхода к определению понятия „количество информации“» [24]. Эти три подхода назывались там комбинаторным, вероятностным и алгоритмическим.

При *алгоритмическом* подходе количество информации в сообщении измеряется его колмогоровской сложностью (как мы говорим теперь; естественно, в оригинальной статье такого названия не было). При *вероятностном* подходе сообщение рассматривается как одно из значений случайной величины, и количество информации в нём определяется как шенноновская энтропия этой величины. Но самым первым упоминался *комбинаторный* подход, основанный на таком тривиальном наблюдении: если имеется N различных сообщений, то нужно предусмотреть $\log N$ битов, чтобы их закодировать. (Вариант: если требуется отгадать один из N объектов, задавая да-нет-вопросы, то нужно $\log N$ вопросов.)

Мы уже приводили некоторые результаты о связи этих трёх подходов. Например, теорема 8 (с. 25) связывает комбинаторный и алгоритмический подходы, будучи уточнением такого (абсурдного при буквальном понимании) утверждения: «слово x имеет сложность меньше n, если и только если оно принадлежит множеству из менее чем 2^n элементов». С другой стороны, результаты раздела 7.3 связывают колмогоровскую сложность и шенноновскую энтропию.

В этой главе мы хотим установить более формальные связи между тремя подходами к определению количества информации, ограничившись достаточно узким классом утверждений: линейными неравенствами для энтропии и сложности (и соответствующими им комбинаторными утверждениями).

Пусть $x_1, \ldots, x_n$ — двоичные слова. Для каждого непустого множества индексов $I \subset \{1, 2, \ldots, n\}$ рассмотрим набор слов с индексами из I, который мы будем обозначать x_I, и его колмогоровскую сложность. Например, при $n = 3$ имеется 7 таких наборов и, соответственно, 7 сложностей:

$$\mathit{KS}(x_1), \mathit{KS}(x_2), \mathit{KS}(x_3), \mathit{KS}(x_1, x_2), \mathit{KS}(x_1, x_3), \mathit{KS}(x_2, x_3), \mathit{KS}(x_1, x_2, x_3).$$

Несколько примеров линейных неравенств, их связывающих:

- $\mathit{KS}(x_1, x_2) \leqslant \mathit{KS}(x_1) + \mathit{KS}(x_2) + O(\log N)$;
- $\mathit{KS}(x_1, x_2, x_3) \leqslant \mathit{KS}(x_1) + \mathit{KS}(x_2, x_3) + O(\log N)$;
- $\mathit{KS}(x_1, x_2, x_3) + \mathit{KS}(x_1) \leqslant \mathit{KS}(x_1, x_2) + \mathit{KS}(x_1, x_3) + O(\log N)$;
- $2\,\mathit{KS}(x_1, x_2, x_3) \leqslant \mathit{KS}(x_1, x_3) + \mathit{KS}(x_2, x_3) + \mathit{KS}(x_1, x_2) + O(\log N)$

(мы предполагаем, что x_1, x_2, x_3 — слова длины не больше N).

Общий вид линейного неравенства для сложностей:

$$\sum_I \lambda_I\,\mathit{KS}(x_I) \leqslant O(\log N)$$

(суммирование по всем непустым подмножествам множества $\{1,\ldots,n\}$, коэффициенты λ_I могут быть любого знака; предполагается, что все слова имеют длину не больше N).

Нас интересует, какие неравенства такого вида верны, то есть, говоря более формально, при каких наборах действительных чисел λ_I найдётся такое число c, что

$$\sum_I \lambda_I\,\mathit{KS}(x_I) \leqslant c\log N$$

для любого N и для любых слов $x_1,\ldots,x_n$ длины не больше N.

Ответ на этот вопрос неизвестен, есть лишь некоторые частичные результаты.

Первый из них говорит, что неравенство для сложностей верно тогда и только тогда, когда верно неравенство для шенноновских энтропий с теми же коэффициентами, которое получается, если вместо слов $x_1,\ldots,x_n$ рассматривать случайные величины $\xi_1,\ldots,\xi_n$ с произвольным совместным распределением:

$$\sum_I \lambda_I H(\xi_I) \leqslant 0.$$

Здесь ξ_I — случайная величина, составленная из величин ξ_i при $i \in I$, или, другими словами, проекция случайного вектора $\langle\xi_1,\ldots,\xi_n\rangle$ на I-координаты.

В одну сторону это сразу же вытекает из результатов раздела 7.3: теорема 123 (с. 199) говорит, что энтропия есть математическое ожидание сложности, и если линейное неравенство верно для сложностей, то оно верно и для их математических ожиданий (с нулём в правой части, поскольку при $N\to\infty$ отношение $O(\log N)/N$ стремится к нулю).

Более подробно. Пусть величина ξ_i принимает значения в множестве X_i. Тогда значение $\xi = \langle\xi_1,\ldots,\xi_n\rangle$ можно изобразить в виде столбца высоты n. Готовясь применить теорему 123, для произвольного N рассмотрим N независимых величин, имеющих то же распределение, что и ξ. Вместе они образуют величину ξ^N, значениями которой являются прямоугольные таблицы шириной N и высотой n. По теореме 123 математическое ожидание сложности такой таблицы есть $NH(\xi)+O(\log N)$ (в этой теореме речь шла о префиксной сложности при условии N, но с точностью до $O(\log N)$ это не играет роли).

Такую таблицу можно также читать по строкам, рассматривая её как набор из n строк, каждая из которых есть слово длины N в соответствующем алфавите. При этом теорему 123 не обязательно применять ко всем строкам: можно оставить лишь строки с номерами из некоторого множества $I\subset\{1,2,\ldots,n\}$. Математическое ожидание сложности этой части таблицы будет $NH(\xi_I)+O(\log N)$.

Если неравенство

$$\sum_I \lambda_I\,\mathit{KS}(x_I) \leqslant O(\log N)$$

справедливо для любых слов $x_1,\ldots,x_n$, то оно справедливо и для строк нашей таблицы. Поэтому, переходя к математическим ожиданиям, имеем

$$\sum_I \lambda_I NH(\xi_i) \leqslant O(\log N)$$

Левая часть равна

$$\left(\sum_I \lambda_I H(\xi_i)\right) \cdot N$$

поэтому это возможно лишь при

$$\sum_I \lambda_I H(\xi_i) \leqslant 0,$$

что и требовалось.

Более сложен обратный переход (если неравенство верно для энтропий, то оно верно и для сложностей с логарифмической точностью). Здесь надо, начав с некоторого набора слов $x_1, \ldots, x_n$, построить набор случайных величин $\xi_1, \ldots, \xi_n$ так, чтобы энтропии этих величин и их комбинаций были близки к соответствующим сложностям. Это делается с помощью предложенного А. Ромащенко метода «типизации»: мы рассматриваем множество всех наборов $x'_1, \ldots, x'_n$, которые имеют не бо́льшие сложности и условные сложности, чем $x_1, \ldots, x_n$, и затем рассматриваем случайный элемент этого множества. Подробности см. ниже, в разделе 10.6 (теорема 186).

Дальнейшие результаты так или иначе связаны с комбинаторной интерпретацией неравенств. Для начала рассмотрим простейшее неравенство

$$\mathit{KS}(x_1, x_2) \leqslant \mathit{KS}(x_1) + \mathit{KS}(x_2) + O(\log N)$$

и попытаемся понять, что ему соответствует при комбинаторном подходе. Пусть X_1 и X_2 — конечные множества, из которых берутся сообщения x_1 и x_2, а $A \subset X_1 \times X_2$ — множество возможных пар сообщений. Тогда для пары $\langle x_1, x_2 \rangle$ возможно $|A|$ вариантов (здесь $|A|$ — число элементов в множестве A). Для первой компоненты x_1 число возможностей равно числу элементов в первой проекции множества A (в множестве тех x_1, для которых $\langle x_1, x_2 \rangle \in A$ при некотором $x_2 \in X_2$). Обозначая это число $m(1)$, а размер проекции на вторую ось $m(2)$, можно записать комбинаторный аналог рассматриваемого неравенства:

$$\log |A| \leqslant \log m(1) + \log m(2)$$

или, в мультипликативном варианте,

$$|A| \leqslant m(1)m(2)$$

(размер множества не больше произведения размеров его проекций, что очевидно).

Менее очевидное неравенство получается из другого неравенства для сложностей (теорема 26; аналогичное неравенство для энтропий составляет содержание задачи 172):

$$2\,\mathit{KS}(x_1, x_2, x_3) \leqslant \mathit{KS}(x_1, x_2) + \mathit{KS}(x_1, x_3) + \mathit{KS}(x_2, x_3) + O(\log N)$$

Действуя по аналогии, можно предположить, что для любого подмножества A декартова произведения $X_1 \times X_2 \times X_3$ справедливо неравенство:

$$2 \log |A| \leqslant \log m(1,2) + \log m(1,3) + \log m(2,3)$$

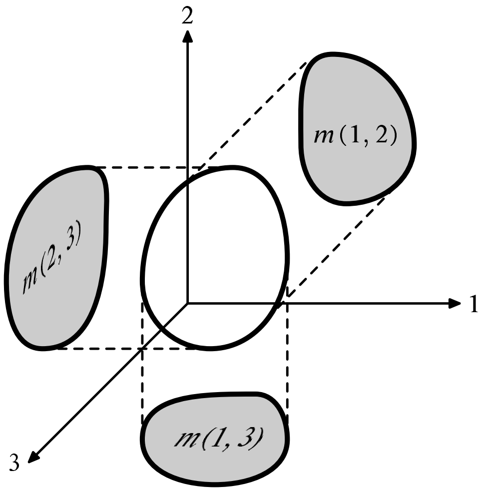

Рис. 24. Три проекции

[ineq.1]

(здесь $m(i,j)$ — число элементов в проекции множества A на оси i и j, рис. 24). В мультипликативной записи:

$$|A|^2 \leqslant m(1,2)m(1,3)m(2,3).$$

И действительно, это не только верно, но может быть выведено из неравенства для сложностей с помощью следующего простого рассуждения. Рассмотрим произвольное натуральное N и множество A^N. Изображая элемент $\langle x_1, x_2, x_3\rangle \in A$ в виде столбца высоты 3, мы представляем элемент A^N как таблицу ширины N и высоты 3. Таких таблиц имеется $|A|^N$, и потому среди них есть таблицы сложности не меньше $\log(|A|^N) = N\log|A|$ (с точностью $O(1)$). Но каждую такую таблицу можно считать тройкой строк $\bar{x}_1, \bar{x}_2, \bar{x}_3$ (каждая из трёх строк имеет длину N) и применить неравенство для сложностей:

$$2\,\mathit{KS}(\bar{x}_1,\bar{x}_2,\bar{x}_3) \leqslant \mathit{KS}(\bar{x}_1,\bar{x}_2) + \mathit{KS}(\bar{x}_1,\bar{x}_3) + \mathit{KS}(\bar{x}_2,\bar{x}_3) + O(\log N).$$

Далее можно оценить каждое слагаемое: например, пара $\langle \bar{x}_1, \bar{x}_2\rangle$, которую можно представить таблицей ширины N и высоты 2, представляет собой набор из N столбцов, каждый из которых принадлежит проекции множества A на первую и вторую координату. Для каждого столбца имеется $m(1,2)$ возможностей, а для всей таблицы $m(1,2)^N$ возможностей, и потому её сложность (при известных N и A) не превосходит $N\log m(1,2) + O(1)$. Множество A не зависит от N, а сложность N есть $O(\log N)$, поэтому в итоге мы получаем

$$2N\log|A| \leqslant N\log m(1,2) + N\log m(1,3) + N\log m(2,3) + O(\log N)$$

что при $N\to\infty$ даёт нам искомое неравенство

$$2\log|A| \leqslant \log m(1,2) + \log m(1,3) + \log m(2,3).$$

208 Докажите то же неравенство, исходя из неравенства

$$2H(\xi_1,\xi_2,\xi_3) \leqslant H(\xi_1,\xi_2) + H(\xi_1,\xi_3) + H(\xi_2,\xi_3).$$

[Указание: рассмотрите тройку случайных величин, равномерно распределённую в множестве A, и воспользуйтесь тем, что энтропия любой случайной величины с k значениями не превосходит $\log k$.]

$\boxed{\textbf{209}}$ Докажите то же неравенство непосредственно, без использования сложностей или энтропий. [Указание. Его можно вывести из неравенства теоремы 139.]

Аналогичное рассуждение можно применить к любому линейному неравенству для сложностей, имеющему в левой части (слева от знака $\leqslant$) лишь одно слагаемое с положительным коэффициентом, а в правой части произвольное количество слагаемых с неотрицательными коэффициентами. При этом можно разрешить в правой части не только безусловные сложности, но и условные сложности. Например, неравенству

$$\mathit{KS}(x_1, x_2) \leqslant \mathit{KS}(x_1) + \mathit{KS}(x_2|x_1)$$

соответствует (очевидное) неравенство

$$m(1,2) \leqslant m(1) \cdot m(2|1),$$

выполненное для произвольного множества $A \subset X_1 \times X_2$, если под $m(1,2)$ понимать число элементов в A, под $m(1)$ понимать число элементов в проекции A на первую координату, а под $m(2|1)$ понимать максимальный размер сечения множества A, получаемого фиксацией первой координаты.

[Объясним, почему именно это неравенство естественно считать комбинаторным аналогом неравенства для сложностей. «Комбинаторное количество информации» в x_1 есть $\log m(1)$; при фиксированном x_1 у нас имеется не более $m(2|1)$ возможных сообщений x_2, и потому количество информации в x_2 при известном x_1 (согласно комбинаторному подходу) не превосходит $\log m(2|1)$. А количество информации в паре $\langle x_1, x_2\rangle \in A$ считается равным $\log A = \log m(1,2)$.]

$\boxed{\textbf{210}}$ Покажите, что любому линейному неравенству с неотрицательными коэффициентами, справедливому для сложностей (безусловных и условных), в котором в левой части стоит только одно слагаемое, соответствует (описанным образом) истинное комбинаторное неравенство.

Неравенства такого вида, в которые не входят условные сложности, можно полностью описать. Рассмотрим неравенство

$$\mathit{KS}(x_1, \ldots, x_n) \leqslant \sum_I \lambda_I \,\mathit{KS}(x_I) + O(\log N)$$

где в правой части все коэффициенты неотрицательны, а сумма берётся по непустым множествам, не совпадающим с полным множеством индексов $\{1, 2, \ldots, n\}$. (Ясно, что только такие неравенства представляют интерес: если в левой части отсутствует некоторое слово, то это слово можно удалить и из правой части: при его замене на пустое слово сложность правой части только уменьшится.)

Theorem 178. *Это неравенство выполнено тогда и только тогда, когда для любого $i \in \{1, 2, \ldots, n\}$ сумма коэффициентов в правой части при членах, содержащих i, не меньше единицы.*

◁ Если сумма коэффициентов при x_i меньше единицы, то неравенство не выполняется даже в случае, когда все остальные слова (кроме x_i) пусты.

С другой стороны, пусть при всех i сумма коэффициентов в правой части не меньше единицы. Разложим каждое слагаемое в сумму, заменив, скажем,

$$KS(x_1, x_2, x_3, \ldots, x_n)$$

на

$$KS(x_1) + KS(x_2|x_1) + KS(x_3|x_1, x_2) + \ldots + KS(x_n|x_1, \ldots, x_{n-1})$$

при этом во всех случаях будем использовать один и тот же порядок (возрастание индексов). Посмотрим отдельно на члены вида $KS(x_i|\ldots)$ с различными условиями. В левой части в качестве условия используются все предыдущие слова $x_1, \ldots, x_{i-1}$, а в правой части могут быть разные подмножества этого условия, но от уменьшения условия сложность лишь возрастает. Остаётся вспомнить, что по предположению сумма коэффициентов не меньше единицы. ▷

$\boxed{\textbf{211}}$ Покажите, что для префиксной сложности неравенства, о которых идёт речь в только что доказанной теореме, выполнены с точностью до $O(1)$ (без логарифма длин). [Указание. Приведённое рассуждение показывает, что это неравенство является линейной комбинацией базисных неравенств, которые верны для префиксной сложности с точностью до $O(1)$ (теорема 64, с. 109). В самом деле, если временно положить $KP(y|x)$ равным $KP(x, y) - KP(x)$, то неравенство $KP(x|y, z) \leqslant KP(y)$ сводится к базисному неравенству.]

Нам, однако, хочется понять комбинаторный смысл произвольных линейных неравенств для сложностей (энтропий), а не только тех, у которых лишь один член в левой части. Тут мы сталкиваемся с некоторой трудностью.

Посмотрим на базисное неравенство

$$KS(x_1) + KS(x_1, x_2, x_3) \leqslant KS(x_1, x_2) + KS(x_1, x_3) + O(\log N).$$

По аналогии с предыдущим можно было бы предположить, что для произвольного множества $A \subset X_1 \times X_2 \times X_3$ выполнено неравенство

$$m(1) \cdot m(1, 2, 3) \leqslant m(1, 2) \cdot m(1.3)$$

Однако это не так. Это неравенство верно (и обращается в равенство) для любого «параллелепипеда» $a \times b \times c$: в этом случае $m(1) = a$, $m(1, 2, 3) = abc$, $m(1, 2) = ab$ и $m(1, 3) = ac$. Но если мы добавим к параллелепипеду $a \times b \times c$ с бо́льшими a, b, c ещё и параллелепипед $a' \times 1 \times 1$, взяв a' много больше a, но много меньше ab и ac, то от такого добавления $m(1, 2)$, $m(1, 3)$ и $m(1, 2, 3)$ изменятся мало, но $m(1)$ сильно возрастёт, и неравенство нарушится.

Другой пример: рассмотрим обратное неравенство для сложности пары:

$$KS(x_1) + KS(x_2|x_1) \leqslant KS(x_1, x_2) + O(1).$$

Как перевести его на комбинаторный язык? Неравенство

$$m(1) \cdot m(2|1) \leqslant m(1, 2),$$

(которое могло бы быть переводом) неверно: $m(1,2)/m(1)$ есть *средний* размер (непустого) сечения, и этот средний размер может быть существенно меньше *максимального*, который мы обозначаем через $m(2|1)$.

В чём тут дело и каков выход из положения? Есть несколько вариантов. Можно ограничиться некоторыми специальными множествами (однородными или почти однородными), для которых этой проблемы не возникает. Другой вариант состоит в том, чтобы попытаться лучше понять, какое именно комбинаторное утверждение соответствует неравенству. Оба подхода будут рассмотрены ниже. Начнём с первого — однородных множеств.

10.2. Однородные множества

Напомним использованные нами обозначения. Пусть $A \subset X_1 \times \ldots \times X_n$ — некоторое непустое подмножество декартова произведения n конечных множеств $X_1, \ldots, X_n$. Для каждого множества индексов $I \subset \{1, \ldots, n\}$ можно рассмотреть проекцию A на соответствующие координаты. Она является подмножеством произведения $\prod_{i \in I} X_i$. Число элементов в этой проекции мы будем обозначать $m_A(I)$. Помимо проекций, можно рассматривать и их сечения. Пусть I и J — два непересекающихся множества индексов. Фиксируем произвольным образом I-координаты (выбрав некоторую точку в $\prod_{i \in I} X_i$) и рассмотрим множество всех J-координат точек из A с заданными I-координатами. Таким образом каждой точке множества $\prod_{i \in I} X_i$ соответствует некоторое подмножество произведения $\prod_{j \in J} X_j$. Максимальный размер такого подмножества мы будем обозначать $m_A(J|I)$. (Если множество A ясно из контекста, мы будем опускать индекс A в этих обозначениях.)

Естественно считать, что $m(\varnothing) = 1$, а также $m(\varnothing|J) = 1$ при любом J. С другой стороны, $m(I|\varnothing)$ естественно положить равным $m(I)$.

Пусть, например, имеется некоторое множество $A \subset X_1 \times X_2$ (рис. 25).

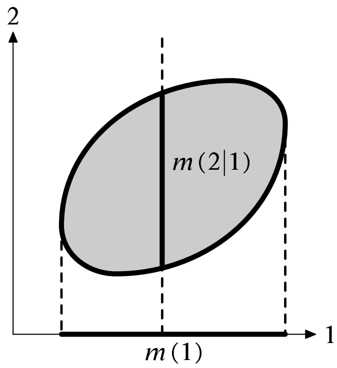

Рис. 25. Плоское множество и его характеристики

[ineq.2]

Тогда $m_A(\{1\})$ — число элементов в проекции множества A на горизонтальную ось, $m_A(\{2\}$ — число элементов в проекции на вертикальную ось, $m_A(\{2\}|\{1\})$ — максимальное число элементов в вертикальных сечениях, а $m_A(\{1\}|\{2\})$ — в горизонтальных. Общее число элементов в множестве есть $m_A(\{1,2\})$.

Имеет место очевидное неравенство:

$$m(1,2) \leqslant m(1) \cdot m(2|1)$$

(для краткости мы опускаем индекс A и фигурные скобки в множествах индексов). В самом деле, каждое из $m(1)$ вертикальных сечений содержит не более $m(2|1)$ элементов.

Для n-мерного множества аналогичное неравенство выглядит так:

$$m(1,2,\ldots,n) \leqslant m(1) \cdot m(2|1) \cdot m(3|1,2) \cdot \ldots \cdot m(n|1,2,\ldots,n-1).$$

В самом деле, для каждого из $m(1)$ возможных значений первой координаты есть не более $m(2|1)$ значений второй, для каждого из которых есть не более $m(3|1,2)$ значений третьей и так далее. Порядок координат роли не играет:

$$m(k_1,\ldots,k_n) \leqslant m(k_1) \cdot m(k_2|k_1) \cdot m(k_3|k_1,k_2) \cdot \ldots \cdot m(k_n|k_1,\ldots,l_{n-1}).$$

для любой перестановки $k_1,k_2,\ldots,k_n$ чисел $1,2,\ldots,n$ (в левой части так или иначе записывается общее число элементов в множестве A).

Будем называть множество A *однородным*, если все эти неравенства (для любой перестановки $k_1,\ldots,k_n$) обращаются в равенства. Простейший пример однородного множества — «параллелепипед», то есть произведение подмножеств $A_i \subset X_i$. Однако бывают и другие однородные множества, Например, шестиэлементное двумерное множество рис. 26 является однородным (все ненулевые сечения состоят из двух элементов, а проекции на обе оси — из трёх).

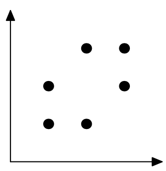

Рис. 26. Однородное множество

[ineq.3]

Пусть I,J,K — непересекающиеся множества индексов. Тогда для произвольного множества A выполнено неравенство

$$m(J\cup K|I) \leqslant m(J|I) \cdot m(K|I\cup J)$$

(при фиксированных I-координатах есть не более $m(J|I)$ возможных значений J-координат, для каждого из которых есть не более $m(K|I\cup J)$ значений K-координат).

С помощью этого неравенства можно доказать ранее упоминавшее неравенство

$$m(k_1,\ldots,k_n) \leqslant m(k_1) \cdot m(k_2|k_1) \cdot m(k_3|k_1,k_2) \cdot \ldots \cdot m(k_n|k_1,\ldots,l_{n-1}),$$

группируя сомножители в правой части. Например, произведение

$$m(k_3|k_1,k_2) \cdot m(k_4|k_1,k_2,k_3)$$

не меньше

$$m(k_3, k_4 | k_1, k_2),$$

после чего произведение

$$m(k_2 | k_1) \cdot m(k_3, k_4 | k_1, k_2),$$

можно (не увеличивая) заменить на

$$m(k_2, k_3, k_4 | k_1)$$

и так далее, пока не получится левая часть. Для однородного множества все эти неравенства обращаются в равенства (так как два крайних члена в цепочке неравенств равны). Из этого рассуждения видно, что для однородных множеств неравенство

$$m(J \cup K | I) \leqslant m(J | I) \cdot m(K | I \cup J)$$

обращается в равенство для любых I, J, K, поскольку можно подобрать цепочку неравенств, в которой оно встречается. Равенство

$$m(J \cup K | I) = m(J | I) \cdot m(K | I \cup J)$$

можно считать определением однородных множеств (требуя его выполнения для любых непересекающихся множеств индексов I, J, K).

212 Докажите, что это свойство действительно равносильно однородности.

213 Докажите, что проекция однородного множества на любое множество индексов является однородным множеством.

214 Докажите, что сечение однородного множества (мы фиксируем одну координату и рассматриваем множество возможных значений остальных координат) является однородным множеством.

Однородные множества важны как источники случайных величин. Пусть $A \subset X_1 \times \ldots \times X_n$ — произвольное множество. Рассмотрим случайную точку в A, принимающую все значения в A с равной вероятностью. Её проекция на i-ю координату есть случайная величина ξ_i со значениями в X_i.

Theorem 179. *Множество A однородно тогда и только тогда, когда для любого $I = \{i_1, \ldots, i_k\}$ величина $\xi_I = \langle \xi_{i_1}, \ldots, \xi_{i_k} \rangle$ принимает все свои значения с равными вероятностями.*

$\triangleleft$ Пусть I — некоторое множество индексов, а J — его дополнение до $\{1, 2, \ldots, n\}$. Тогда равенство

$$m(1, 2, \ldots, n) = m(I) \cdot m(J | I)$$

означает, что средний размер (непустого) сечения, получающегося фиксацией I-координат, то есть $m(1, 2, \ldots, n)/m(I)$, равен его максимальному размеру $m(J|I)$, то есть что все сечения одинаковы. А это и значит, что все значения случайной величины ξ_I равновероятны.

Напротив, пусть для некоторого множества A для любого множества индексов I все значения случайной величины ξ_I равновероятны. В частности, при $I = \{1, \ldots, n-1\}$ мы получаем, что все (непустые) сечения, получаемые фиксацией первых $n-1$ координат, имеют одинаковый размер, и потому

$$m(1, 2, \ldots, n) = m(n|1, 2, \ldots, n-1) \cdot m(1, 2, \ldots, n-1) \qquad (*)$$

Кроме того, поскольку совместное распределение величин $\xi_1, \xi_2, \ldots, \xi_{n-1}$ равномерно в проекции множества A на координаты $1, 2, \ldots, n-1$, то в этой проекции мы имеем ту же картину: $\xi_1, \ldots, \xi_{n-1}$ суть случайные величины, получающиеся проектированием на разные оси случайной точки некоторого множества, равномерно распределённой в этом множестве. Рассуждая по индукции, можно предполагать, что это множество однородно, и тогда равенство $(*)$ можно продолжить:

$$m(1, 2, \ldots, n) = m(n|1, 2, \ldots, n-1) \cdot m(n-1|1, 2, \ldots, n-2) \cdot \ldots \cdot m(3|1, 2) \cdot m(2|1) \cdot m(1)$$

Поскольку порядок координат мог быть любым, мы заключаем, что множество A является однородным в смысле нашего определения. $\triangleright$

Следствие Для построенных таким образом величин $\xi_1, \ldots, \xi_n$ энтропия набора ξ_I равна $\log m(I)$ (при любом $I \subset \{1, 2, \ldots, n\}$).

Отсюда мы заключаем, что

Theorem 180. `[sets-to-variables]` *Каждому неравенству для энтропий случайных величин соответствует неравенство, выполненное для размеров проекций однородных множеств.*

Например, для однородного множества $A \subset X_1 \times X_2 \times X_3$ выполнено неравенство

$$m(1) \cdot m(1, 2, 3) \leqslant m(1, 2) \cdot m(1, 3),$$

соответствующее базисному неравенству для сложностей (теорема 24) и, вообще говоря, неверное для произвольных (не однородных) множеств.

В следующем разделе мы докажем обратное к теореме 180 утверждение, для чего по данному набору случайных величин построим однородное множество, имеющее соответствующие размеры проекций.

10.3. Построение однородного множества

Пусть имеется некоторый набор случайных величин $\eta_1, \ldots, \eta_n$ с конечным числом значений. Мы хотим построить однородное множество A, для которого размеры проекций соответствуют энтропиям величин η_I. Идеально было бы, чтобы

$$\log m_A(I) = H(\eta_I)$$

для всех $I \subset \{1, \ldots, n\}$. Тогда можно было бы сразу заключить, что неравенство, верное для (логарифмов) размеров проекций однородных множеств, верно и для любых случайных величин.

Однако это, как легко понять, невозможно: энтропия вовсе не обязана быть логарифмом целого числа. Но если нас интересуют линейные неравенства, достаточно, чтобы энтропии были бы *пропорциональны* логарифмам проекций, и даже чтобы они были *приблизительно* (с малой ошибкой) пропорциональны им. Сейчас мы построим однородное множество для случая, когда вероятности всех значений для набора $\langle\eta_1, \ldots, \eta_n\rangle$ рациональны. Ясно, что для неравенств рассматриваемого нами вида этого достаточно: по непрерывности любое неравенство, выполненное для случайных величин с рациональными вероятностями, верно и для любых случайных величин.

Итак, пусть имеется набор случайных величин $\eta_1, \ldots, \eta_n$. Каждое его значение будем изображать столбцом высоты n, в котором сверху вниз записаны значения величин. Каждый столбец имеет некоторую (по предположению рациональную) вероятность. Приведя все эти вероятности к общему знаменателю N, можно составить таблицу из N столбцов, в которой частота (доля) каждого столбца равна его вероятности как значения случайной величины $\langle\eta_1, \ldots, \eta_n\rangle$.

Такую таблицу можно читать и «по строкам», тогда i-я строка есть слово длины N в алфавите, буквами которого являются возможные значения величины η_i. Множество таких слов мы обозначим X_i (длина N у нас фиксирована, поэтому мы её не указываем в обозначении). А вся таблица есть набор из n слов длины N (каждое в соответствующем алфавите), то есть элемент множества $X_1 \times \ldots \times X_n$.

Рассмотрим теперь всевозможные таблицы, получающиеся из данной перестановкой столбцов. Другими словами, рассмотрим все таблицы ширины N, в которых частоты столбцов такие же, то есть соответствуют распределению вероятности для случайной величины. Получится некоторое подмножество U множества $X_1 \times \ldots \times X_n$. Оно и будет интересующим нас однородным подмножеством.

Чтобы убедиться в этом, заметим, что любой элемент множества U можно получить, применив к исходной таблице перестановку столбцов. Если эту перестановку выбирать случайно, считая все $N!$ перестановок равновероятными, то вероятность получить данный элемент множества U не зависит от выбора этого элемента. (В самом деле, число перестановок, дающих этот элемент, равно числу перестановок, оставляющих его на месте, и определяется лишь количествами равных столбцов в таблице, а не тем, где они расположены.)

Это свойство останется верным, если из таблицы удалить некоторые строки. Следовательно, проекция случайной равномерно распределённой в U точки на любой набор координат также равномерно распределена, так что множество является однородным.

Теперь надо найти размеры проекций. Для начала найдём размер самого множества. Пусть в нём имеется m различных типов столбцов, которые встречаются с частотами $q_1, \ldots, q_m$. Тогда общее число вариантов, которые можно получить перестановками столбцов, равно

$$\frac{N!}{(q_1N)!(q_2N)!\ldots(q_mN)!}$$

и его логарифм по формуле Стирлинга оценивается как

$$Nh(q_1, \ldots, q_m) + O(\log N),$$

где $h(q_1, \ldots, q_m) = \sum q_i(-\log q_i)$ — шенноновская энтропия величины, принимающей m значений с вероятностями $q_1, \ldots, q_m$, то есть — в нашем случае — величины $\langle\eta_1, \ldots, \eta_n\rangle$.

Так что логарифм размера множества примерно в N раз больше энтропии набора случайных величин. То же самое рассуждение применимо и к любой проекции этого множества и показывает, что логарифм размера проекции на множество индексов I примерно в N раз больше энтропии величины η_I.

Если теперь у нас есть линейное неравенство, справедливое для логарифмов проекций однородных множеств, то оно будет справедливо для множества U. Увеличивая N (умножая его на большое целое число) и переходя к пределу, мы видим, что $O(\log N)/N$ стремится к нулю, и заключаем, что то же линейное неравенство верно и для энтропий любых n случайных величин с рациональным распределением вероятностей. По непрерывности это верно и для любого (не обязательно рационального) распределения вероятностей. Приходим к такому утверждению

Theorem 181. [random-to-uniform] *Всякому неравенству для размеров проекций однородных множеств, соответствует неравенство, выполненное для произвольных случайных величин.*

10.4. Однородные множества и орбиты

Возникает вопрос, за счёт чего нам удалось построить однородное множество в предыдущем разделе. Обычно такого рода однородность достигается за счёт какой-либо алгебраической структуры на рассматриваемых объектах. Нетрудно обнаружить такую структуру и в данном случае.

А именно, у нас имеется группа перестановок S_N, которая действует на столбцах таблицы. Однородное множество — это орбита некоторой точки (таблицы с заданными частотами) при этом действии. В общем случае конструкция выглядит следующим образом.

Пусть имеется произвольная конечная группа G и конечные множества $X_1, \ldots, X_n$. Пусть заданы действия группы G на каждом из X_i. Они вместе задают действие G на множестве $X_1 \times \ldots \times X_n$. Рассмотрим произвольную точку $\langle x_1, \ldots, x_n \rangle \in X_1 \times \ldots \times X_n$ и орбиту U этой точки.

Theorem 182. *Орбита U является однородным подмножеством множества $X_1 \times \ldots \times X_n$.*

$\lhd$ Подействуем на точку $x = \langle x_1, \ldots, x_m \rangle$ случайным элементом группы G (считая все элементы равновероятными). Результат будет случайной величиной, значениями которой являются точки из U. При этом все точки равновероятны. В самом деле, элементы группы, переводящие x в заданную точку y, образуют смежный класс при факторизации по стабилизатору точки x (подгруппе, состоящей из тех элементов G, которые оставляют точку x на месте), а все смежные классы имеют один и тот же размер.

Аналогичное утверждение верно и для любой части индексов, поэтому случайный элемент множества U имеет равную вероятность спроектироваться в любую точку проекции, и потому множество U является однородным. $\rhd$

Во что превращаются при этом неравенства для размеров проекций однородных множеств? Размер орбиты U равен отношению размера G и размера стабилизатора точки x. Этот стабилизатор есть пересечение стабилизаторов точек $x_1, \ldots, x_n$ при соответствующих действиях. Аналогичным образом размер проекции на индексы из I равен отношению размера G и размера пересечения стабилизаторов для x_i при $i \in I$. Заметим, что любая подгруппа H группы G может быть стабилизатором некоторого элемента при некотором действии (достаточно рассмотреть действие G на множестве смежных классов по H). Поэтому любое неравенство для размеров проекций однородных множеств превращается в неравенство, справедливое для любых подгрупп произвольной конечной группы.

Например, неравенство $m(1,2) \leqslant m(1) \cdot m(2)$ превращается в неравенство, верное для любых подгрупп H_1 и H_2 произвольной конечной группы G:

$$\frac{|G|}{|H_1 \cap H_2|} \leqslant \frac{|G|}{|H_1|} \cdot \frac{|G|}{|H_2|}$$

или $|H_1 \cap H_2| \geqslant |H_1| \cdot |H_2|/|G|$. А неравенство $m(1,2,3)^2 \leqslant m(1,2)m(1,3)m(2,3)$ даёт неравенство

$$|H_1 \cap H_2 \cap H_3|^2 \geqslant |H_1 \cap H_2| \cdot |H_1 \cap H_3| \cdot |H_2 \cap H_3|/|G|$$

Доказательство теоремы 181 показывает, что верно и обратное: всякому неравенству для размеров пересечений подгрупп (в котором они перемножаются в некоторых степенях) соответствует неравенство для случайных величин (поскольку мы можем приближать случайные величины с помощью орбит), а потому и для проекций однородных множеств. Получаем следующий удивительный результат:

Theorem 183. *Всякому линейному неравенству для энтропий случайных величин соответствует неравенство для размеров подгрупп конечной группы и их пересечений, и наоборот.*

10.5. Почти однородные множества

Мы называли множество $A \subset X_1 \times \ldots \times X_n$ однородным, если неравенство

$$m(k_1,\ldots,k_n) \leqslant m(k_1) \cdot m(k_2|k_1) \cdot m(k_3|k_1,k_2) \cdot \ldots \cdot m(k_n|k_1,\ldots,k_{n-1}).$$

обращается в равенство для любой перестановки $k_1,\ldots,k_n$ чисел $1,\ldots,n$. Будем говорить, что множество c-однородно, если правая часть этого неравенства не более чем в c раз превосходит левую (для любой перестановки).

Таким образом, 1-однородные множества — это однородные множества в смысле старого определения.

Многие свойства однородных множеств переносятся (с «потерей точности») на c-однородные множества.

Theorem 184. `[quasi-uniform]` *Пусть множество A является c-однородным.*

(а) Если I, J, K — непересекающиеся множества индексов, то в неравенстве

$$m(J \cup K|I) \leqslant m(J|I) \cdot m(K|I \cup J)$$

(верном для любого множества A) правая часть превосходит левую не более чем в c раз.

(б) Проекция c-однородного множества на любой набор координат c-однородна.

(в) Пусть A' — подмножество множества A, составляющее в нём долю не менее ε. Тогда множество A' является c/ε-однородным.

(г) Пусть ξ — случайная величина, равномерно распределённая в A. Тогда её проекция ξ_I на любое множество индексов I имеет энтропию не больше $\log m(I)$ и не меньше $\log m(I) - \log c$.

(д) Пусть I и J — непересекающиеся множества индексов. Тогда $H(\xi_J|\xi_I)$ не больше $\log m(J|I)$ и не меньше $\log m(J|I) - \log c$.

$\triangleleft$ (а) В правой части неравенства

$$m(k_1, \ldots, k_n) \leqslant m(k_1) \cdot m(k_2|k_1) \cdot m(k_3|k_1, k_2) \cdot \ldots \cdot m(k_n|k_1, \ldots, k_{n-1}).$$

можно группировать сомножители, пользуясь неравенством

$$m(J|I) \cdot m(K|I \cup J) \geqslant m(J \cup K|I).$$

При этом произведение будет постепенно уменьшаться, пока правая часть не превратится в левую. Если исходное неравенство отличалось от равенства не более чем в c раз, то и на каждом шаге уменьшение будет не более чем в c раз, а порядок действий можно выбрать так, чтобы пройти через любую тройку I, J, K.

(б) По предположению

$$m(n|1, \ldots, n-1) \cdot m(n-1|1, \ldots, n-2) \cdot \ldots \cdot m(2|1) \cdot m(1) \leqslant cm(1, \ldots, n)$$

и это неравенство можно продолжить:

$$cm(1, \ldots, n) \leqslant cm(n|1, \ldots, n-1) \cdot m(1, \ldots, n-1).$$

Сокращая теперь на $m(n|1, \ldots, n-1)$, получаем условие c-однородности для проекции на координаты $1, 2, \ldots, n-1$ (порядок координат может быть любым, а не только $1, 2, \ldots, n$, как в нашем примере).

(в) От перехода к подмножеству максимальные размеры сечений увеличиться не могут, а размер всего множества уменьшается не более чем в $1/\varepsilon$ раз.

(г) Случайная величина ξ_I принимает $m(I)$ значений, поэтому её энтропия не превосходит $\log m(I)$. С другой стороны, если $J = \{1, 2, \ldots, n\} \setminus I$, то

$$m(I) \cdot m(J|I) \leqslant cm(1, 2, \ldots, n)$$

поэтому каждое значение величины ξ_I имеет вероятность не больше $m(J|I)/m(1, 2, \ldots, n) \leqslant c/m(I)$ и потому её энтропия не меньше $\log m(I) - \log c$.

(д) Сравним равенство

$$H(\xi_1, \ldots, \xi_n) = H(\xi_n|\xi_1, \ldots, \xi_{n-1}) + H(\xi_{n-1}|\xi_1, \ldots, \xi_{n-2}) + \ldots + H(\xi_2|\xi_1) + H(\xi_1)$$

с неравенством

$$\log m(1, \ldots n) \leqslant$$
$$\leqslant \log m(n|1, \ldots, n-1) + \log m(n-1|1, \ldots, n-2) + \ldots + \log m(2|1) + \log m(1).$$

Левые части у них совпадают, поскольку ξ равномерно распределена в множестве A. Каждая энтропия в правой части не превосходит соответствующего логарифма (условная энтропия есть математическое ожидание энтропии при данном значении условия, которая не больше логарифма соответствующего сечения). Кроме того, в силу однородности неравенство отличается от равенства не более чем на $\log c$. Отсюда следует, что каждая энтропия может отличаться от соответствующего логарифма не более чем на $\log c$. Переставляя и группируя члены подходящим образом, мы можем тем же способом доказать неравенство

$$\log m(J|I) - \log c \leqslant H(\xi_J|\xi_I) \leqslant \log m(J|I)$$

для произвольных непересекающихся множеств I, J. $\triangleright$

Непосредственное следствие утверждений (г) и (д): если некоторое линейное неравенство выполнено для энтропий, то оно выполнено для логарифмов проекций и сечений c-однородных множеств с ошибкой не более $m \log c$, где m — число слагаемых в неравенстве. (Чем больше c, тем менее однородно множество, и тем больше возможная ошибка.)

Это свойство будет использовано нами в следующем разделе, когда мы по набору слов построим почти однородное множество, а затем случайные величины.

10.6. Метод типизации

[romashchenko-theorem]

Следующая теорема по любому набору слов $x_1, \ldots, x_n$ строит почти однородное подмножество A некоторого декартова произведения $X_1 \times X_2 \times \ldots \times X_n$ конечных множеств, для которого $\log m(J|I) \approx \mathit{KS}(x_J|x_I)$ с логарифмической (от сложности слов) погрешностью. Вот точная формулировка.

Theorem 185. [romashchenko-typical] *Для всякого n существует число d, при котором верно следующее утверждение: для любого числа $N > 1$ и любого набора слов $x_1, \ldots, x_n$ сложности не более N найдутся конечные множества $X_1, \ldots, X_n$ и N^d-однородное подмножество $A \subset X_1 \times \ldots \times X_n$, для которых $|\log m(J|I) - \mathit{KS}(x_J|x_I)| \leqslant d \log N$ при любых непересекающихся множествах $I, J \subset \{1, \ldots, n\}$.*

Заметим, что множества X_i в этой теореме не очень по существу: можно говорить о конечном множестве n-ок произвольной природы с нужными размерами проекций. Отметим ещё, что странное условие $N > 1$ объясняется тем, что при $N = 1$ оценка N^c не растёт с ростом c.

В доказательстве используется понятие *сложностного вектора* набора слов. А именно, сложностным вектором слов $x_1, \ldots, x_n$ называется список всех сложностей $\mathit{KS}(x_I|x_J)$ для всех пар (I, J) непересекающихся подмножеств множества индексов $\{1, \ldots, n\}$. (Заметим, что этот вектор имеет экспоненциальное по n число компонент: одних безусловных сложностей (при $J = \varnothing$) уже набирается $2^n - 1$, не считая пустого множества.) Будем обозначать сложностной вектор $\varkappa(x_1, \ldots, x_n)$.

$\triangleleft$ Для каждого набора слов $x_1, \ldots, x_n$ рассмотрим множество $A(x_1, \ldots, x_n)$ всех наборов $y_1, \ldots, y_n$, у которых

$$\varkappa(y_1, \ldots, y_n) \leqslant \varkappa(x_1, \ldots, x_n)$$

при покомпонентном сравнении векторов. При $n=1$ это будет множество всех слов не большей сложности, чем x_1. При $n=2$ мы рассматриваем все пары слов y_1, y_2, для которых

$$KS(y_1) \leqslant K(x_1), \quad KS(y_2) \leqslant K(x_2), \quad KS(y_1,y_2) \leqslant K(x_1,x_2),$$
$$KS(y_1|y_2) \leqslant K(x_1|x_2), \quad KS(y_2|y_1) \leqslant K(x_2|x_1).$$

Множество $A(x_1,\ldots,x_n)$ заведомо непусто, поскольку содержит набор $\langle x_1,\ldots,x_n\rangle$. Мы хотим показать, что оно содержит достаточно много (примерно $2^{KS(x_1,\ldots,x_n)}$) элементов. (Больше оно содержать и не может, поскольку все его элементы по построению имеют сложность не больше $KS(x_1,\ldots,x_n)$.)

В самом деле, зная сложностной вектор $\varkappa(x_1,\ldots,x_n)$, мы можем перечислять множество $A(x_1,\ldots,x_n)$. Задание сложностного вектора требует $O(\log N)$ битов (заметим, что число компонент в этом векторе, хотя и экспоненциально большое, зависит лишь от n и потому может считаться постоянным). Поэтому любой элемент $T(x_1,\ldots,x_n)$ можно задать, указав (помимо сложностного вектора) его порядковый номер в перечислении, и сложность любого элемента в $A(x_1,\ldots,x_n)$ не превосходит

$$\log|A(x_1,\ldots,x_n)| + O(\log N)$$

В частности, это относится и исходному набору $\langle x_1,\ldots,x_n\rangle$, откуда и следует требуемая оценка.

Покажем, что множество $A(x_1,\ldots,x_n)$ является c-однородным для некоторой константы c, полиномиально зависящей от N. Для этого надо сравнить обе части неравенства

$$m(1,2,\ldots,n) \leqslant m(1)\cdot m(2|1)\cdot m(3|1,2)\cdot\ldots\cdot m(n|1,2,\ldots,n-1)$$

Логарифмы сомножителей в правой части можно оценить сверху соответствующими сложностями: $m(1) \leqslant 2^{KS(x_1)}$, поскольку по построению $KS(y_1) \leqslant KS(x_1)$ для любого набора $\langle y_1,\ldots,y_n\rangle \in A(x_1,\ldots,x_n)$ (формально говоря, следовало бы написать $2^{KS(x_1)+1}$ в оценке, но с нашей логарифмической точностью это не имеет значения). Аналогично $m(2|1) \leqslant 2^{KS(x_2|x_1)}$ и так далее. В итоге получается, что логарифм правой части не превосходит

$$KS(x_1) + KS(x_2|x_1) + KS(x_3|x_1,x_2) + \ldots + KS(x_n|x_1,\ldots,x_{n-1}) + O(\log N)$$

что равно $KS(x_1,\ldots,x_n) + O(\log N)$. Но и для левой части мы знаем, что её логарифм не меньше $KS(x_1,\ldots,x_n) - O(\log N)$, так что разница между логарифмами есть $O(\log N)$, а отношение самих значений ограничено многочленом от N, что и требовалось. Одновременно мы устанавливаем, что

$$KS(x_i|x_1,\ldots,x_{i-1}) = \log m(i|1,\ldots,i-1) + O(\log N),$$

а аналогичное рассуждение с группировкой слагаемых даёт

$$KS(x_J|x_I) = \log m(J|I) + O(\log N)$$

для любых непересекающихся I и J. $\triangleright$

Использованный при доказательстве приём можно назвать *типизацией*: от одного слова мы переходим к целому множеству, в котором это слово является «типичным представителем» (с точки зрения сложностей и мощностей).

Теперь уже легко завершить доказательство обещанной теоремы:

Theorem 186. `[prob-compl]` *Всякое линейное неравенство*

$$\sum_I \lambda_I H(\xi_I) \leqslant 0,$$

выполненное для произвольных случайных величин $\xi_1, \ldots, \xi_n$, *выполнено для колмогоровских сложностей произвольных слов сложности не более* N *с погрешностью* $O(\log N)$*:*

$$\sum_I \lambda_I K(\xi_I) \leqslant O(\log N).$$

Заметим, что константа в $O(\log N)$ зависит от n (и быстро растёт с ростом n), но не от слов $x_1, \ldots, x_n$.

$\triangleleft$ В одну сторону (если неравенство верно для сложностей, то оно же верно и для энтропий) мы это уже доказали в разделе 10.1.

Для обратного перехода у нас тоже всё готово. Пусть неравенство верно для энтропий. Рассмотрим произвольные слова $x_1, \ldots, x_n$ и множество $A = A(x_1, \ldots, x_n)$ из только что доказанной теоремы. Рассмотрим случайную величину $\langle \xi_1, \ldots, \xi_n \rangle$, равномерно распределённую в множестве A. Поскольку множество A N^c-однородно, то энтропии будут отличаться от логарифмов размеров сечений не более чем на $O(\log N)$ по теореме 184. С другой стороны, величины $\log m(I|J)$ совпадают с соответствующими сложностями с точностью до $O(\log N)$ по теореме 185, что и завершает доказательство. $\triangleright$

10.7. Комбинаторная интерпретация: примеры

Вернёмся теперь к комбинаторной интерпретации. Напомним основную идею: «слово x имеет сложность меньше n» надо понимать как «слово x принадлежит множеству из менее чем 2^n элементов». (Поскольку сложность определена с точностью до $O(1)$, мы не будем аккуратно различать строгие и нестрогие неравенства.)

Строго говоря, таким образом мы получаем комбинаторную интерпретацию не для функции KS, а для двуместного предиката $KS(x) < n$ с аргументами x и n, и все утверждения о сложности (в частности, неравенства) надо предварительно записать в этих терминах.

Разберём несколько примеров.

- Неравенство $KS(x) \leqslant KS(y)$ записывается так: для всех n из $KS(y) < n$ следует $KS(x) < n$.

- Неравенство $KS(x) \leqslant 2\,KS(y)$ записывается так: для всех n из $KS(y) < n$ следует $KS(x) < 2n$.

- Неравенство $KS(z) \leqslant KS(x) + KS(y)$ записывается так: для всех u и v из $KS(x) < u$ и $KS(y) < v$ следует $KS(z) < u + v$.

Используя последний пример в качестве образца, попытаемся перевести на комбинаторный язык неравенство для сложности пары: $KS(x, y) \leqslant KS(x) + KS(y)$. Первый шаг: если $KS(x) < u$ и $KS(y) < v$, то $KS(x, y) < u + v$. Далее естественно переводить так: если x принадлежит заданному множеству из менее чем 2^u элементов, а y принадлежит заданному множеству из менее чем 2^v элементов, то можно указать множество из менее чем 2^{u+v} элементов, которому принадлежит пара $\langle x, y\rangle$ (а именно, произведение множеств). Так что ничего особенно нового не получается.

Для условной сложности $KS(y|x)$ аналогичный перевод утверждения $KS(y|x) < v$ выглядел бы так: пара $\langle x, y\rangle$ принадлежит известному множеству, у которого все сечения (при фиксированном x) имеют размер менее 2^v. Так что и неравенство $KS(x, y) \leqslant KS(x) + KS(y|x)$ тоже ясно, как переводить.

Ситуация существенно меняется, если заменить знак неравенства. Неравенство $KS(z) \geqslant KS(x) + KS(y)$ можно переписать так: если $KS(x) \geqslant u$ и $KS(y) \geqslant v$, то $KS(z) \geqslant u + v$. Но в нашем предикате KS *меньше* границы, а не больше (и эта асимметрия существенна: мы можем перечислять слова малой сложности, но не слова большой сложности!) Поэтому надо перейти к отрицаниям: если неверно, что $KS(x) < u$ и неверно, что $KS(y) < v$, то неверно, что $KS(z) < u + v$. Другими словами: если $KS(z) < u + v$, то $KS(x) < u$ *или* $KS(y) < v$.

Попробуем применить тот же метод перевода для неравенства

$$KS(x, y) \geqslant KS(x) + KS(y|x),$$

с которым у нас были трудности. Получаем вот что: если пара $\langle x, y\rangle$ принадлежит заданному множеству из менее чем 2^{u+v} элементов, то либо x принадлежит какому-то множеству из менее чем 2^u элементов, либо $\langle x, y\rangle$ принадлежит какому-то множеству, у которого любое сечение содержит не более 2^v элементов.

Более точно: для всякого множества пар A из менее чем 2^{u+v} элементов можно указать

- множество B из менее чем 2^u элементов;
- множество C пар, которое содержит менее 2^v элементов с одним и тем же первым членом,

причём так, что для всякого $\langle x, y\rangle \in A$ выполнено по крайней мере одно из двух: либо $x \in B$, либо $\langle x, y\rangle \in C$.

Если теперь вспомнить доказательство теоремы о сложности пары, то там как раз по существу использовались такие множества: для данных x, y мы смотрели, много ли пар с тем же x имеют малую сложность. Если немного, то $KS(y|x)$ оказывалось малым (в наших теперешних обозначениях: такие пары входят в C), а если много, то тогда $KS(x)$ мало, поскольку это бывает для немногих x. (В наших теперешних обозначениях такие x попадают в B).

$\boxed{\textbf{215}}$ Повторите это рассуждение и формально докажите сформулированное комбинаторное утверждение.

Теперь рассмотрим неравенство, в котором и в левой, и в правой части несколько слагаемых:

$$KS(x_1) + KS(x_1, x_2, x_3) \leqslant KS(x_1, x_2) + KS(x_2, x_3)$$

(базисное неравенство, выполненное с точностью до $O(\log N)$ для слов сложности не выше N).

Переходя к двуместному предикату $KS(x) < n$, это неравенство можно переписать так:

> если $KS(x_1, x_2) < a$, $KS(x_1, x_3) < b$ и $a + b = p + q$, то выполнено по крайней мере одно из неравенств $KS(x_1) < p$ и $KS(x_1, x_2, x_3) < q$.

Естественно предположить, что соответствующее комбинаторное утверждение выглядит так:

> если $A \subset X_1 \times X_2 \times X_3$, $m_A(1,2) \leqslant 2^a$, $m_A(1,3) \leqslant 2^b$ и $a + b = p + q$, то существуют такие $B, C \subset X_1 \times X_2 \times X_3$, что $A \subset B \cup C$, $m_B(1) \leqslant 2^p$ и $m_C(1,2,3) \leqslant 2^q$.

Или, переходя к мультипликативной записи и исключая лишние переменные:

> если $A \subset X_1 \times X_2 \times X_3$ и
>
> $$m_A(1,2) \cdot m_A(1,3) = l \cdot V$$
>
> для некоторых чисел $l, V > 0$, то множество A можно покрыть множествами B и C, для которых $m_B(1) \leqslant l$ и $m_C(1,2,3) \leqslant V$.

Геометрически это можно прочесть так: если у множества A проекции на плоскости $1,2$ и $1,3$ малы, то это ещё не мешает его длине в направлении 1 (проекции на первую координату) и объёму (общему числу точек в A) быть большими. Однако A можно разбить на две части B и C, у первой из которых мала длина по первой координате, а у второго мал объём.

(Вспоминая пример с двумя параллелепипедами, мы видим, что в том случае толстый параллелепипед мог бы составить B, а тонкий — C.)

Всё это не более чем аналогии, которые не заменяют доказательств. Но последнее утверждение действительно оказывается верным, хотя и не вполне тривиальным. Вот как его можно доказать.

Рассмотрим проекцию множества A на плоскость 1,2; она является некоторым подмножеством $X_1 \times X_2$, которое мы обозначим A_{12}. Для каждого $x \in X_1$ рассмотрим сечение этой проекции; пусть это сечение содержит $n_2(x)$ элементов. Тогда

$$m(1,2) = |A_{12}| = \sum_x n_2(x)$$

(размер множества равен сумме всех его сечений). Аналогичным образом

$$m(1,3) = |A_{13}| = \sum_x n_3(x)$$

Длина $m_A(1)$ есть число ненулевых слагаемых в этих суммах, а $m(1,2,3)$ (общее число элементов в A) можно оценить сверху:

$$m(1,2,3) = |A| \leqslant \sum_x n_2(x) n_3(x).$$

(Без ограничения общности можно считать, что это неравенство обращается в равенство, добавив в A недостающие элементы без изменения его проекций.)

Нам надо разбить A на части B и C, при этом у нас есть ограничения на длину B (в направлении первой координаты) и объём C. Естественно, что при данной длине B нужно забрать в это множество как можно больше точек, поэтому мы включим в B самые большие сечения (у которых $n_2(x)n_3(x)$ максимально) в количестве l штук, а остальное отправим в C. После этого надо доказать лишь, что число элементов в C не больше $|A_{12}| \cdot |A_{13}|/l$.

Как это сделать? Для множества C мы можем оценить размеры двух его плоских проекций; они не превосходят $|A_{12}|$ и $|A_{13}|$. Мы знаем также, что все сечения множества C (при любом $x \in X_1$) по площади не превосходят S_l, где S_l — l-е по величине (в порядке убывания) сечение множества A. Рассмотрим неравенство

$$2\,\mathit{KS}(x_1, x_2, x_3) \leqslant \mathit{KS}(x_1, x_2) + \mathit{KS}(x_1, x_3) + \mathit{KS}(x_2, x_3|x_1)$$

(которое легко доказать, перейдя к условным сложностям относительно x_1). Оно содержит в левой части только один член, поэтому мы уже знаем, что из него вытекает комбинаторное утверждение

$$m(1,2,3)^2 \leqslant m(1,2) \cdot m(1,3) \cdot m(2,3|1)$$

и потому

$$|C|^2 \leqslant |A_{12}| \cdot |A_{13}| \cdot S_l$$

Осталось, таким образом, доказать, что

$$S_l \leqslant \frac{|A_{12}| \cdot |A_{13}|}{l^2}$$

Вспомним, что множество B состоит из l прямоугольников, каждый из которых имеет площадь не меньше S_l. Сумма «оснований» $n_2(x)$ этих прямоугольников не превосходит $|A_{12}|$, а сумма их «высот» $n_3(x)$ не превосходит $|A_{13}|$, так что среднее (арифметическое) основание не больше $|A_{12}|/l$, а средняя (арифметическая) высота не больше $|A_{13}|/l$. Остаётся заметить, что если $S \leqslant a_i b_i$ при всех $i = 1, 2, \ldots, l$, то

$$S \leqslant \frac{a_1 + \ldots + a_l}{l} \cdot \frac{b_1 + \ldots + b_l}{l}$$

(что легко следует из выпуклости логарифма).

Таким образом, в нашем случае сформулированное по аналогии с базисным неравенством утверждение о размерах множеств действительно оказалось верным.

10.8. Комбинаторная интерпретация: общий случай

Переходя от примеров к общему утверждению, рассмотрим произвольное линейное неравенство для сложностей и разделим его на две части с положительными коэффициентами

$$\sum \lambda_I\,\mathit{KS}(x_I) \leqslant \sum \mu_J\,\mathit{KS}(x_J)$$

(суммы в левой и правой части берутся по непересекающимся множествам индексов $I, J \subset \{1, 2, \ldots, n\}$; никакое множество индексов не входит одновременно в обе части; коэффициенты λ_I и μ_J положительны).

Как выглядит соответствующее комбинаторное утверждение? По аналогии с рассмотренными примерами его можно сформулировать так:

Пусть $A \subset X_1 \times \ldots \times X_n$ и пусть n_I — произвольный набор чисел, для которого

$$\prod_I (n_I)^{\lambda_I} = \prod_J m_A(J)^{\mu_J}$$

Тогда множество A можно покрыть множествами B_I, для которых

$$m_{B_I}(I) \leqslant n_I$$

К сожалению, про сформулированное таким образом утверждение не удаётся доказать, что оно выполнено одновременно с соответствующим неравенством для сложностей. Нам придётся несколько ослабить утверждение, включив в него множитель, соответствующий $O(\log N)$ в неравенстве. Вот этот ослабленный вариант:

> Существует такая константа d, что для произвольных $X_1, \ldots, X_n$, для произвольного $A \subset X_1 \times \ldots \times X_n$ и для произвольного набора чисел n_I, для которого
>
> $$\prod_I (n_I)^{\lambda_I} = \prod_J m_A(J)^{\mu_J},$$
>
> существует покрытие множества A множествами B_I, для которых
>
> $$m_{B_I}(I) \leqslant n_I \cdot (\log |A|)^d$$

Theorem 187. [combinatorial1] *Сформулированное утверждение верно для данных наборов λ_I и μ_J тогда и только тогда, когда*

$$\sum \lambda_I \, \mathit{KS}(x_I) \leqslant \sum \mu_J \, \mathit{KS}(x_J) + O(\log N)$$

для любых слов $x_1, \ldots, x_n$ сложности не больше N.

$\lhd$ Предположим, что неравенство верно и покажем, как можно покрыть произвольное множество A частями требуемого размера. Прежде всего заметим, что без ограничения общности можно считать элементы множества A наборами $\langle x_1, \ldots, x_n \rangle$ слов длины не больше $N = \log |A|$ (слов хватит, а конкретная природа элементов роли не играет).

Предположим временно, что множество A является простым (имеет сложность $O(\log N)$). В этом случае простыми являются и все его проекции, и потому сложности их элементов не превосходят логарифма размера (с точностью до $O(\log N)$). Поэтому для произвольного элемента $\langle x_1, \ldots, x_n \rangle \in A$ выполнены неравенства

$$K(x_J) \leqslant \log m_A(J) + O(\log N)$$

для всех J; складывая их с коэффициентами μ_J, получаем, что

$$\sum_J \mu_J K(x_J) \leqslant \log \left(\prod_J m_A(J)^{\mu_J} \right) + O(\log N)$$

Пользуясь неравенством (которое мы считаем верным) и вспоминая свойство чисел n_I, заключаем, что

$$\sum_I \lambda_I K(x_I) \leqslant \log\left(\prod_I n_I^{\lambda_I}\right) + O(\log N) = \sum_I \lambda_I \log n_I + O(\log N)$$

для любого элемента $\langle x_1, \ldots, x_n \rangle \in A$. А отсюда следует, что (для каждого элемента множества A) хотя бы одно слагаемое в левой части меньше соответствующего слагаемого в правой: для всякого $\langle x_1, \ldots, x_n \rangle$ найдётся I, при котором

$$KS(x_I) \leqslant n_I + O(\log N)$$

то есть x_I принадлежит множеству всех объектов сложности меньше $n_I + O(\log N)$. Соответствующие элементы A и образуют искомое множество B_I.

Это завершает рассуждение для случая простого множества A. Общий случай сводится к этому с помощью стандартного приёма: для каждого N рассмотрим всевозможные множества $A \subset X_1 \times \ldots \times X_n$, где все X_i состоят из слов длины не больше N, и рассмотрим то из них, которое (при данных λ_I и μ_J) покрывается хуже всего (с наибольшей разностью между левой и правой частью для оптимального покрытия). Это множество является простым (сложности $O(\log N)$) поскольку его можно найти перебором, для которого достаточно знать N и коэффициенты λ_I, μ_J, и поэтому к нему применимо изложенное выше рассуждение. Поскольку это множество было самым трудным для покрытия, для остальных множеств утверждение тем более верно.

(Сказанное требует некоторого уточнения. Вообще говоря, мы не предполагаем, что коэффициенты λ_I и μ_J являются рациональными или даже вычислимыми. Но их достаточно рассматривать с точностью до $1/N$, поскольку логарифмы всех мощностей не превосходят N и общая ошибка не больше $O(1)$. А задание их с такой точностью требует $O(\log N)$ битов.)

В одну сторону утверждение теоремы доказано. Осталось показать, что если верно утверждение о покрытии, то верно и неравенство для сложностей. Это делается с помощью того же метода типизации.

Рассмотрим произвольный набор слов $x_1, \ldots, x_n$, каждое из которых имеет сложность не более N. Пусть для него неравенство не выполняется и левая часть заметно (более чем на $O(\log N)$) больше правой. Включим наш набор в почти однородное множество $A = A(x_1, \ldots, x_n)$. Рассмотрим в качестве n_I величины $KS(x_I)$, уменьшенные поровну настолько, чтобы сравнять левую часть неравенства с правой; поскольку $\log m_A(J)$ не превосходит $KS(x_J)$, можно применить комбинаторное утверждение (увеличив n_I не более чем на $O(\log N)$). В результате множество A будет покрыто множествами B_I, у которых I-проекции содержат не более 2^{n_I} элементов, что заметно меньше соответствующей проекции множества A, содержащей примерно $2^{KS(x_I)}$ элементов (вспомним, что мы заметно уменьшали n_I). В силу почти однородности множества A всякая его часть, составляющая небольшую долю в одной из проекций, составляет небольшую долю по отношению ко всему A, и потому фиксированное число множеств B_I не может покрыть множества A. $\triangleright$

$\boxed{\textbf{216}}$ Проведите аккуратно все требуемые в этом доказательстве оценки (скрытые за словами «заметно», «небольшая доля» и т. п.)

$\boxed{217}$ Покажите, что теорема 187 и её доказательство переносятся на неравенства, содержащие не только безусловные, но и условные сложности.

10.9. Комбинаторная интерпретация: другой вариант

Проведённое рассуждение подсказывает другой вариант комбинаторной интерпретации неравенств — менее естественный, но проще формулируемый, поскольку не требуется разбивать неравенство на две части с положительными коэффициентами и по-разному их рассматривать.

А именно, рассмотрим набор чисел λ_I любого знака и следующее комбинаторное утверждение:

> Существует такая константа d, что для произвольных конечных множеств $X_1, \ldots, X_n$ и для произвольного $A \subset X_1 \times \ldots \times X_n$ можно представить A в виде объединения не более чем $(\log |A|)^d$ множеств, для каждого из которых выполнено неравенство
> $$\prod m(I)^{\lambda_I} \leqslant (\log |A|)^d.$$
> (для размеров его проекций).

Theorem 188. *Сформулированное утверждение верно для некоторого набора коэффициентов λ_I тогда и только тогда, когда*

$$\sum \lambda_I \, KS(x_I) \leqslant O(\log N)$$

для любых слов $x_1, \ldots, x_n$ сложности не более N.

$\triangleleft$ Пусть верно комбинаторное утверждение, и мы хотим доказать неравенство для сложностей. Рассмотрим произвольные слова x_i сложности не больше N. Применим метод типизации и получим множество $A = A(x_1, \ldots, x_n)$. При этом $\log |A|$ будет ограничен многочленом от N.

По предположению, множество A можно представить в виде объединения полиномиального (от $\log |A|$, а потому и от N) числа множеств. Рассмотрим то из них, которое содержит больше всего элементов. Обозначим его B. Множество B включает в себя полиномиальную долю элементов множества A, которое является c-однородным с полиномиальным значением c, а потому и B является таковым (с бóльшим, но по-прежнему полиномиальным значением c), причём логарифмы размеров проекций у A и B отличаются на $O(\log N)$. Следовательно, из неравенства для B можно вывести неравенство для A (с точностью $O(\log N)$), а потому, как мы видели, и для сложностей.

В другую сторону доказываемая теорема вытекает из такой леммы:

Лемма. Любое множество $A \subset X_1 \times \ldots \times X_n$ можно представить в виде объединения полиномиального от $N = \log |A|$ числа частей, каждая из которых является c-однородным множеством с полиномиальным (от N) значением c.

Заметим, что в этой лемме ничего не говорится ни о колмогоровской сложности, ни о неравенствах. Но из неё следует требуемое нам утверждение, поскольку для каждой из частей неравенство верно для энтропий, а логарифмы размеров проекций отличаются не более чем на $O(\log c) = O(\log N)$.

Интересно, что (пожалуй, самое) простое доказательство использует колмогоровскую сложность. Без ограничения общности можно предполагать, что элементы $\langle x_1, \ldots, x_n \rangle \in A$ состоят из двоичных слов. Для каждого элемента $x \in A$ рассмотрим множество всех $\langle y_1, \ldots, y_n \rangle \in A$, для которых сложностной вектор, составленный из условных (относительно A) сложностей, не превосходит такого же вектора для x.

Эта конструкция отличается от ранее рассмотренной «типизации» в двух отношениях: во-первых, мы рассматриваем лишь элементы $y \in A$ (а раньше никакого A не было); во-вторых, мы используем сложности при условии A.

Заметим, что полученное множество определяется сложностями слов $x_1, \ldots, x_n$ и их комбинаций, а не самими этими словами, и потому получится лишь полиномиальное число таких множеств. Осталось проверить, что каждое из них является c-однородным для полиномиального (от N) значения c.

Это делается так же, как и раньше: число элементов в таком множестве не сильно меньше $2^{KS(x_1,\ldots,x_n|A)}$, а логарифмы сечений оцениваются условными сложностями, так что эти два изменения оставляют доказательство однородности в силе. $\triangleright$

Интересно найти чисто комбинаторное доказательство только что рассмотренной леммы, не использующее колмогоровскую сложность. Это не вполне тривиально даже для двумерных множеств. Пусть мы имеем некоторое конечное множество $A \subset \mathbb{N}^2$ «Почти однородность» в этом случае означает, что

$$m(1,2) \approx m(1)m(2|1), \quad m(1,2) \approx m(2)m(1|2)$$

Другими словами, среднее (непустое) вертикальное сечение $m(1,2)/m(1)$ должно не слишком сильно отличаться от максимального сечения $m(2|1)$ — и то же самое для горизонтальных сечений.

Естественная идея — разбить множество на части в зависимости от вертикальных сечений, при этом в каждой части размеры сечений отличаются друг от друга не более чем (скажем) вдвое. В этом случае для каждой части максимальное и среднее сечение тоже будет отличаться не более чем вдвое. Проблема в том, что это нужно сделать не только для вертикальных сечений, но и для горизонтальных — после чего вертикальные сечения могут снова испортиться.

Как преодолеть эту трудность? Во-первых, можно заметить, что достаточно найти почти однородное множество, составляющее не очень малую (полиномиальную) долю в исходном множестве. В самом деле, после этого можно повторить то же рассуждение с остатком множества, получив вторую почти однородную часть (непересекающуюся) и так далее: если каждая очередная часть забирает ε-долю элементов, то после $1/\varepsilon$ шагов число оставшихся элементов уменьшится примерно в e раз. Повторяя это полиномиальное число раз, можно добиться, чтобы осталось менее одного элемента (то есть чтобы вообще ничего не осталось).

Как же построить не очень малую часть, почти однородную в обоих направлениях? После деления на части по вертикали выберем наибольшую часть (остальные не используются), поделим её по горизонтали и снова выберем наибольшую часть (выбросив остальные). Нужно только заметить, что она останется почти однородной по вертикали (пункт (в) теоремы 184).

Это рассуждение проходит для любой размерности и имеет два преимущества по сравнению с предыдущим (использующим колмогоровскую сложность). Во-первых, мы получаем непересекающиеся части. Во-вторых, оно даёт части, неоднородность которых есть полином от логарифма мощности не всего множества, а этой части (более сильное условие).

Ещё немного усложнив рассуждение, можно достичь ещё одного улучшения: неоднородность каждой части ограничена константой (зависящей от размерности задачи n, но не от размера разбиваемого на части множества). Вот как это делается.

Для каждого разбиения на части определим его *вес* таким образом, что разбиение наименьшего веса (которое существует, поскольку число разбиений конечно) будет удовлетворять всем нужным условиям.

Вес разбиения будет суммой весов всех элементов в этом разбиении. А вес элемента x, попавшего в часть X, определяется формулой

$$\sum_{A,B} \log m_X(B|A) - d\log|X|,$$

где сумма берётся по парам непересекающихся множеств индексов $A, B \subset \{1, 2, \ldots, n\}$, а d — некоторый постоянный коэффициент (заметим, кстати, что в сумме также есть член $\log|X|$ — при $A = \varnothing$, $B = \{1, 2, \ldots, n\}$). Подчеркнём, что веса всех элементов внутри одной части одинаковы.

Покажем, что если выбрать d достаточно большим, то разбиение наименьшего веса содержит немного частей. А именно, мы покажем, что выгодно соединить две части, у которых все параметры $\log m_X(B|A)$ достаточно близки (отличаются не более чем на единицу). В самом деле, при соединении каждое $m_X(B|A)$ возрастёт не более чем втрое (по сравнению с меньшей частью), а $\log|X|$ возрастёт не менее чем в полтора раза, и при достаточном d это перевесит, и при слиянии частей веса всех элементов уменьшатся. Заметим, что нужное значение d определяется лишь числом слагаемых (которое, в свою очередь, зависит лишь от n и не зависит от размера множества).

Пусть d выбрано таким образом. Будем классифицировать части по целым частям значений $\log m_X(B|A)$ при всех A и B. Как мы видели, никакие две части не попадут в один класс, а число классов ограничено полиномом от максимального значения $\log|X|$, которое не превосходит логарифма числа элементов в разбиваемом множестве. Таким образом мы получаем желаемую оценку на число частей.

Осталось доказать, что в разбиении наименьшего веса все части почти однородны. Другими словами, надо показать, что достаточно неоднородная часть может быть разбита на две таким образом, что вес разбиения уменьшится. При разбиении части на две все остальные части и веса входящих в них точек никак не затрагиваются, поэтому можно изучать лишь изменение веса в пределах одной части. В формуле

$$\sum_{A,B} \log m_X(B|A) - d\log|X|,$$

для веса точки все члены (и слагаемые, и вычитаемое) уменьшаются. Нам надо разбить неоднородную часть на две так, чтобы уменьшение слагаемых (для обеих частей) пересилило бы уменьшение вычитаемого. Последнее легко подсчитать: если часть из m элементов разбивается на две части из pm и qm элементов (так что $p+q = 1$), то вычитаемое уменьшится на $dmh(p,q)$, где

$$h(p,q) = p(-\log p) + q(-\log q)$$

не превосходит единицы (это энтропия случайной величины с двумя значениями). Таким образом, прибавка в (удельном, в расчёте на точку) весе за счёт уменьшения вычитаемого не больше d.

Если множество (рассматриваемая нами часть) сильно неоднородно, то найдутся такие множества индексов A и B, что $m(A \cup B)$ много больше $m(A) \cdot m(B|A)$. Это означает, что если мы рассмотрим проекцию на $A \cup B$, то у этой проекции сечения, параллельные B, сильно различаются по размеру, и максимальное из них значительно (скажем, в l раз) превосходит среднее.

Эту проекцию (и тем самым всё множество) можно разбить на две, считая в качестве граничного размера сечения среднее геометрическое между максимальным и средним размером. У части с «малыми» сечениями тогда уменьшится максимальный размер сечения в $\sqrt{l}$ раз. С другой стороны, большие сечения по неравенству Чебышёва составляют не более доли $1/\sqrt{l}$ (иначе бы среднее было больше), поэтому у части с большими сечениями уменьшится (также в $\sqrt{l}$ раз) размер A-проекции.

Таким образом, после разбиения для каждой из новых частей хотя бы одно слагаемое в формуле для весов уменьшится на $\log\sqrt{l}$ (а остальные, как мы уже говорили, не возрастут). Поэтому при достаточно большом l (когда $\log\sqrt{l} > d$) общий вес заведомо уменьшится. Значит, в разбиении минимального веса все части c-однородны для некоторой константы c, которая (вслед за d и l) определяется лишь числом n (хотя и быстро растёт с ростом n), что и завершает доказательство леммы с использованием весов.

10.10. Неравенства для двух и трёх слов

Мы убедились, что имеется некоторый класс неравенств, которые можно определять самыми разными способами: неравенства для энтропий, для сложностей, для размеров проекций однородных множества, для размеров подгрупп и т. д. и т. п. Но хорошо бы понять всё-таки, каковы эти неравенства.

Это удалось сделать лишь для простейших случаев (при $n \leqslant 3$). Для $n = 1$ ситуация вообще тривиальна. При $n = 2$ имеются неравенства

$$0 \leqslant H(\xi_1) \leqslant H(\xi_1,\xi_2),\ 0 \leqslant H(\xi_2) \leqslant H(\xi_1,\xi_2),\ H(\xi_1,\xi_2) \leqslant H(\xi_1) + H(\xi_2)$$

которые означают, что три величины

$$H(\xi_2|\xi_1),\ H(\xi_1|\xi_2),\ I(\xi_1 : \xi_2)$$

неотрицательны. С другой стороны, ясно, что значения этих трёх величин могут быть любыми неотрицательными числами: возьмём три независимые величины α, β, γ с такими энтропиями и положим

$$\xi_1 = \langle \alpha, \beta \rangle, \xi_2 = \langle \beta, \gamma \rangle,$$

тогда, как легко проверить, $H(\xi_1|\xi_2) = H(\alpha)$, $I(\xi_1 : \xi_2) = H(\beta)$, $H(\xi_2|\xi_1) = H(\gamma)$. Таким образом, в данном случае выясняется, что указанные неравенства необходимы и достаточны, чтобы набор из трёх чисел мог быть равен

$$H(\xi_1), H(\xi_2), H(\xi_1,\xi_2),$$

и потому никаких других неравенств не нужно (любое другое неравенство будет следствием этих; заметим также, что линейное программирование учит, что всякое следствие есть линейная комбинация с неотрицательными коэффициентами).

Прежде чем переходить к случаю $n = 3$, отметим, что для $n = 2$ мы сделали нечто большее, чем просто описали все верные линейные неравенства для энтропий: мы доказали также, что все тройки чисел, удовлетворяющие этим неравенствам, могут быть значениями энтропий.

Геометрически это можно описать так. Для всякого набора случайных величин есть точка в (трёхмерном при $n = 2$) линейном пространстве, состоящая из энтропий комбинаций этих величин. Рассматривая разные наборы случайных величин, мы получаем некоторое множество $\mathcal{E}$ в этом линейном пространстве.

Линейные неравенства для энтропий — это замкнутые полупространства, содержащие целиком множество $\mathcal{E}$. В данном случае (при $n = 2$) мы установили, что $\mathcal{E}$ есть в точности многогранный конус, заданный тремя неравенствами.

В общем случае пересечение всех полупространств, содержащих некоторое множество, может быть и больше самого множества (например, если оно невыпукло). Поэтому, даже зная эти подпространства (зная все неравенства), мы имеем лишь частичную информацию о множестве возможных значений энтропий.

С другой стороны, именно неравенства (полупространства, содержащие $\mathcal{E}$) представляют наибольший интерес, поскольку результат о совпадении классов неравенств для сложностей, энтропий, размеров проекций однородных множеств и т. д. относится именно к ним. Сами множества возможных значений там другие: например, размеры проекций однородных множеств обязательно являются логарифмами целых чисел; для колмогоровских сложностей вообще всё определено лишь с точностью до $O(1)$ и так далее.

После этих комментариев перейдём к случаю $n = 3$. В этом случае имеется семимерное пространство (соответствующее семи непустым подмножествам трёхэлементного множества). Удобнее перейти к новым координатам $a_1, \ldots, a_7$ в этом векторном пространстве, описанным на с. 51 (рис. 5) для случая колмогоровских сложностей. В этих координатах известные нам базисные неравенства означают, что все a_i, кроме «центральной части» a_5, которую мы также обозначали $I(\xi_1 : \xi_2 : \xi_3)$, неотрицательны, а эта центральная часть неотрицательна в сумме с любой из трёх частей a_2, a_4, a_6. Другими словами, можно сказать, что множество $\mathcal{E} \subset \mathbb{R}^7$ всех возможных наборов энтропий содержится в множестве $\mathcal{F}$ всех семёрок, удовлетворяющих этому неравенству.

Множество $\mathcal{F}$, как легко понять, состоит из неотрицательных линейных комбинаций конечного числа векторов (образующих этого выпуклого конуса). А именно, образующими будут векторы, в которых одно из a_i равно единице, а остальные нулю. Кроме того, есть ещё один особый вектор e: у него $a_5 = -1$, $a_2 = a_4 = a_6 = 1$, остальные a_i равны нулю. Этого достаточно: хотя a_5 может быть отрицательно, но по модулю не может превосходит a_2, a_4 и a_6, поэтому надо взять особый вектор e с коэффициентом $|a_5|$ и добавить остальные по мере необходимости.

Теперь ясно, что других неравенств нет: поскольку все образующие реализуются как наборы энтропий (особый вектор соответствует независимым величинам ξ_1, ξ_2 с двумя равновероятными значениями и $\xi_3 = \xi_1 + \xi_2 \pmod 2$), то любое верное неравенство верно для образующих, и потому для всего F, то есть вытекает из базисных неравенств.

$\boxed{218}$ [nonconvex-e] Покажите, что множество $\mathcal{E}$ невыпукло: а именно, λe принадлежит $\mathcal{E}$ тогда и только тогда, когда λ есть логарифм целого числа.

$\boxed{219}$ [adding-e] Докажите, что множество $\mathcal{E}$ (для любого числа случайных величин n) замкнуто относительно сложения: если два вектора $e, e' \in \mathbb{R}^{2^n-1}$ принадлежат $\mathcal{E}$, то и их сумма $e+e'$ принадлежит $\mathcal{E}$. [Указание. Рассмотрите два независимых набора случайных величин, задающих векторы e и e', и соединим их в один.]

$\boxed{220}$ [convex-closure-e] Докажите, что замыкание множества $\mathcal{E}$ (для любого числа случайных величин) выпукло. [Указание. Если e и e' принадлежат множеству $\mathcal{E}$, то при больших целых k и l вектор $ke+le'$ также принадлежит множеству $\mathcal{E}$, осталось лишь научиться умножать вектор на число. На целые числа мы его умножать умеем, а приближённо делить на большое число можно так: с малой вероятностью ε берём данные случайные величины, в противном случае тривиальные (имеющие только одно значение).]

10.11. Размерности и неравенство Инглетона

В случае двух и трёх случайных величин (или слов) мы описали все истинные линейные неравенства для энтропий (и тем самым для сложностей); при этом для случая $n=2$ мы описали даже множество $\mathcal{E}$, а не только двойственное к нему множество всех неравенств, выполненных для всех элементов из $\mathcal{E}$.

Для случая четырёх и более слов такого сделать уже не удаётся, и полный набор неравенств до сих пор не известен. Попытаемся описать, что про это известно.

Напомним, что мы рассматриваем набор случайных величин $\xi = \xi_1, \ldots, \xi_n$ и через ξ_I (для данного множества индексов $I \subset \{1,\ldots,n\}$) обозначаем часть этого набора, состоящую из величин с индексами из I. Через $H(\xi_I)$ обозначается энтропия этой части. Условные энтропии $H(\xi_I|\xi_J)$ (которые имеют смысл при непересекающихся множествах I и J, так как общие элементы можно удалить из I) выражаются через безусловные и потому их нет необходимости рассматривать отдельно.

Каждому набору случайных величин $\langle \xi_1, \ldots, \xi_n\rangle$ соответствует точка в пространстве $\mathbb{R}^{2^n-1}$, составленная из чисел $H(\xi_I)$ при всех непустых I. Эти точки (для всех наборов) образуют множество, которое мы обозначали через $\mathcal{E}$. Как мы видели в задачах 218, 219 и 220, это множество не обязательно выпукло, но его замыкание является выпуклым конусом (вместе с любыми двумя точками содержит их положительные линейные комбинации).

От множества $\mathcal{E}$ можно перейти к двойственному множеству всех линейных неравенств, выполненных для энтропий случайных величин (то есть для элементов $\mathcal{E}$). Геометрически это означает, что мы рассматриваем все полупространства, содержащие $\mathcal{E}$. Обратный переход даёт множество всех точек, для которых выполнены все эти неравенства (пересечение всех упомянутых полупространств); линейное программирование учит, что это пересечение есть минимальный замкнутый выпуклый конус, содержащий $\mathcal{E}$.

Среди неравенств заведомо есть следующие:

$$\begin{aligned} H(\xi_I) &\geqslant 0 && \text{для любого } I\\ H(\xi_I) &\leqslant H(\xi_J) && \text{для любых } I \subset J\\ H(\xi_{I\cap J}) + H(\xi_{I\cup J}) &\leqslant H(\xi_I) + H(\xi_J) && \text{для любых } I, J \end{aligned}$$

Будем называть эти неравенства *базисными*. Этот термин раньше употреблялся по отношению к неравенству

$$H(\xi_1) + H(\xi_1, \xi_2, \xi_3) \leqslant H(\xi_1, \xi_2) + H(\xi_1, \xi_3),$$

которое соответствует случаю $I = \{1, 2\}$, $J = \{1, 3\}$, и к которому сводится неравенство для произвольных I и J, если объединить случайные величины в группы. Для удобства мы включаем в базисные неравенства и неравенства первых двух типов.

Таким образом, множество $\mathcal{E}$ содержится в многогранном конусе, состоящем из всех наборов, удовлетворяющих базисным неравенствам. В случае $n = 2$ имеет место совпадение; при $n = 3$, как мы говорили, совпадения нет, но $\mathcal{E}$ плотно в этом конусе. При $n = 4$ это, как мы увидим, уже не так.

Выяснение ситуации естественно начать с описания структуры конуса. Линейное программирование учит, что всякое множество, задаваемое системой однородных линейных неравенств, есть множество всех положительных линейных комбинаций своих крайних точек (рёбер конуса). Если бы (как это происходит при $n = 3$) все эти крайние точки попали в $\mathcal{E}$, то мы бы установили, что все неравенства для энтропий следуют из базисных (потому что всякое неравенство, верное для рёбер, верно и для их положительных линейных комбинаций).

Для случая $n = 4$ такие рёбра можно найти вручную или с помощью компьютерной программы; это было сделано (см. статью [16], где приведён список всех рёбер). Выяснилось, что большинство рёбер действительно принадлежит $\mathcal{E}$ (и соответствующие случайные величины легко найти, см. [16]), но есть и несколько особых рёбер. Эти особые рёбра по существу все одинаковы (отличаются перестановкой переменных), и мы приведём одно из таких особых рёбер:

$$\begin{aligned}
H(\xi_1) = H(\xi_2) = H(\xi_3) = H(\xi_4) = 2n;\\
H(\xi_1, \xi_2) = 4n;\\
H(\xi_1, \xi_3) = H(\xi_1, \xi_4) = H(\xi_2, \xi_3) = H(\xi_2, \xi_4) = H(\xi_3, \xi_4) = 3n;\\
H(\xi_1, \xi_2, \xi_3) = H(\xi_1, \xi_2, \xi_4) = H(\xi_1, \xi_3, \xi_4) = H(\xi_2, \xi_3, \xi_4) = 4n;\\
H(\xi_1, \xi_2, \xi_3, \xi_4) = 4n.
\end{aligned}$$

Другими словами, каждое слово имеет сложность $2n$, все слова вместе и все тройки слов имеют сложность $4n$, а все пары имеют сложность $3n$, за исключением одной, которая имеет сложность $4n$. (Множитель n указан, поскольку ребро конуса есть луч, и его точки определены с точностью до пропорциональности.)

Довольно трудно представить себе смысл этих условий; картинки для четырёх слов довольно запутаны. Можно заметить, что величины ξ_1 и ξ_2 входят симметрично; аналогично для величин ξ_3 и ξ_4. Можно нарисовать схемы распределения сложности для троек случайных величин (рис. 27): Пытаясь придумать случайные величины (или двоичные слова) с такой энтропией (сложностью), мы сталкиваемся со следующей трудностью: правая картинка показывает, что ξ_3 и ξ_4 хорошо бы иметь n битов общей информации, и вся эта информация входит как в ξ_1, так и в ξ_2. С другой стороны, левая картинка показывает, что ξ_1 и ξ_2 не должны иметь общей информации (в пересечении стоят нули). Конечно, это всего лишь разговоры, поскольку понятие слова «n битов общей информации» ничего

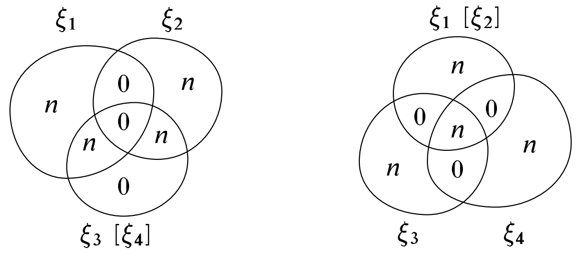

Рис. 27. Распределение сложностей для троек слов.

[ineq.4]

определённого не означают. Однако, как мы увидим ниже, действительно таких величин $\xi_1, \xi_2, \xi_3, \xi_4$ не существует.

Если особые рёбра оказываются недостижимыми, можно предположить, что выполняются некоторые пока неизвестные нам неравенства. Как их найти? Простейшая гипотеза (как мы увидим позже, неверная, но пока мы про это не знаем) состоит в том, что мы уже знаем все рёбра конуса (неособые рёбра из числа найденных) и нужно лишь найти неравенства, образующие его грани. Это также можно сделать с помощью компьютера, и мы получаем новые неравенства

$$I(\xi_3 : \xi_4) \leqslant I(\xi_3 : \xi_4|\xi_1) + I(\xi_3 : \xi_4|\xi_2) + I(\xi_1 : \xi_2).$$

(и аналогичные, получаемые перестановкой переменных). Для наглядности мы записали это неравенство с условными сложностями; выражая их через безусловные, мы получаем

$$12 + 3 + 4 + 134 + 234 \leqslant 13 + 23 + 14 + 24 + 34$$

(для краткости мы пишем лишь индексы: скажем, 134 обозначает $H(\xi_1, \xi_3, \xi_4)$), но в таком виде его смысл ещё менее ясен.

Оказывается, что это неравенство хорошо известно в теории матроидов и называется неравенством Инглетона:

Theorem 189. [ingleton] *Для любых конечномерных подпространств H_1, H_2, H_3, H_4 векторного пространства выполнено неравенство*

$$\begin{aligned}\dim(H_1 + H_2) + \dim H_3 + \dim H_4 + \dim(H_1 + H_3 + H_4) + \dim(H_2 + H_3 + H_4) \leqslant \\ \leqslant \dim(H_1 + H_3) + \dim(H_2 + H_3) + \dim(H_1 + H_4) + \dim(H_2 + H_4) + \dim(H_3 + H_4)\end{aligned}$$

Прежде чем доказывать эту теорему, полезно понять связи между неравенствами для энтропий и для размерностей.

Рассмотрим конечномерное пространство X над конечным полем $\mathbb{F}$ и его подпространства. Этим подпространствам соответствуют случайные величины следующим образом. Вероятностным пространством будет множество всех линейных функционалов $X \to \mathbb{F}$, и каждому подпространству $Y \subset X$ соответствует случайная величина ξ_Y, которая сопоставляет с каждым функционалом его ограничение на Y. (Менее формально: случайная величина, соответствующая Y, есть ограничение случайного линейного функционала на подпространство Y.) Значениями такой случайной величины являются элементы пространства Y^*, сопряжённого к Y, и все они равновероятны, их число равно $\mathbb{F}$ в степени $\dim Y^* = \dim Y$, так что энтропия такой случайной величины равна $\dim Y \cdot \log |\mathbb{F}|$.

Заметим, что у нас не просто для каждого подпространства есть случайная величина, а есть некоторое их совместное распределение. Поэтому можно задать такой вопрос. Пусть Y и Z — два подпространства. Какова будет энтропия пары величин $\langle \xi_Y, \xi_Z \rangle$? Заметим, что ограничения случайного функционала на Y и Z однозначно определяют его ограничение на сумму подпространств $Y + Z = \{y + z | y \in Y, z \in Z\}$ и наоборот. Поэтому энтропия пары равна $\dim(Y + Z) \cdot \log |\mathbb{F}|$.

Из этого наблюдения сразу же следует такая

Theorem 190. [entropy-dimension] *Всякому неравенству, выполненному для энтропий случайных величин и их наборов, соответствует неравенство для размерностей конечномерных подпространств векторного пространства над конечным полем (в котором энтропии группы величин соответствует размерность суммы подпространств).*

$\boxed{\textbf{221}}$ Покажите, что аналогичное утверждение верно для размерностей конечномерных подпространств векторного пространства над $\mathbb{R}$ и над $\mathbb{C}$. [Указание. Комплексный случай сводится к вещественному, так как размерности вдвое больше. В вещественном случае будем считать пространство евклидовым и рассмотрим случайные величины, являющиеся проекциями случайной точки единичного шара на каждое из подпространств, причём в каждом пространстве значения округляются с точностью ε. При этом возникают разные погрешности (проекция случайной точки шара не равномерна в круге; из-за округления проекции на X и Y не определяют проекцию на $X + Y$ однозначно, а лишь с точностью до конечного числа вариантов), но главные члены при $\varepsilon \to 0$ всё-таки пропорциональны размерностям.]

[ВОПРОС: а как это доказать для произвольного поля (если это вообще верно? Может, есть какие-то простые алгебраические соображения, что это выражается какими-то определителями и т.п.?]

Заметим ещё, что переход от подпространств к соответствующим случайным величинам является довольно общим приёмом построения примеров точек из $\mathcal{E}$. В частности, все неособые рёбра конуса, о котором говорилось выше (см. [16]), могут быть получены как раз таким способом. Это же относится и ко всем до сих пор встречавшимся примерам элементов $\mathcal{E}$; существенно другие примеры появятся лишь в следующем разделе, когда будет идти речь об условно независимых случайных величинах.

$\triangleleft$ Перейдём теперь к доказательству неравенства Инглетона. Его нельзя непосредственно вывести из теоремы 190 (поскольку для энтропий оно не выполняется, как мы впоследствии увидим). Чтобы его доказать, надо установить ещё некоторые связи между энтропиями и размерностями.

В наших неравенствах для энтропий можно считать условную энтропию $H(\alpha|\beta)$ сокращением для $H(\alpha, \beta) - H(\beta)$. В соответствующем неравенстве для размерностей условная энтропия переводится как $\dim(A + B) - \dim B$, что равно размерности образа A при линейном отображении, ядро которого равно B (образа A при факторизации по B). Как учит линейная алгебра, это можно переписать как $\dim A - \dim(A \cap B)$. Другое сокращение, $I(\alpha : \beta)$, расшифровывается как $H(\alpha) + H(\beta) - H(\alpha, \beta)$, что соответствует

$\dim A + \dim B - \dim(A+B)$. Последнее выражение равно $\dim(A \cap B)$. Наконец, $I(\alpha : \beta|\gamma)$ можно развернуть (не до конца) в

$$H(\alpha|\gamma) + H(\beta|\gamma) - H(\alpha,\beta|\gamma)$$

и ему соответствует

$$\dim A/C + \dim B/C - \dim(A+B)/C$$

если через X/C обозначить пространство, получаемое из X при линейном отображении с ядром C. Заметим, что последнее выражение нельзя переписать в виде $\dim(A \cap B)/C$: это выражение равно размерности пересечения образов подпространств A и B при факторизации по C, а это пересечение содержит образ $A \cap B$ при факторизации по C, но не обязано с ним совпадать. (Пусть, например, A, B, C — три различных одномерных подпространства двумерного пространства.)

Возвращаясь к неравенству Инглетона для размерностей подпространств, мы можем переписать его так:

$$\dim(A \cap B) \leqslant I(A : B|C) + I(A : B|D) + \dim(C \cap D)$$

(где $I(A : B|C)$ обозначает размерность пересечения образов A и B при факторизации по C). Обозначая $A \cap B$ через X, мы видим, что достаточно доказать

$$\dim X \leqslant \dim X/C + \dim X/D + \dim(C \cap D),$$

поскольку $\dim X/C \leqslant I(A : B|C)$ (образ пересечения при факторизации содержится в пересечении образов, хотя может быть и меньше). А это неравенство соответствует легко проверяемому неравенству

$$H(\xi) \leqslant H(\xi|\gamma) + H(\xi|\delta) + I(\gamma : \delta)$$

для энтропий, и остаётся применить теорему 190. $\triangleright$

$\boxed{222}$ [`quasi-ingleton`] Докажите неравенство $H(\xi) \leqslant H(\xi|\gamma) + H(\xi|\delta) + I(\gamma : \delta)$ для энтропий. [Указание. Как легко заметить по картинке и проверить прямым вычислением,

$$H(\xi) + H(\xi|\gamma,\delta) + I(\gamma : \delta|\xi) = H(\xi|\gamma) + H(\xi|\delta) + I(\gamma : \delta),$$

так что рассматриваемое неравенство является суммой базисных.]

$\boxed{223}$ Строго говоря, наше доказательство неравенства Инглетона годится для пространств над конечным полем (или над полем $\mathbb{R}$, если воспользоваться соответствующей задачей). Как перенести его на случай произвольного поля? [Указание. Поле было существенно, когда мы переходили от неравенства для энтропий к неравенству для размерностей. Но это неравенство для энтропий является комбинацией базисных, а базисные неравенства верны для размерностей над произвольным полем.]

$\boxed{224}$ Мы знаем, что неравенствам для энтропий соответствуют неравенства для размеров подгрупп и их пересечений. Покажите, что неравенству Инглетона соответствует неравенство, верное для подгрупп *коммутативной* группы. [Указание: Действуйте как при доказательстве теоремы 189, роль пересечения подпространств играет сумма подгрупп (которая есть подгруппа в коммутативном случае).]

В качестве побочного продукта наших рассмотрений получаем такое любопытное утверждение:

225 Докажите, что любое линейное неравенство, выполненное для произвольных четвёрок подпространств, следует из базисных неравенств и неравенства Инглетона. [Указание. Как мы уже говорили, все рёбра конуса решений базисных неравенств, кроме упоминавшихся особых, реализуемы не только случайными величинами, но и подпространствами, а выбросив из числа образующих особые рёбра, мы получим конус, гранями которого являются базисные неравенства и неравенства Инглетона (для разных порядков переменных). Как обойтись здесь без длинных (ручных или компьютерных) вычислений, неясно.]

226 Сформулируйте и докажите аналогичное утверждение для четырёх конечных подгрупп коммутативной группы.

10.12. Условно независимые случайные величины

[conditional-independence]

Покажем теперь (как мы давно обещали), что неравенство Инглетона для произвольных случайных величин может нарушаться. А именно, иногда правая его часть обращается в нуль, а левая нет.

Говорят, что случайные величины α и β ***независимы при условии*** γ (где γ — третья случайная величина на том же вероятностном пространстве), если $I(\alpha : \beta|\gamma) = 0$. Легко проверить, что это равносильно следующему: для любого значения γ_0 величины γ, принимаемого с ненулевой вероятностью, условные распределения α и β внутри события $\gamma = \gamma_0$, независимы.

227 Проверьте это. [Указание: $I(\alpha : \beta|\gamma)$ есть среднее значение (по всем γ_0 с весами, равными их вероятностей) взаимной информации соответствующих распределений.]

Будем говорить, что случайные величины α и β ***условно независимы***, если найдутся (на том же вероятностном пространстве или на его измельчении) случайные величины γ и δ с такими свойствами:

- γ и δ независимы;
- α и β независимы при условии γ;
- α и β независимы при условии δ.

Благодаря нашей оговорке (разрешению измельчать вероятностное пространство, разбивая элементарные события на несколько событий нужной суммарной вероятности) свойство условной независимости становится свойством совместного распределения величин α и β и не зависит от выбора вероятностного пространства. (Напомним, что здесь, как и всюду, мы рассматриваем лишь случайные величины с конечным числом значений.)

Три требования, входящие в определение условной независимости, означают, что три слагаемых в правой части неравенства Инглетона равны нулю. Осталось показать, что отсюда не следует, что левая его часть равна нулю:

Theorem 191. [conditionally-only] *Существуют условно независимые случайные величины, не являющиеся независимыми.*

$\lhd$ Для доказательства достаточно предъявить четвёрку случайных величин $\alpha, \beta, \gamma, \delta$, для которой бы выполнялись требования из определения условной независимости, но α и β были бы зависимыми. Укажем такой пример. Каждая из четырёх величин принимает значения 0 и 1 с вероятностью $1/2$. Величины γ и δ независимы, и каждая из четырёх комбинаций имеет вероятность $1/4$.

Величины α и β определяются так: если $\gamma = \delta$, то это общее значение будет одновременно значением величин α и β (которые в этом случае равны). Если же $\gamma \neq \delta$, то условное распределение вероятностей величин α и β) (в каждом из случаев $\gamma = 1, \delta = 0$ и $\gamma = 0, \delta = 1$ задаётся таблицей

	0	1
0	1/8	3/8
1	3/8	1/8

Легко подсчитать, что (скажем) при фиксированном $\gamma = 0$ условное распределение вероятностей для α и β будет полусуммой этой матрицы и матрицы

	0	1
0	1	0
1	0	0

Оно равно

	0	1
0	9/16	3/16
1	3/16	1/16

что есть распределение двух независимых величин, равных нулю с вероятностью $3/4$. Вместе с тем совместное распределение величин α и β есть среднее арифметическое всех четырёх матриц условных распределений и равно

	0	1
0	5/16	3/16
1	3/16	5/16

так что величины α и β зависимы. $\rhd$

Следствием этой теоремы является такое утверждение: при $n = 4$ неотрицательные линейные комбинации векторов, соответствующих размерностям пространств, не исчерпывают всего $\mathcal{E}$ (поскольку для них выполнено неравенство Инглетона, которое верно не всегда). Наша нижняя оценка для замыкания множества $\mathcal{E}$ при $n = 4$ (неособые рёбра и их комбинации) оказывается, таким образом, не наилучшей.

Как мы увидим в следующем разделе, и верхняя оценка (конус решений базисных неравенств) тоже не является точной: при $n \geqslant 4$ существуют неравенства, не вытекающие из базисных, но верные для всех элементов $\mathcal{E}$

10.13. Неравенства, не сводящиеся к базисным

Theorem 192. [rv-inequality] *Для любых случайных величин* $\alpha, \beta, \gamma, \delta, \varepsilon$ *выполняется неравенство*

$$I(\alpha : \beta) \leqslant I(\alpha : \beta|\gamma) + I(\alpha : \beta|\delta) + I(\gamma : \delta) + \\ + I(\alpha : \beta|\varepsilon) + I(\alpha : \varepsilon|\beta) + I(\beta : \varepsilon|\alpha)$$

Смысл и природа этого неравенства, увы, не вполне понятны. Но некоторые комментарии к нему (хотя бы для его запоминания) всё-таки сделать можно.

Его правая часть делится на две группы (записанные нами в разных строках). Первая группа знакома нам по неравенству Инглетона (если бы второй группы не было, то так бы и получилось). Таким образом, наше неравенство является ослаблением неравенства Инглетона за счёт дополнительного (неотрицательного) слагаемого

$$W(\alpha, \beta, \varepsilon) = I(\alpha : \beta|\varepsilon) + I(\alpha : \varepsilon|\beta) + I(\beta : \varepsilon|\alpha)$$

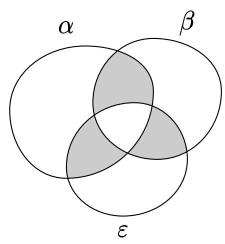

Рис. 28. Величина $W(\alpha, \beta, \varepsilon)$.

[ineq.6]

Это слагаемое (символически показанное на рис. 28) содержит величину ε, которая в оставшуюся часть не входит. Так что можно было бы сказать и так: неравенство Инглетона справедливо с погрешностью, которая не превосходит $\inf_\varepsilon W(\alpha, \beta, \varepsilon)$.

Другое наблюдение: если взять в качестве α, β и ε одну и ту же случайную величину ξ (не одинаково распределённые величины, а именно одну и ту же), то $W(\alpha, \beta, \varepsilon)$ обращается в нуль, и мы получаем неравенство

$$H(\xi) \leqslant H(\xi|\gamma) + H(\xi|\delta) + H(\gamma : \delta) \qquad (*)$$

которое уже встречалось нам при доказательстве неравенства Инглетона для размерностей линейных пространств.

Ещё одно замечание: частным случаем теоремы 192 является такое утверждение: если $W(\alpha, \beta, \varepsilon) = 0$, то

$$I(\alpha : \beta) \leqslant I(\alpha : \beta|\gamma) + I(\alpha : \beta|\delta) + I(\gamma : \delta)$$

для любых γ и δ. Кстати, это утверждение можно доказать и непосредственно, применив неравенство $(*)$ к величине ξ из следующей теоремы (которую нужно применять к величинам α, β и ε).

Theorem 193. [w-zero] *Если $W(\alpha,\beta,\gamma) = 0$, то у тройки величин α,β,γ выделяется общая информация в следующем смысле: найдётся такая величина ξ, что*

$$H(\xi|\alpha) = H(\xi|\beta) = H(\xi|\gamma) = 0;$$

$$I(\alpha:\beta|\xi) = I(\beta:\gamma|\xi) = I(\alpha:\gamma|\xi) = 0$$

$\lhd$ Рассмотрим трёхмерную таблицу распределений вероятностей для величин α,β,γ. Условие теоремы означает, что любое двумерное сечение этой таблицы (параллельное осям координат) имеет ранг 1 (если оно имеет ранг нуль, то соответствующее значение вообще не встречается и эту плоскость можно выбросить).

Предположим сначала, что все элементы таблицы не равны нулю и покажем, что в этом случае величины α,β,γ независимы. В самом деле, рассмотрим, скажем, все одномерные сечения, параллельные первой координате, как векторы. Получится двумерная таблица, в каждой клетке которой стоит ненулевой вектор. По условию в каждом столбце все векторы столбце пропорциональны, и то же самое для строк. Поскольку векторы ненулевые, то и все они пропорциональны, то есть все плоскости таблицы (в перпендикулярном векторам направлении) пропорциональны и имеют ранг 1, что и требовалось доказать.

Рассмотрим теперь общий случай. Покажем, что наша трёхмерная таблица «блочно-диагональная». Это означает, что можно (при некотором k) представить таблицу как соединение k блоков, каждый из которых заполняет «параллелепипед» (произведение множеств по каждой координате), и проекции этих k параллелепипедов на каждую координату не пересекаются, а вне блоков стоят нули.

Из ранее доказанного следует, что внутри каждого блока имеет место независимость, то есть это как раз соответствует утверждению теоремы (величина ξ есть номер блока, является функцией от любой из трёх величин α,β,γ и при известном значении ξ величины α,β,γ независимы).

Чтобы доказать, что таблица блочно-диагональная, достаточно использовать такое свойство множества ненулевых позиций (мест таблицы, где стоят положительные числа): если в двух противоположных углах прямоугольника, параллельного осям координат, стоят не-нули, то и в других двух стоят не-нули. (Иначе определитель бы не обращался в нуль и ранг был бы по крайней мере 2.)

В самом деле, будем постепенно увеличивать блок (начав с одноэлементного). Если какая-то ненулевая точка совпадает по одной из координат с точками блока, то из сформулированного свойства вытекает (как нетрудно проверить), что блок можно расширить (оставляя его параллелепипедом). Так будем расширять пока можно, после чего применим то же рассуждение к оставшейся части матрицы. $\rhd$

$\boxed{\textbf{228}}$ Проведите это рассуждение подробно.

$\boxed{\textbf{229}}$ (а) Докажите следующее утверждение (иногда называемое Double Markov property): если $I(\beta:\gamma|\alpha) = 0$ и $I(\alpha:\gamma|\beta) = 0$, то найдётся такая величина ξ, что $I(\xi|\alpha) = 0$, $I(\xi|\beta) = 0$ и $I(\langle\alpha,\beta\rangle:\gamma|\xi) = 0$. (б) Выведите отсюда утверждение теоремы 193

Всё сказанное, однако, позволяет доказать теорему 192 лишь в частных случаях. Вот общее доказательство:

$\lhd$ Мы используем (до конца ещё, видимо, не понятый) приём «независимизации», который состоит в следующем.

Переменные, входящие в неравенство, делятся на три группы:

$$(1)\ \alpha,\beta; \qquad (2)\ \gamma,\delta; \qquad (3)\ \varepsilon$$

Заметим, что переменные второй и третьей групп никогда не встречаются вместе (хотя и те, и другие по отдельности сочетаются с переменными первой группы). Поэтому *без ограничения общности можно предполагать, что пара $\langle\gamma,\delta\rangle$ независима с ε при известном $\langle\alpha,\beta\rangle$*. В самом деле, если мы изменим совместное распределение, выполнив (для каждого значения пары $\langle\alpha,\beta\rangle$) «принудительную независимизацию» величин $\langle\gamma,\delta\rangle$ и ε, то совместное распределение величин первой и второй групп, а также совместное распределение величин первой и третьей групп останутся неизменными, и все величины, входящие в неравенство, не изменятся.

Раз так, то достаточно доказать (вообще говоря, более слабое) неравенство

$$\begin{aligned} I(\alpha:\beta) \leqslant{} & I(\alpha:\beta|\gamma) + I(\alpha:\beta|\delta) + I(\gamma:\delta) + \\ & + W(\alpha,\beta,\varepsilon) + \\ & + I(\langle\gamma,\delta\rangle:\varepsilon|\langle\alpha,\beta\rangle) + \\ & + I(\gamma:\varepsilon|\langle\alpha,\beta\rangle) + \\ & + I(\delta:\varepsilon|\langle\alpha,\beta\rangle) \end{aligned}$$

(заметим, что если пара $\langle\gamma,\delta\rangle$ независима с ε при известных α,β, то каждая из величин γ и δ по отдельности также независима с δ при известных α,β), а это неравенство удивительных образом является суммой восьми базисных неравенств

$$\begin{aligned} I(\alpha,\beta:\varepsilon|\gamma,\delta) &\geqslant 0; \\ I(\alpha:\beta|\varepsilon,\gamma) &\geqslant 0; \\ I(\alpha:\beta|\varepsilon,\delta) &\geqslant 0; \\ I(\gamma:\delta|\varepsilon) &\geqslant 0; \\ I(\gamma:\varepsilon|\alpha) &\geqslant 0; \\ I(\gamma:\varepsilon|\beta) &\geqslant 0; \\ I(\delta:\varepsilon|\alpha) &\geqslant 0; \\ I(\delta:\varepsilon|\beta) &\geqslant 0. \end{aligned}$$

Это несложно проверить: если раскрыть все взаимные информации и сложить, то все лишние члены замечательным образом сокращаются. (Но почему так получается и как до этого можно было догадаться, неясно.) $\triangleright$

Таким образом, наше неравенство, хотя и не в прямом смысле, является следствием базисных. Можно сказать, что мы обнаружили (помимо сложения неравенств с неотрицательными коэффициентами) ещё одно «правило вывода»: если удаётся разбить переменные в каком-то неравенстве на три группы, причём переменные второй и третьей групп не входят одновременно ни в одно слагаемое, то можно вывести это неравенство из (вообще говоря) более слабого неравенства, куда добавлена ещё и общая информация переменных второй и третьей групп при известных переменных первой группы.

[Какие-нибудь другие применения этого общего правила вывода?]

[История вопроса? Условные неравенства? Альтернативное доказательство методом Ромащенко?]

[Упоминания: результаты Макарычевых об условной независимости]

11. Общая информация

[cominf]

11.1. Представление слов в несжимаемом виде

В какой степени можно представлять себе «информацию, содержащуюся в данном слове», как нечто материальное? Допустим, у нас есть слово x, «содержащее n битов информации», то есть имеющее сложность n. Можно ли — как если бы эти биты были камешками — разложить их на две равные кучки? Этому вопросу можно придать формальный смысл: можно ли найти такие слова x_1 и x_2 сложности $n/2$, для которых $K(x_1|x) \approx 0$, $K(x_2|x) \approx 0$ (слова x_1, x_2 «не содержат новой информации по сравнению с x») и $K(x|x_1, x_2) \approx 0$ («никакая информация не потеряна»)? Если, как обычно, понимать приближённое равенство как совпадение с точностью до $O(\log n)$, где n — максимальная длина (или сложность) рассматриваемых слов, то ответ на этот вопрос положительный, как мы сейчас увидим.

Будем говорить, что слова x и x' *эквивалентны с точностью c*, или *c-эквивалентны*, если $\mathit{KS}(x|y) \leqslant c$ и $\mathit{KS}(y|x) \leqslant c$. Это отношение не является, конечно, настоящим отношением эквивалентности: если x эквивалентно y, а y эквивалентно z с точностью c, то мы можем утверждать лишь, что x эквивалентно z с точностью $2c+O(\log c)$. (Можно было бы рассматривать не слова, а последовательности слов полиномиально растущей длины и говорить, что последовательность $x_0, x_1, \ldots$ эквивалентна $y_0, y_1, \ldots$, если $\mathit{KS}(x_i|y_i) = O(\log i)$ и $\mathit{KS}(y_i|x_i) = O(\log i)$; тогда бы это было настоящее отношение эквивалентности.)

Легко проверить, что сложности c-эквивалентных слов мало отличаются (не более чем на $O(c)$ и даже $c + O(\log c)$). Более общо, при замене слова на c-эквивалентное ему все сложности с его участием меняются на $O(c)$, так что, скажем, величина $I(x:y|z)$ меняется не более чем на $O(c)$ при замене любого из слов x, y, z (или всех трёх) на c-эквивалентное. Теперь мы можем сделать следующее (почти очевидное) наблюдение:

Theorem 194. [one-string-structure] *Для всякого слова x существует слово x' длины $\mathit{KS}(x)$, эквивалентное x с точностью до $O(\log \mathit{KS}(x))$. Это слово является «несжимаемым» (сложность близка к длине) с той же точностью.*

$\triangleleft$ В самом деле, возьмём в качестве x' кратчайшее описание слова x, которое как раз имеет длину $\mathit{KS}(x)$. Ясно, что $\mathit{KS}(x|x') = O(1)$, поскольку x восстанавливается по x'. С другой стороны, цепочка неравенств

$$\mathit{KS}(x) \leqslant \mathit{KS}(x, x') \leqslant \mathit{KS}(x') \leqslant l(x') = \mathit{KS}(x)$$

(выполненных с точностью до $O(\log \mathit{KS}(x))$), показывает, что все эти неравенства обращаются в равенства (с той же точностью), и потому $\mathit{KS}(x, x') \approx \mathit{KS}(x)$ и по теореме о сложности пары $\mathit{KS}(x'|x) \approx 0$. $\triangleright$

Из этого утверждения очевидно следует ответ на поставленный нами вопрос о разбиении слова на части: можно заменить слово на эквивалентное несжимаемое, и в качестве x_1 и x_2 взять просто-напросто левую и правую половины.

$\boxed{\textbf{230}}$ Проверьте, что выполнены все требуемые свойства.

$\boxed{231}$ Докажите, что если $KS(y|x) = n$, то существует такое («промежуточное») z, что $KS(z|x) \approx n/2$ и $KS(y|z) \approx n/2$ с точностью до $O(\log KS(x, y))$.

Так что пока биты информации ведут себя вполне материальным образом. Более того, это верно и в более общем случае, когда одновременно рассматривается информация в слове и её часть.

Пусть даны слова x и y, причём $KS(y|x) \approx 0$ («информация в слове y является частью информации в слове x»). Тогда существует несжимаемое слово x', эквивалентное x, некоторое начало y' которого эквивалентно y. (Такое начало автоматически будет несжимаемым и иметь длину $KS(y)$.) Более точно это утверждение формулируется так:

Theorem 195. *Для любых слов x и y существуют слова x' и y', эквивалентные x и y (соответственно) с точностью до $O(KS(y|x) + \log KS(x, y))$ и несжимаемые с той же точностью, причём y' является началом x'.*

$\triangleleft$ Рассмотрим в качестве y' кратчайшее описание слова y, оно несжимаемо и имеет длину примерно $KS(y)$. Теперь рассмотрим кратчайшее описание z' слова x при известном слове y; оно также будет несжимаемым, а длина его примерно равна $KS(x|y)$. Зная y' и z', можно восстановить сначала y, а потом и x, поэтому сложность пары y', z' не меньше $KS(x, y)$ (что примерно равно $KS(x)$, поскольку $KS(y|x)$ считается малым). С другой стороны, суммарная длина слов y' и z' равна $KS(y) + KS(x|y) \approx KS(x, y)$, так что слово $x' = y'z'$ несжимаемо. Сложность любого начала слова относительно этого слова логарифмическая, поэтому $KS(y'|x') \approx 0$ и, как легко проверить, $KS(x|x') \approx 0$. Кроме того, $KS(x'|x) \approx 0$, что легко следует из такой леммы:

Лемма. Если z' — кратчайшее описание x при известном y, то $KS(z'|x, y) = O(\log KS(x, y))$.

В самом деле,

$$KS(x, y, z') \leqslant K(y, z') \leqslant KS(y) + l(z') = KS(y) + KS(x|y) = KS(x, y)$$

(все равенства с точностью до $O(\log KS(x, y))$, поэтому $KS(z'|x, y) \approx 0$.

Лемма, а с ней и наша теорема, доказана. $\triangleright$

Эта теорема показывает, что пару слов x, y с $KS(y|x) \approx 0$ действительно можно себе представлять так: слово x состоит из $KS(x)$ почти что материальных битов, из которых первые $KS(y)$ битов образуют y.

11.2. Выделение общей информации

[cominf-profile]

Можно пойти ещё дальше и интересоваться аналогичным утверждением для двух произвольных слов. Напомним, что три основных сложностных характеристики пары слов x, y — это их сложности $KS(x)$, $KS(y)$ и сложность пары $KS(x, y)$; через них, как мы говорили в разделе 2.3, выражаются (с логарифмической точностью) условные сложности и взаимная информация:

$$\begin{aligned} KS(x|y) &= KS(x, y) - KS(y), \\ KS(y|x) &= KS(x, y) - KS(y), \\ I(x : y) &= KS(x) + KS(y) - KS(x, y) \end{aligned}$$

(см. рис. 3 на с. 47). Там же говорилось, что иногда этот рисунок можно понимать буквально, взяв в качестве x и y пересекающиеся подслова длинного случайного слова. Можно предположить, что это общий случай, то есть верна следующая

> **Гипотеза**: для любых слов x и y существует несжимаемое слово u длины $KS(x,y)$, эквивалентное (с логарифмической точностью) паре $\langle x,y\rangle$, причём начало этого слова длины $KS(x)$ эквивалентно x, а конец длины $KS(y)$ эквивалентен y (с той же точностью).

Но это, как выясняется, не так. Чтобы убедиться в этом, заметим, что из нашей гипотезы (для слов x и y) следует, что существует слово z (средняя часть), для которой

$$\begin{aligned} KS(z|x) &= 0, \\ KS(z|y) &= 0, \\ KS(z) &= I(x:y) \end{aligned}$$

(все равенства понимаются с логарифмической точностью). Можно сказать, что в этом случае «из слов x и y выделяется общая информация z». Как мы вскоре увидим, это возможно далеко не для всех пар слов x и y.

232 Покажите, что если общая информация выделяется (найдётся слово z с указанными тремя свойствами), то для пары слов x, y выполнена сформулированная выше гипотеза (найдётся слово u с указанными свойствами).

Построим опровергающий нашу гипотезу пример. Это можно сделать самыми разными способами, о которых мы ещё будем говорить, но, видимо, самый простой из них таков. Для определённости договоримся, что строимые нами слова x и y с невыделяемой общей информацией будут иметь сложность $2n$, а общая информация будет равна n (с логарифмической точностью), то есть пара $\langle x,y\rangle$ должна иметь сложность $3n$.

Мы хотим, чтобы из этой пары не выделялась общая информация, то есть чтобы не существовало слова z сложности n, для которого $KS(z|x)$ и $KS(z|y)$ обращаются в нуль (все равенства — с точностью $O(\log n)$). Нам удобнее будет записать эти условия иначе: $KS(x|z) = n$ и $KS(y|z) = n$. Это утверждение будет выполняться с некоторым запасом (мы разрешим сложностям быть даже немного больше n):

Theorem 196. [muchnik-example] *Для любого n существуют слова x и y, для которых $KS(x) = 2n + O(\log n)$, $KS(y) = 2n + O(\log n)$, $I(x:y) = n + O(\log n)$, но не существует такого слова z сложности меньше $1{,}1n$, что $KS(x|z) < 1{,}1n$ и $KS(y|z) < 1{,}1n$.*

◁ Как мы знаем, список всех слов сложности меньше k сам имеет сложность не больше $k + O(\log k)$ (чтобы его задать, достаточно, помимо числа k, указать число таких слов, которое имеет порядок $O(2^k)$ и потому требует k битов).

Аналогичное рассуждение позволяет установить, что список всех пар слов $\langle u,v\rangle$, для которых $KS(u) < 1{,}1n$ и $KS(v|u) < 1{,}1n$, имеет сложность не больше $2{,}2n + O(\log n)$: помимо n, достаточно указать число таких пар, а оно есть $O(2^{2{,}2n})$.

По тем же причинам список всех слов сложности меньше $2n$ имеет сложность не больше $2n + O(\log n)$, а список слов сложности меньше $3n$ имеет сложность не больше $3n + O(\log n)$.

Что можно сказать о сложности всех этих списков вместе взятых? Её можно оценить суммой сложностей. Но оказывается, что есть лучшая оценка: сложность всех списков вместе *не превосходит максимальной из наших оценок* (с точностью до $O(\log n)$). В самом деле, при известном n каждый из этих списков задаётся числом своих элементов. Более того, достаточно знать суммарное число элементов в этих списках, поскольку их можно перечислять и ждать появления нужного числа элементов: достигнув нужной суммы, мы автоматически исчерпаем все списки. А сумма нескольких чисел требует для записи лишь на несколько битов больше, чем самое длинное из них.

Итак, все вместе эти списки имеют сложность $3n+O(\log n)$. Пусть эти списки у нас есть. Используя их, мы перебором находим пару слов x, y длины $2n+2$ каждое, для которых:

- $\mathit{KS}(x) \geqslant 2n$;
- $\mathit{KS}(y) \geqslant 2n$;
- $\mathit{KS}(x,y) \geqslant 3n$;
- не существует слова z, для которого одновременно $\mathit{KS}(z) < 1{,}1n$, $\mathit{KS}(x|z) < 1{,}1n$ и $\mathit{KS}(y|z) < 1{,}1n$.

Все эти условия проверяются по нашим спискам, так что надо только убедиться, что такая пара существует. В самом деле, первое условие бракует не более четверти всех пар (всего слов x имеется 2^{2n+2}, из них негодных не больше 2^{2n}). Аналогично со вторым условием. Третье отбрасывает не более 2^{3n} пар, что ещё меньше. Наконец, четвертое условие бракует не более $2^{1{,}1n} \times 2^{1{,}1n}$ пар для каждого из не более чем $2^{1{,}1n}$ значений z, то есть всего не более $2^{3{,}3n}$ пар, что тоже очень мало по сравнению с общим числом пар.

Сложность найденной таким образом (первой в порядке перебора) пары будет не меньше $3n$ и не больше сложности наших списков, то есть $3n + O(\log n)$; сложность каждого слова будет не меньше $2n$ и не больше его длины, то есть $2n + O(1)$, и построение гарантирует, что общая информация не выделяется. $\triangleright$

Видно, что в нашем рассуждении есть «запас»; проведём подсчёты более аккуратно. Пусть мы хотим построить слова x и y сложности $2n$ с общей информацией n, но так, чтобы нельзя было найти слово z, для которого одновременно

$$\mathit{KS}(z) < \alpha,\ \ \mathit{KS}(x|z) < \beta \text{ и } \mathit{KS}(y|z) < \gamma.$$

Список всех пар u, v с $\mathit{KS}(u) < \alpha$ и $\mathit{KS}(v|u) < \beta$ имеет сложность $\alpha+\beta$, поэтому возникает условие $\alpha+\beta < 3n$. Аналогично появляется условие $\alpha+\gamma < 3n$. Наконец, для существования пары нужно, чтобы запрещённых пар было меньше 2^{4n}; это даёт условие $\alpha+\beta+\gamma < 4n$. Если все эти условия выполнены, то искомую пару x, y построить можно.

Более того, мы можем параллельно проводить отбраковку пар для всех троек целых чисел α, β, γ, удовлетворяющих указанным неравенствам. Таких троек $O(n^3)$, поэтому если неравенства выполняются с логарифмическим запасом, то останутся неотбракованные пары; сложность объединения полиномиального числа перечислимых списков также увеличится на $O(\log n)$. Получаем такой результат:

Theorem 197. `[muchnik-example-general]` *Для любого n существуют слова x и y сложности $2n+O(\log n)$, для которых $KS(x,y) = 3n+O(\log n)$, и при этом для любого слова z выполнено хотя бы одно из трёх неравенств:*

(а) $KS(z) + KS(x|z) \geqslant 3n - O(\log n)$;

(б) $KS(z) + KS(y|z) \geqslant 3n - O(\log n)$;

(в) $KS(z) + KS(x|z) + KS(y|z) \geqslant 4n - O(\log n)$.

Более подробно утверждение теоремы формулируется так: для некоторой константы c и для всех n существуют слова x и y, сложности которых отличаются от $2n$ не более чем на $c\log n$, сложность пары отличается от $3n$ не более чем на $c\log n$ и для любого слова z выполнено одно из трёх неравенств с заменой $O(\log n)$ в правой части на $c\log n$.

Построенная в этой теореме пара слов является «наихудшей» с точки зрения возможности выделения общей информации. Это можно уточнить следующим образом. Для данной пары x, y можно рассмотреть множество $C(x,y) \subset \mathbb{N}^3$, состоящее из всех троек α, β, γ, для которых существует слово z с

$$KS(z) < \alpha,\ \ KS(x|z) < \beta \text{ и } KS(y|z) < \gamma.$$

Это множество «наследственно вверх» (вместе с каждой тройкой содержит и все покоординатно бóльшие) и зависит от пары x, y; даже если ограничиться парами с $KS(x) = KS(y) = 2n$ и $KS(x,y) = 3n$, соответствующие множества могут сильно различаться. При этом для пары пересекающихся подслов несжимаемого слова множество $C(x,y)$ будет максимально возможным, а для построенного в теореме 197 — минимально возможным.

Рассмотрим этот вопрос более детально. Пусть даны слова x, y сложности $2n$ с общей информацией n (как обычно, мы допускаем отклонения порядка $O(\log n)$, не оговаривая этого особо). Любая тройка $\langle\alpha,\beta,\gamma\rangle$, принадлежащая $C(x,y)$, должна удовлетворять (с логарифмической точностью) очевидным неравенствам

$$\alpha+\beta \geqslant 2n;\ \ \alpha+\gamma \geqslant 2n;\ \ \alpha+\beta+\gamma \geqslant 3n$$

(поскольку $KS(x) \leqslant KS(z) + KS(x|z)$ и так далее). Это означает, что множество $C(x,y)$ содержится в множестве C_M всех троек, удовлетворяющих этим неравенствам. (И здесь аккуратная формулировка требовала бы оговорок о логарифмических погрешностях, которые мы для краткости опускаем.)

Как изобразить это множество наглядно? При каждых β и γ есть некоторое минимальное значение $\alpha_0(\beta,\gamma)$, начиная с которого $\langle\alpha,\beta,\gamma\rangle$ попадает в C_M. График функции $\langle\beta,\gamma\rangle \mapsto \alpha_0(\beta,\gamma)$, представляет собой поверхность, состоящую из трёх плоских частей (соответствующих трём неравенствам), которую можно изобразить с помощью линий уровня (рис. 29). Каждая линия уровня соответствует некоторому значению α и отделяет на плоскости β,γ удовлетворяющие неравенствам пары от не удовлетворяющих.

Глядя на эту картинку, можно убедиться, что для случая пары $\langle x,y\rangle$ с полностью выделяемой общей информацией (пересекающиеся куски несжимаемого слова) множество $C(x,y)$ равно C_M, так что верхняя оценка C_M для этого множества достижима. Несложно понять, как нужно выбирать слово z для данных α, β и γ: при $\alpha < 1$ выбираем z внутри общей части, тем самым сокращая условную сложность обоих слов x и y на α. При $\alpha > 1$, помимо общей части, можно присоединить к z куски из остатков слов x и y (в произвольной пропорции; изменение пропорции соответствует сдвигу по наклонному отрезку на рисунке).

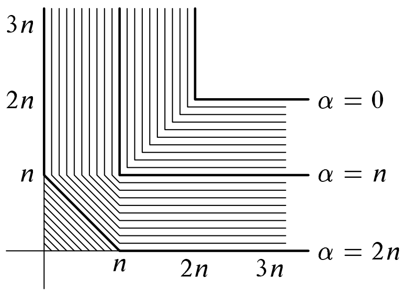

Рис. 29. Множество C_M

[cominf.1]

Теорема 197 даёт пример пары x, y с меньшим множеством $C(x, y)$. В самом деле, для построенной в этой теоремы пары множество $C(x, y)$ содержится не только в C_M, но и в объединении множеств решений неравенств

$$\alpha + \beta \geqslant 3n;\ \alpha + \gamma \geqslant 3n;\ \alpha + \beta + \gamma \geqslant 4n$$

(соответствующих неравенствам (а)–(в) этой теоремы). Пересекая C_M с этим объединением, мы получаем меньшее множество, которое мы назовём C_m; оно изображено на рисунке 30.

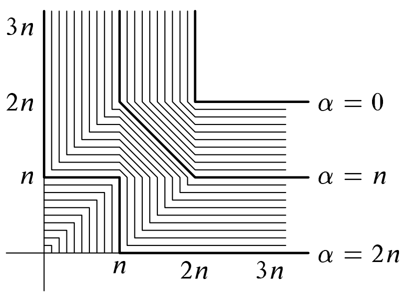

Рис. 30. Множество C_m

[cominf.2]

Таким образом, для построенной в теореме пары x, y множество $C(x, y)$ содержится в C_m. Оказывается, что оно совпадает с C_m; более того, для любой пары x, y (а не только для пары из теоремы) множество C_m содержится в $C(x, y)$.

Чтобы проверить это, надо для каждой точки $\langle \alpha, \beta, \gamma \rangle$ из C_m указать подходящее z. Достаточно делать это для минимальных троек (поскольку увеличение не выводит из C_m). Точки на наклонных линиях на рисунке соответствуют словам z, которые соединяют часть слова x с частью слова y в некоторой пропорции. Например, точка $(\frac{3}{2}n, \frac{3}{2}n)$ при $\alpha = n$ соответствует слову z длины n, состоящему из двух половин; левая половина содержит $n/2$ битов из кратчайшего описания слова x, а правая содержит $n/2$ битов кратчайшего описания слова y. Аналогичным образом точка $(n.n)$ соответствует слову длины $2n$, содержащему по n битов того и другого описаний. Точка $(n+h,h)$ (для h от 0 до n) соответствует слову z длины $2n-h$, являющемуся началом кратчайшего описания слова y. Тогда $KS(y|z)$ равно числу битов этого описания, не вошедших в z, то есть h. С другой стороны, $KS(x|z)$ не превосходит $KS(x,y|z)$; сложность пары $\langle x,y\rangle$ равна $3n$, слово z имеет нулевую сложность относительно $\langle x,y\rangle$ (и даже относительно y) и сложность $2n-h$, поэтому $KS(x,y|z)\leqslant 3n-(2n-h)=n+h$. Наконец, точки вида $(h,0)$ при h от 0 до n соответствуют словам z сложности $3n-h$, которые содержат полностью всё слово y и $n-h$ битов кратчайшего описания слова x при известном y.

В итоге получаем такое утверждение:

Theorem 198. [cominf-region] *Для любой пары слов x,y множество $C(x,y)$ заключено (с обычными оговорками о логарифмической точности) между нижней оценкой C_m и верхней C_M; обе оценки достигаются для некоторых пар.*

Зная множество $C(x,y)$ для данных слов x,y, можно уже чисто формально получать разные следствия. Вот один пример:

Theorem 199. *Пусть $C(x,y)=C_M$ (пара из теоремы 197). Тогда для любого слова z выполняется неравенство*

$$KS(z)\leqslant 2\,KS(z|x)+2\,KS(z|y)$$

$\lhd$ Можно считать, что $KS(z)=O(n)$ (для слов большой сложности $KS(z|x)$ и $KS(z|y)$ близки к $KS(z)$ и уж заведомо больше половины $KS(z)$).

Перепишем неравенство, выразив $KS(z|x)$ и $KS(z|y)$ через величины, входящие в определение $C(x,y)$:

$$KS(z)\leqslant 2\,KS(x,z)-2\,KS(x)+2\,KS(y,z)-2\,KS(y)$$

и далее

$$KS(z)\leqslant 2\,KS(z)+2\,KS(x|z)-2\,KS(x)+2\,KS(z)+2\,KS(y|z)-2\,KS(y),$$

то есть

$$2\,KS(x)+2\,KS(y)\leqslant 3\,KS(z)+2\,KS(x|z)+2\,KS(y|z).$$

Левая часть равна $8n$, а правая не может быть меньше $8n$, в чём легко убедиться, рассматривая каждую линию уровня (причём на ней достаточно проверить точки с минимальной суммой координат). $\rhd$

$\boxed{233}$ Покажите, что при малых значениях $KS(z)$ верна лучшая оценка

$$KS(z)\leqslant KS(z|x)+KS(z|y),$$

но если не ограничивать z, то константу 2 улучшить нельзя.

11.3. Комбинаторный смысл общей информации

[cominf-combin] Построенный в предыдущем пример пары слов, из которых не выделяется общая информация, всё же не даёт наглядного объяснения, чем именно отличаются такие слова. Полностью удовлетворительного ответа на этот (неформальный) вопрос мы не знаем, но некоторые наблюдения сделать можно.

Что означает, что для данных слов x и y найдётся слово z, при котором $KS(z) < \alpha$, $KS(x|z) < \beta$ и $KS(y|z) < \gamma$? Обозначим через $U_n(z)$ множество всех слов, сложность которых при известном z меньше n. Таких слов (примерно) 2^n. Наше условие означает, что пара $\langle x, y\rangle$ принадлежит одному из множеств $U_\beta(z)\times U_\gamma(z)$ при $KS(z) < \alpha$; всего таких множеств имеется (примерно) 2^α.

Таким образом, условие $\langle \alpha, \beta, \gamma\rangle \in C(x, y)$ означает, что пара $\langle x, y\rangle$ принадлежит объединению перечислимого семейства из 2^α «прямоугольников» размера $2^\beta \times 2^\gamma$ (прямоугольником мы называем декартово произведение любых двух множеств слов). Верно и обратное: если пара $\langle x, y\rangle$ покрывается объединением перечислимого семейства из 2^α прямоугольников размера $2^\beta \times 2^\gamma$, то тройка $\langle \alpha, \beta, \gamma\rangle$ (плюс сложность порождающего это семейство алгоритма и логарифмическая добавка) принадлежит $C(x, y)$. Таким образом, можно сказать, что множество $C(x, y)$ для данной пары зависит от того, какие (порождаемые простыми алгоритмами) перечислимые семейства прямоугольников её покрывают.

Отсюда ясен план построения другого примера пары слов, из которых не выделяется информация: надо взять множество пар, плохо покрываемое прямоугольниками, а потом взять его случайный элемент.

Удобно представлять себе множества пар как бинарные отношения, или двудольные графы, изображая пару $\langle x, y\rangle$ как ребро, соединяющее точку x в левой доле графа с точкой y в его правой доле. Прямоугольник, таким образом, задаётся множествами вершин слева и справа и покрывает все рёбра, соединяющие вершины этих множеств.

Свойство, которое будет гарантировать, что граф плохо покрывается прямоугольниками, совсем простое: мы будем рассматривать графы, в которых нет четырёхугольников (не существует таких вершин $a.b$ в одной доле и c, d в другой, что все четыре ребра ac, ad, bc, bd входят в граф). Имеет место следующая простая комбинаторная

[lemma-rectangles] **Лемма**. Рассмотрим произвольный двудольный граф с l вершинами слева и L вершинами справа (считая для определённости, что $l \leqslant L$). Если этот граф не содержит четырёхугольников, то плотность рёбер в нём (число рёбер, делённое на Ll) не превосходит большей из оценок $O(1)/\sqrt{L}$ и $O(1)/l$.

(Другими словами, если в прямоугольной таблице расставлены звёздочки так, что никакие четыре из них не стоят в вершинах прямоугольника, параллельного сторонам таблицы, то доля звёздочек не превосходит $O(1)$, делённого либо на меньшую сторону, либо на квадратный корень из большей стороны.)

Доказательство леммы. Каждой из l вершин с левой стороны соответствует множество её соседей справа. Условие леммы (отсутствие четырёхугольников) означает, что пересечение любых двух таких множеств содержит не более одного элемента. Формула включений и исключений позволяет в таком случае оценить число элементов в объединении: оно не меньше суммы размеров всех множеств (эта сумма есть общее число рёбер графа) минус число пар (которое не больше l^2). С другой стороны, объединение содержит не более L элементов.

Отсюда мы заключаем, что общее число рёбер графа не превосходит $L+l^2$, а их плотность не больше $1/l + l/L$. Это даёт необходимую оценку при $l \leqslant \sqrt{L}$, когда первый член больше второго. Если же $l \geqslant \sqrt{L}$, то мы получаем оценку $O(1)l/L$, что недостаточно (мы хотим $O(1)/\sqrt{L}$). Чтобы получить требуемое, заметим, что мы можем оставить от графа лишь $\sqrt{L}$ вершин в левой доле, для которых число соседей максимально. От этого его плотность лишь возрастёт, а после этого наше рассуждение даёт необходимую оценку. Лемма доказана.

Теперь построим граф без четырёхугольников и установим с помощью только что доказанной леммы, что он плохо покрывается прямоугольниками.

Такой граф проще всего построить геометрически. Рассмотрим произвольное конечное поле $\mathbb{F}$ и плоскость (двумерное векторное пространство) над этим полем. Элементами левой доли будут точки плоскости, элементами правой доли — прямые. Четырёхугольников нет, потому что (как учил ещё Евклид) через две различные точки проходит не более одной прямой.

Оценим число вершин и рёбер такого графа. Если поле содержит (примерно) 2^n элементов, то вершин с каждой стороны будет примерно 2^{2n}, а рёбер — примерно 2^{3n} (на каждой прямой примерно 2^n точек, и через каждую точку проходит примерно 2^n прямых). Таким образом, для большинства ребер (x, y) этого графа сложности $\mathit{KS}(x)$ и $\mathit{KS}(y)$ близки к $2n$, сложность пары $\mathit{KS}(x, y)$ близка к $3n$, а $I(x:y)$ примерно равно n.

234 Покажите, что $I(x:y) = n + O(\log n)$ для всех рёбер этого графа, сложность которых больше $3n - O(\log n)$.

Чтобы дать новое доказательство теоремы 196, посмотрим, какую часть построенного графа (множества пар) можно покрыть $2^{1,1n}$ прямоугольниками размера $2^{1,1n} \times 2^{1,1n}$ (мы используем те же) значения параметров α, β и γ, что и теореме 196). К каждому такому прямоугольнику можно применить доказанную лемму, по которой плотность рёбер в нём не превосходит $2^{-0,55n}$, так что общее число рёбер, покрытых всеми прямоугольниками, не больше

$$2^{1,1n} \times 2^{1,1n} \times 2^{1,1n} \times 2^{-0,55n} = 2^{2,75n} \ll 2^{3n}$$

Поэтому подавляющее большинство рёбер не покрыто (а также имеет нужные сложности, как мы говорили) и любое из таких рёбер является примером, существование которого утверждает теорема 196. Таким образом, мы получаем новое доказательство этой теоремы.

235 Покажите, что любое ребро (x, y), сложность которого близка к $3n$, не допускает выделение общей информации (в смысле теоремы 196). [Указание: множество покрытых рёбер перечисляется простым алгоритмом.]

В этом построении (если рассматривать его как альтернативное доказательство теоремы 196) есть один тонкий момент: нам нужно иметь поле из 2^n элементов (или хотя бы из близкого числа элементов). Поле из 2^n элементов может быть построено как расширение степени n поля из двух элементов (поле разложения многочлена $x^{2^n} - x$); кроме того, можно воспользоваться известным из теории чисел фактом, что между m и $2m$ всегда есть простое число («постулат Бертрана», который для бóльших m следует, например, из теоремы о плотности распределения простых чисел).

Тем самым пара слов, из которых не выделяется общая информация, приобретает конкретные очертания: первое из них — это точка конечной плоскости над конечным полем, а второе — проходящая через неё прямая, причём берётся случайная из таких пар.

Более симметрично было бы перейти к проективной плоскости (что мало влияет на сложность, так как бесконечно удалённые точки составляют небольшую долю). Или можно говорить о случайной паре ортогональных (относительно скалярного произведения произведения $x_1y_1 + x_2y_2 + x_3z_3$) прямых в трёхмерном векторном пространстве над конечным полем.

[ВОПРОС: можно ли построить аналогичный пример на основе точек на вещественной сфере (прямых в трёхмерном пространстве)? Можно дискретизировать с каким-то шагом, но возможная проблема: когда x и y близки, то им сразу много точек почти ортогональны, надо это как-то оценивать.]

Как и раньше, наугад взятые коэффициенты 1,1, конечно, не являются наилучшими. Ровно то же самое рассуждение можно проводить для любых α, β, γ, и оно окончится успешно (то есть число покрытых рёбер будет меньше $3n$), если числа α, β, γ не слишком велики. Применяя лемму (ровно в той форме, как она сформулирована), мы получаем, что множество $C(x, y)$ для нашего примера (случайной пары точка – прямая) содержится (с точностью до $O(\log n)$) в множестве M, граница которого изображёна (в виде линий уровня) на рисунке 31. (Линия уровня для $\alpha = 3n$ состоит из начала координат.)

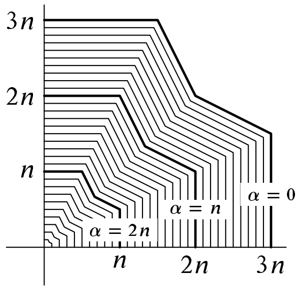

Рис. 31. Множество C_M

[cominf.3]

Это множество можно ещё пересечь с C_M (поскольку $C(x, y) \subset C_M$ для любой пары (x, y); получатся более замысловатые линии уровня, которые мы рисовать не будем.

Заметим, что в отличие от ранее приведённых примеров, это лишь верхняя оценка для множества $C(x, y)$; каково это множество в точности, мы не знаем.

[Упражнение: теорема Чернова о конечном поле особого вида и точке на границе $C(x, y)$.]

Как и в предыдущем разделе, из доказанных утверждений следует неравенство, оценивающее безусловную сложность z через условные:

Theorem 200. *Пусть x, y — случайная пара точка – прямая на плоскости над конечным полем из (примерно) 2^n элементов. Тогда для любого слова z выполняется неравенство*

$$KS(z) \leqslant 2\,KS(z|x) + 2\,KS(z|y) + O(\log n).$$

$\lhd$ Как и раньше, это сводится к проверке неравенства

$$8n \leqslant 3\alpha + 2\beta + 2\gamma$$

для всех $\langle \alpha, \beta, \gamma \rangle \in C(x, y)$. Но это неравенство верно для всех точек множества $C \cap C_M$, как можно проверить по рисунку. $\triangleright$

Итак, у нас есть два разных построения пар с невыделяемой общей информацией. Первое из них даёт лучшие оценки для множества $C(x, y)$; в чём же тогда преимущества второго? Наиболее явное, хотя и неформализуемое преимущество состоит в том, что мы нашли комбинаторную причину невыделяемости общей информации (существование графов, плохо покрываемых прямоугольниками), а также простое достаточное условие для этого (отсутствие четырёхугольников).

Более формально можно сказать, что во втором случае мы построили пример, являющийся «стохастическим»: наша пара является элементом максимальной сложности в некотором простом множестве.

[ВОПРОС: как доказать, что в первом случае (неконструктивное доказательство) пара заведомо нестохастическая?]

Ещё один способ объяснить, чем второе доказательство лучше — это рассмотреть сложность с оракулом. Фактически мы доказали, что для любого оракула A существуют слова x, y сложности (обычной) $2n$ с общей информацией n, для которых не существует слова z с $KS^A(z) < 1{,}1n$, $KS^A(x|z) < 1{,}1n$ и $KS^A(y|z) < 1{,}1n$. В самом деле, для любого оракула получаются также прямоугольники, и потому они покрывают лишь небольшую часть множества пар, и в качестве x, y нужно взять какую-либо непокрытую (и не являющуюся простой) пару. (Естественно, результат будет зависеть от оракула, поскольку любая пара становится простой при подходящем оракуле.)

$\boxed{\textbf{236}}$ Сформулируйте аналогичное утверждение, используя вместо оракула дополнительное условие u произвольной сложности.

Аналогичное рассуждение применимо и к другим алгебраическим конструкциям. Можно, например, рассмотреть пару случайных перпендикулярных одномерных подпространств в четырёхмерном пространстве или (что то же самое с точностью до бесконечно удалённых точек) случайную пару (точка трёхмерного пространства, проходящая через неё плоскость). Про возникающий граф, правда, уже нельзя утверждать, что в нём нет четырёхугольников: для любых двух одномерных подпространств есть двумерное подпространство, им ортогональное, и любое его одномерное подпространство будет соединено с обоими подпространствами левой доли. Тем не менее таких одномерных подпространств довольно мало, и можно применить аналогичную оценку с формулой включений и исключений.

[Сюда хорошо бы вписать подробности или хотя бы подробное упражнение. А также про другие примеры (две пересекающиеся прямые? Почему, кстати, в этих случаях удаётся получить оценку $KS(z) \leqslant O(KS(z|x) + KS(z|y))$ непосредственно, а в других нет?

Графы и итерации налево-направо (Ромащенко?)]

[Ещё хорошо бы написать о том, как построить стохастический пример с минимальным множеством — надо доказывать с помощью оценок вероятности, что множество с какими-то свойствами существует, но с какими? вроде как если хотеть, чтобы в каждом прямоугольнике доля была бы близка к средней, то не получается.]

11.4. Условная независимость и общая информация

Существует ещё один способ построения пары слов, из которых не выделяется общая информация, довольно загадочный. Для начала вспомним неравенство задачи 222 (с. 307):

$$H(\xi) \leqslant H(\xi|\alpha) + H(\xi|\beta) + I(\alpha : \beta),$$

а точнее, соответствующее неравенство для колмогоровских сложностей:

$$KS(z) \leqslant KS(z|x) + KS(z|y) + I(x : y).$$

(логарифмические члены мы опускаем для краткости). В случае, когда $I(x : y) = 0$, из него следует оценка сложности через условные сложности:

$$KS(z) \leqslant KS(z|x) + KS(z|y),$$

что не удивительно: если взаимной информации нет, так она и не выделяется. Казалось бы, это ничего дать не может. Но оказывается, что аналогичную оценку можно получить и в случае, когда x и y *условно независимы*, то есть существуют слова u и v, для которых $I(x : y|u) = 0$, $I(x : y|v) = 0$ и $I(u : v) = 0$. А именно, имеет место такое неравенство:

Theorem 201. [iterated-inequality] *Неравенство*

$$KS(z) \leqslant 2\,KS(z|x) + 2\,KS(z|y) + I(x : y|u) + I(x : y|v) + I(u : v)$$

справедливо для произвольных слов x, y, z, u, v *с точностью до* $O(\log KS(x, y, u, z, v))$.

Заметим, что это неравенство можно было бы вывести из предыдущего и из неравенства Инглетона

$$I(x : y) \leqslant I(x : y|u) + I(x : y|v) + I(u : v),$$

но беда в том, что для произвольных слов неравенство Инглетона может не выполняться. Поэтому придётся действовать в другом порядке.

◁ Начнём с уже надоевшего нам неравенства

$$KS(z) \leqslant KS(z|u) + KS(z|v) + I(u : v)$$

и оценим $KS(z|u)$ и $KS(z|v)$, пользуясь релятивизованными аналогами того же самого неравенства:

$$KS(z|u) \leqslant KS(z|x, u) + KS(z|y, u) + I(x : y|u)$$

и

$$KS(z|v) \leqslant KS(z|x, v) + KS(z|y, v) + I(x : y|v)$$

После этого остаётся лишь заметить, что добавление нового условия лишь уменьшает сложность: $KS(z|x, u) \leqslant KS(z|x)$ и т.п. ▷

Эту теорему можно применить для построения слов с плохо выделяемой общей информацией. Возьмём условно независимые, но не независимые случайные величины (теорема 191, с. 309). Пусть это величины α и β, и они независимы при известных γ или δ, причём γ и δ независимы.

Произведём N независимых испытаний четвёрки величин $\langle\alpha,\beta,\gamma,\delta\rangle$ и получим четыре слова x,y,u,v (точнее, получим четыре случайные величины, значениями которых являются слова). Как мы знаем из раздела 7.3.4 (теорема 125), с близкой к единице вероятностью сложности этих случайных слов и их комбинаций равны энтропиям соответствующих случайных величин (умноженным на N) с точностью $O(\sqrt{N})$. Поэтому для любого N можно найти четыре слова x,y,u,v, для которых

$$I(x:y|u)=O(\sqrt{N}),\ I(x:y|v)=O(\sqrt{N}),\ I(u:v|u)=O(\sqrt{N}),$$

и при этом

$$I(x:y)=NI(\alpha:\beta)+O(\sqrt{N}),$$

причём, напомним, $I(\alpha:\beta)\neq 0$. Остаётся воспользоваться теоремой 201 и заключить, что для любого z выполняется неравенство

$$\mathit{KS}(z)\leqslant 2\,\mathit{KS}(z|x)+2\,\mathit{KS}(z|y)$$

с точностью до $O(\sqrt{N})$, в то время как сложности слов x и y и их взаимная информация с той же точностью пропорциональны первой степени N (с ненулевыми коэффициентами пропорциональности)

Таким образом, мы получаем новую конструкцию слов с невыделяемой взаимной информацией. Эти слова, как и в нашем геометрическом примере, являются стохастическими, и удивительным образом никаких комбинаторных оценок нам делать не пришлось. Правда, мы имеем лишь более слабое утверждение (с корнем вместо логарифма).

[Здесь хорошо бы разъяснить историческую связь с теоремой Гача, объяснив, что результат Макарычевых об итерированной условной независимости позволяет получить это для любых неблочных случайных величин]

12. Алгоритмическая теория информации для нескольких источников

[multi]

12.1. Постановка задачи о передаче информации

Рассмотрим ориентированный граф, рёбра которого играют роль «линий связи», а вершины — узлов обработки информации. В некоторые вершины извне поступает какая-то информация; в результате обработки информации (в вершинах) и её передачи (по рёбрам) некоторая другая информация должна попасть в заданные вершины. Такая задача традиционно рассматривается в шенноновской теории информации с несколькими источниками; здесь мы хотим проанализировать её с точки зрения алгоритмической теории информации.

Более формально. Пусть задан некоторый конечный ориентированный ациклический граф. Некоторые его вершины объявлены *входами*, и для каждого входа указано *входное* слово. Другие его вершины объявлены *выходами*, и для каждой из них указано *выходное* слово. Возникает задача: на каждом ребре графа написать некоторое двоичное слово (*передаваемое* по данному ребру), причём для каждой вершины v должно выполняться такое условие:

> если X — любое выходящее из вершины v слово (указанное изначально выходное слово или передаваемое по исходящим из v рёбрам), а $Y_1, \ldots, Y_k$ — все входящие в эту вершину слова (заданные изначально входные слова или передаваемые по входящим в v рёбрам), то
>
> $$KS(X|Y_1, \ldots, Y_k) \approx 0.$$

Неформально говоря, это условие означает, что в вершине v происходит лишь переработка информации, а никакой новой информации не возникает.

Приближённое равенство нулю, как обычно, означает, что указанная сложность есть $O(\log N)$, где N — суммарная длина всех входящих и выходящих слов, указанных в формулировке задачи. Тем самым мы фактически рассматриваем не одну такую задачу передачи информации, а семейство задач, соответствующих разным значениям N.

При этом мы будем накладывать ограничения на «пропускные способности» рёбер, то есть длины передаваемых по ним слов. Такие ограничения будут накладываться лишь на некоторые рёбра; пропускная способность остальных не ограничивается. Нас будет интересовать, возможно ли решить задачу при указанных ограничениях.

Начнём с самого простого примера: пусть имеется граф, состоящий из двух вершин и соединяющего их ребра (рис. 32; договоримся сразу, что на наших рисунках все рёбра ведут сверху вниз)

В верхнюю вершину входит слово A, а из нижней должно выйти слово B, при этом соединяющий эти вершины канал связи имеет пропускную способность k. Другими словами, мы для данных слов A и B ищем такое слово X, что

$$KS(X|A) \approx 0\,; \quad KS(B|X) \approx 0\,; \quad l(X) \leqslant k.$$

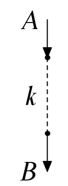

Рис. 32. Простейшая задача передачи информации.

[multi-pic1]

Ясно, что это возможно только при $KS(B|A) \approx 0$ и $KS(B) \leqslant k$ (последнее неравенство также понимается «с точностью до логарифма»); с другой стороны, эти необходимые условия являются также и достаточными, так как в качестве X можно взять кратчайшее описание для B.

Более формально это можно сказать так: пусть имеется последовательность слов A_n и B_n, а также последовательность натуральных чисел k_n, причём длины слов A_n и B_n, а также числа k_n, ограничены сверху некоторым многочленом от n. Тогда следующие два свойства равносильны:

(1) существует последовательность слов X_n, для которых $l(X_n) \leqslant k_n + O(\log n)$, а также $KS(X_n|A_n) = O(\log n)$ и $KS(B_n|X_n) = O(\log n)$;

(2) $KS(B_n|A_n) = O(\log n)$ и $KS(B_n) \leqslant k_n + O(\log n)$.

Эквивалентность (1) и (2) вытекает из того, что

$$KS(B|A) \leqslant KS(B|X) + KS(X|A) + O(\log KS(A,B,X));$$
$$KS(B) \leqslant l(X) + KS(B|X) + O(\log KS(B|X))$$

при любых A, B, X (отсюда следует импликация $(1) \Rightarrow (2)$), а также того, что

для любых A, B, k найдётся слово X, при котором $l(X) \leqslant KS(B)$, $KS(X|A) \leqslant KS(B|A) + O(\log KS(B))$ и $KS(B|X) = O(1)$,

что даёт импликацию $(2) \Rightarrow (1)$; во втором случае в качестве X надо взять кратчайшее описание слова B.

При $A = B$ наше утверждение приобретает совсем простой вид: передача слова A по каналу связи с пропускной способностью k возможна, когда сложность слова A не превосходит k.

Перейдём к более содержательным примерам.

12.2. Условное кодирование

Следующая задача может быть названа «задачей передачи слова A при известном слове B» (рис. 33). В ней требуется закодировать слово A не более чем k битами, передать и

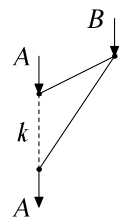

Рис. 33. Передача A при известном B.

[multi-pic2]

раскодировать обратно; при этом и при кодировании, и при декодировании известно слово B (пропускная способность рёбер, изображённых на рисунке сплошными линиями, не ограничивается, так что по ним можно передать слово B полностью).

Эта задача, как легко видеть, разрешима тогда и только тогда, когда $KS(A|B) \leqslant k$. В самом деле, в пункте декодирования, помимо слова B (или какого-то слова, из него полученного, и не содержащего новой по сравнению с B информации) известны k битов, поэтому раскодирование возможно, лишь если $KS(A|B) \leqslant k$. С другой стороны, если $KS(A|B) \leqslant k$, то в качестве слова X, передаваемого по каналу ограниченной пропускной способности, надо взять кратчайшее описание B при известном A, а по двум другим каналам передавать слово B. Заметим, что сложность кратчайшего описания относительно пары $\langle A, B\rangle$ логарифмическая, так как зная длину этого описания, можно параллельно пробовать все слова этой длины, пока не найдётся одно из кратчайших описаний.

237 В этом рассуждении достаточно одного кратчайшего описания; покажите, тем не менее, что *все* кратчайшие описания A при известном B имеют относительно пары $\langle A, B\rangle$ сложность, не превосходящую $O(\log KS(A, B))$.

238 Дайте точную формулировку приведённого только что критерия возможности передачи слова A при известном B по аналогии с разобранным ранее примером, рассмотрев последовательности A_n, B_n и k_n, и докажите его.

12.3. Условное кодирование: теорема Мучника

[multi-muchnik1]

Следующий замечательный результат Ан. А. Мучника [58] является аналогом теоремы Вольфа – Слепяна из шенноновской теории информации и говорит, что в предыдущем примере не обязательно использовать слово B при кодировании. В нём рассматривается граф, получающийся из предыдущего удалением одного ребра (рис. 34).

Оказывается, что эта задача разрешима при тех же условиях, что и предыдущая, то есть при $KS(A|B) \leqslant k$. Необходимость этих условий очевидна (граф стал только меньше); вот точная формулировка утверждения о достаточности.

Theorem 202. [multi-muchnik-th1] *Пусть A и B — произвольные слова сложности не более n. Тогда найдётся слово X длины не более $KS(A|B) + O(\log n)$, при котором $KS(X|A) = O(\log n)$ и $KS(A|B, X) = O(\log n)$.*

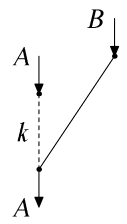

Рис. 34. Передача A при известном B: теорема Мучника.
[multi-pic3]

Имеется в виду, что скрытая в $O(\log n)$ константа не зависит от n и от выбора слов A и B.

Теорему Мучника можно переформулировать так: для любых слов A и B существует программа, переводящая B в A, имеющая логарифмическую сложность относительно A и безусловную сложность $\mathit{KS}(A|B)$ (плюс логарифмическая добавка). Другими словами, дополнительное ограничение, состоящее в простоте программы относительно A, приводит к росту (безусловной) сложности программы не более чем на $O(\log \mathit{KS}(A, B))$.

$\triangleleft$ Пусть слово A имеет сложность a. Заменим это слово на его (кратчайшее) описание длины a. От этого величины $\mathit{KS}(A|B)$, $\mathit{KS}(X|A)$ и $\mathit{KS}(A|B, X)$ изменятся не более чем на $O(\log n)$, поэтому мы можем далее предполагать, что слово A имеет длину (а не сложность) a.

Пусть сложность $\mathit{KS}(A|B)$ равна m. В самом первом приближении идею доказательства можно объяснить так. Рассмотрим хеш-функцию $\chi\colon \mathbb{B}^a \to \mathbb{B}^m$, которая каждому слову длины a ставит в соответствие хеш-значение (отпечаток, fingerprint) длины m.

Для данного слова B имеется 2^m слов Z длины a, для которых $\mathit{KS}(Z|B) \leqslant m$ (на самом деле их $O(2^m)$, но мы пока что пренебрегаем такими мелочами). Пусть $S_B \subset \mathbb{B}^a$ — множество этих слов. Слово A — один из элементов множества S_B.

Пусть хеширование прошло исключительно удачно в том смысле, что у всех слов из S_B хеш-значения оказались различными. Тогда любое слово $P \in S_B$ может быть однозначно восстановлено по $\chi(P)$ при известном B (и χ). Поэтому можно в качестве X, существование которого утверждается в теореме, взять $\chi(A)$: оно имеет правильную длину, просто относительно A (если функция χ проста) и вместе с B позволяет восстановить A (перечисляем множество S_B, пока не обнаружится слово с правильным хеш-значением).

Разумеется, так просто всё не выходит. Какова бы ни была хеш-функция χ, при $a > m$ найдётся как минимум 2^{a-m} слов, имеющих одно и то же хеш-значение (а при простой χ можно найти много простых слов с одним хеш-значением, и они будут просты относительно B при любом B), так что хеширование не будет удачным.

Поэтому мы модифицируем этот план и будем сопоставлять с каждым словом $Z \in \mathbb{B}^a$ не одно хеш-значение, а несколько (в количестве $\mathrm{poly}(n)$). Таким образом, вместо хеш-функции мы рассматриваем двудольный граф $E \subset \mathbb{B}^a \times \mathbb{B}^m$, в котором каждая вершина Z в левой доле (из $\mathbb{B}^a$) имеет $\mathrm{poly}(n)$ соседей в правой доле. Этих соседей мы будем называть *отпечатками* вершины x.

Доказывая теорему, мы будем искать слово X среди отпечатков слова A. Это гарантирует нам, что $\mathit{KS}(X|A) = O(\log n)$, если сам граф E прост (имеет сложность $O(\log n)$).

В самом деле, для задания X при известном A достаточно указать порядковый номер X среди отпечатков слова A.

Если для данного $A \in S_B$ найдётся отпечаток X, уникальный для A внутри S_B (это значит, что в S_B нет других слов, имеющих X своим отпечатком), то слово A можно восстановить, перечисляя S_B и ожидая появления слова, имеющего X среди своих отпечатков. В этом случае $\mathit{KS}(A|B, X) = O(\log n)$ (мы, как и раньше, предполагаем, что граф E имеет сложность $O(\log n)$).

Более того, это верно и в случае, когда в S_B имеется $\mathrm{poly}(n)$ слов с отпечатком X, нужно лишь дополнительно указать порядковый номер A среди этих слов (в порядке перечисления множества S_B), что потребует дополнительных $O(\log n)$ битов.

Таким образом, нам хотелось бы следующего: *у слова A имеется правый сосед, у которого мало* $(\mathrm{poly}(n))$ *левых соседей*. Говоря о соседях, мы рассматриваем сужение графа E на S_B как двудольный граф с левой долей S_B и правой долей $\mathbb{B}^m$. Отметим, что в этом уменьшенном графе число вершин слева и справа примерно одинаково — порядка 2^m.

Это же можно переформулировать так: будем говорить, что вершина справа плоха, если у неё много левых соседей, а вершина слева плоха, если все её соседи справа плохи. Нам хотелось бы, чтобы интересующая нас вершина A не оказалась плохой.

Дальнейший план действий таков: справа плохих вершин мало, так как в графе мало рёбер (число рёбер оцениваем как число вершин слева, умноженное на максимальную степень левых вершин, которая есть $\mathrm{poly}(n)$), а каждая плохая вершина справа даёт большой вклад в число рёбер. Так мы добьёмся, чтобы доля плохих вершин справа была не больше $1/p(n)$ для заданного полинома p (при этом придётся полиномиально увеличить границу, при превышении которой вершина объявляется плохой).

Отсюда можно заключить, что и слева плохих вершин мало, если граф E обладает следующим свойством типа экспандера: *для всякого множества T в левой доле множество $E(T)$ всех соседей всех элементов множества T содержит не меньше элементов, чем само множество T*. В самом деле, возьмём в качестве T множество плохих вершин слева. Все их правые соседи будут плохими, поэтому слева не больше плохих вершин, чем справа. (Заметим, что указанное свойство графа E сохраняется при его сужении на S_B.)

Остаётся объяснить, откуда мы возьмём граф E с нужными свойствами и что делать, если нужное нам слово A попадёт в (пусть и небольшое) число плохих вершин слева.

Как хорошо знают специалисты, существование экспандеров несложно доказать вероятностно (случайный граф обладает нужным свойством с положительной вероятностью); явное их построение — гораздо более сложное и интересное дело, и в последние десятилетия тут есть большие продвижения. Мы, однако, можем обойтись без сложных конструкций с помощью простого трюка. Нужное нам свойство проверяется алгоритмически (за очень большое, но конечное время). Поэтому можно перебирать все графы в каком-либо естественном порядке, пока не обнаружится первый подходящий. Он будет иметь логарифмическую сложность (чтобы организовать перебор, нам достаточно знать размеры множеств). Таким образом из существования *какого-нибудь* графа с нужным свойством автоматически следует существование *простого* графа с тем же свойством.

Последнее: может ли слово A оказаться плохим? Заметим, что плохие слова справа можно алгоритмически перечислять при известных B (и E); по мере обнаружения новых элементов в S_B множество плохих слов справа растёт. Поскольку граф E мы считаем известным, можно перечислять и плохие слова слева. Таким образом, каждое плохое слово можно задать (при известном B) его порядковым номером, что меньше m битов, так как плохих слов заметно меньше 2^m. (Ещё нужно $O(\log n)$ битов, чтобы задать m, a и E, но и с учётом этой добавки получится меньше m, как мы увидим.) А мы с самого начала предполагали, что сложность A при известном B равна m (так выбиралось число m). Поэтому A не может оказаться плохим.

Итак, мы описали схему доказательства теоремы «сверху-вниз». Теперь мы разберём более подробно отдельные шаги этого доказательства, двигаясь «снизу-вверх». Начнём с существования графов со свойствами типа экспандера.

Лемма. Пусть a и m — натуральные числа, причём $a \geqslant m$. Тогда существует двудольный граф $E \subset \mathbb{B}^a \times \mathbb{B}^m$, в котором степень вершин в левой доле ($\mathbb{B}^a$) не превосходит $a + m + 2$, обладающий следующим свойством: для любого множества $T \subset \mathbb{B}^a$, содержащего не более 2^{m-1} элементов, множество $E(T)$ всех правых соседей всех элементов из T содержит больше элементов, чем само T.

Доказательство леммы. Докажем, что случайный граф обладает указанным свойством с положительной вероятностью. Говоря о случайном графе, мы имеем в виду, что для каждой точки в левой доле случайно выбираются $a + m + 2$ соседей, равномерно распределённых в правой доле. При этом этот выбор делается независимо (для разных точек и для разных соседей одной точки).

Если указанное в лемме свойство не выполняется, то существует некоторое (непустое) множество T в левой доле и некоторое множество U в правой доле, для которых $|U| = |T|$, но все соседи элементов из T принадлежат U. Мы подсчитаем вероятность такого события для данных T и U и покажем, что сумма этих вероятностей по всем T и U меньше единицы.

Пусть фиксированы T и U; обозначим число элементов в них через t. По предположению, $t \leqslant 2^{m-1}$, поэтому вероятность того, что случайный элемент правой доли попадёт в U, не больше $1/2$. Вероятность того, что это случится $t(a + m + 2)$ раз при независимых испытаниях (по $a + m + 2$ испытаний для t элементов множества T) не превосходит $2^{-t(a+m+2)}$.

Просуммируем вероятность сначала по всем парам множеств T и U размера t. Число различных T не превосходит $(2^a)^t$ (выбираем t раз один из 2^a элементов; совпадения и возможность перестановки лишь уменьшают число вариантов), число различных U не превосходит $(2^m)^t$. Таким образом, сумма вероятностей по множествам размера t не превосходит

$$2^{mt} \cdot 2^{at} \cdot 2^{-t(m+a+2)} = (1/4)^t.$$

Остаётся заметить, что сумма ряда $\sum (1/4)^t$ (по всем $t \geqslant 1$) меньше 1 (равна $1/3$). Лемма доказана.

Пусть $E_{m,a}$ — первый в каком-нибудь естественном порядке граф, удовлетворяющий лемме (для данных m и a). Его сложность не превосходит $2 \log a + O(1)$ (достаточно указать числа a и m, отведя для них первую и вторую половины описания длины $2 \log a$).

Для данного слова B и для данных m и a рассмотрим множество S_B слов длины a, имеющих относительно B сложность не более m. В нём не более 2^{m+1} слов. Рассмотрим ограничение графа $E_{m,a}$ на S_B. В этом двудольном графе не более $2^{m+1} \cdot (a+m+2) \leqslant a2^{m+3}$ рёбер (каждая из не более чем 2^{m+1} вершин имеет не более чем $a+m+2$ соседей). Объявим *плохими* вершины в правой доле, имеющие не менее a^4 соседей слева. Тогда число плохих вершин будет не более $2^{m+3}/a^3$.

Слева мы объявим *плохими* вершины (слова из S_B), у которых все соседи справа плохие. По построению графа $E_{m,a}$ число плохих вершин слева также не превосходит $2^{m+3}/a^3$. Плохие вершины слева можно перечислять (при известных m, a и B), поэтому плохое слово можно задать его порядковым номером (на что требуется $m-3\log a+O(1)$ битов), так что

$$KS(P|B,m,a) \leqslant m - 3\log a + O(1).$$

для любого плохого слова P. Видно, что запас больше $2\log a$, необходимых для указания m и a, так что все плохие слова имеют условную сложность (относительно B) меньше m и слово A среди них оказаться не может.

Следовательно, у него есть хороший сосед справа. Обозначим его через X. Тогда $KS(X|A) \leqslant (3+\varepsilon)\log a + O(1)$ (надо указать m, a и порядковый номер X среди соседей A в $E_{m,a}$; ε-добавка требуется для кодирования пар). Длина X равна m, то есть $KS(A|B)$. Наконец, $KS(A|B,X)$ не превосходит $(6+\varepsilon)\log a$ (из коих $4\log a$ уходит на указание порядкового номера A среди a^4 левых соседей вершины X, а $2\log a$ требуется для указания m и a; ε нужно для кодирования пар).

Теорема Мучника доказана. $\triangleright$

$\boxed{\mathbf{239}}$ Покажите, что фактически доказано более сильное утверждение, чем объявлено: в условии теоремы требовалось, чтобы слово B имело сложность не более n, но это нигде не использовано.

12.4. Комбинаторный смысл теоремы Мучника

Многие утверждения о колмогоровской сложности имеют комбинаторный эквивалент — равносильное утверждение чисто комбинаторного характера, в котором о сложности ничего не говорится. Во многих случаях эквивалентом является утверждение о существовании выигрышной стратегии в некоторой игре. (См. об этом для случая общей теории алгоритмов в [56].)

[А где написано про это для колмогоровской сложности? может быть, в диссертации Мучника?]

Для доказанной только что теоремы 202 также имеется комбинаторный эквивалент. Рассмотрим для данных чисел a, b и m (считаем, что $m \leqslant a$) игру двух игроков: Математика (**М**) и его Противника (**П**). Игра имеет также параметр c (соответствующий константе в $O(\log n)$, как мы увидим).

Математик имеет право указать для каждого слова A длины a не более чем $c(a+b)^c$ слов длины m, которые он условно именует *простыми относительно* A. Кроме того, он для каждой пары слов B (длины b) и X (длины m) имеет право указать до $c(a+b)^c$ слов длины a, называя их *простыми относительно пары* B, X.

Противник для каждого слова B длины b может указать до 2^m слов длины a, назвав их m-*простыми относительно* B.

Игрок может сделать очередной ход (объявив простыми ещё какие-то слова) в любой момент (независимо от ходов другого игрока). При этом он видит ходы другого игрока, сделанные к этому моменту. Возникающая игра по существу конечна: поскольку объявленные слова нельзя взять назад, партия достигает некоторого предельного положения. (Однако, наблюдая за игрой, невозможно определить, достигнуто ли это предельное положение или ещё нет — игроки сохраняют право хода до бесконечности, даже если им и не пользуются.) В этом предельном положении определяется победитель:

> **М** выиграл, если для всякого слова B длины b и для всякого слова A длины a, которое **П** объявил m-простым относительно B, найдётся слово X длины m, которое **М** объявил простым относительно A и для которого **М** объявил A простым относительно B и X.

Теперь можно сформулировать комбинаторный эквивалент теоремы 202.

Theorem 203. [multi-muchnik-th1a] *Существует такая константа c, что для любых натуральных a, b, m (для которых $m \leqslant a$) в описанной игре с параметрами a, b, m и c выигрывает* **М**.

Объясним, почему это комбинаторное утверждение действительно является комбинаторным эквивалентом теоремы 202. Предположим, что оно верно для некоторого c. Рассмотрим Противника, который делает свои ходы, не обращая внимания на Математика, и объявляет m-простыми слова длины a, которые имеют сложность менее m (относительно соответствующего B). Поведение Противника в этом случае задаётся алгоритмом, для задания которого надо знать значения a, b, m. Выигрышная стратегия Математика (существующая по предположению) может быть найдена перебором (при известных a, b и m). Поэтому слова, которые Математик объявляет простыми, будут и в самом деле иметь малую колмогоровскую сложность (условную) — для их задания, помимо порядкового номера (который записывается $\log c + c\log(a+b)$ битами), достаточно указать a, b, m, что требует ещё $O(\log(a+b))$ битов. Таким образом, мы получаем утверждение теоремы 202. (Техническое пояснение: в оценке $c(a+b)^c$ первый множитель c нужен для малых значений a и b и соответствует слагаемому $O(1)$ в оценках колмогоровской сложности, которое, строго говоря, надо было бы явно добавлять к $O(\log n)$ на случай $n = 1$ и $\log n = 0$.)

В обратную сторону: пусть для любого c комбинаторное утверждение неверно, то есть при некоторых a, b, m Математик проигрывает в игре с параметрами a, b, m, c. В этом случае Противник имеет в этой игре выигрышную стратегию, которую можно найти перебором. Пусть Математик играет против этой стратегии, называя простыми слова X, для которых $\mathit{KS}(X|A) < c\log(a+b) + \log c$, а также слова A, для которых $\mathit{KS}(A|B,X) < c\log(a+b) + \log c$. Ограничения на количество слов, объявляемых простыми, не нарушены. Поэтому проигрыш Математика означает, что найдутся слова A и B длин a и b соответственно, для которых: (1) Противник объявил, что слово A является m-простым относительно B, но (2) не существует X с $\mathit{KS}(X|A) < c\log(a+b) + \log c$ и $\mathit{KS}(A|B,X) < c\log(a+b) + \log c$. Поскольку стратегии Математика и Противника могут быть алгоритмически найдены при известных a, b, m и c, условная сложность $\mathit{KS}(A|B)$ и в самом деле будет мала, не превосходя $m + (3+\varepsilon)\log(a+b) + (1+\varepsilon)\,\mathit{KS}(c) + O(1)$ (каждое из чисел a, b, m содержит не более $\log(a+b)$ битов, ε мы добавляем при соединении беспрефиксных кодов). Можно ограничиться значениями c, являющимися степенями двойки. В этом случае $\mathit{KS}(c) = \log\log c + O(1)$ (замена логарифма на повторный логарифм, как мы увидим дальше, технически существенна).

Поскольку утверждение теоремы 202 мы предполагаем верным (с некоторой константой в $O(\log n)$, которую мы будем обозначать c'), то найдётся слово X' длины $K(A|B) +$

$c' \log(a+b)$, для которого выполняется утверждение этой теоремы. Это слово имеет длину больше m, но ненамного (не более чем на $(c'+3+\varepsilon)\log(a+b)+\log\log c+O(1)$). Оставим от него первые m битов, полученное слово обозначим X. Сложность $KS(X|A)$ превышает $KS(X'|A)$ не более чем на $2\log m + O(1)$, а $KS(X'|A) \leqslant c'\log(a+b)+O(1)$ согласно теореме 202, так что в итоге заведомо

$$KS(X|A) \leqslant (c'+2)\log(a+b)+O(1).$$

Аналогичным образом можно оценить $KS(A|X,B)$: теорема 202 говорит, что $KS(A|X',B)$ не превышает $c'\log(a+b)$, а замена X на X' (удаление не более чем $(c'+4)\log(a+b)+2\log\log c+O(1)$ битов) увеличивает сложность не более чем на $(c'+5)\log(a+b)+3\log\log c+O(1)$ битов. В итогде получается, что

$$KS(A|X,B) \leqslant (2c'+5)\log(a+b)+3\log\log c+O(1).$$

Видно, что при достаточно большом c обе оценки будут меньше $c\log(a+b)+\log c$ (именно здесь важно, что берётся повторный логарифм c), и мы получаем противоречие (оказывается, что выиграл **М**, а не **П**).

Таким образом, мы установили, что комбинаторное утверждение теоремы 203 действительно эквивалентно сложностному утверждению теоремы 202 (и, в частности, истинно).

На комбинаторный язык можно перевести не только формулировку теоремы 202, но и её доказательство. В ходе доказательства Математик не использует своё право указывать простые относительно A слова по ходу игры, а указывает их все сразу (в соответствии с графом-экспандером). Далее он объявляет слово A простым относительно X и B, если среди слов, объявленных Противником простыми относительно B, мало слов, соседних с X. В результате ему — для каждого B — удастся обслужить большинство слов, объявленных простыми относительно этого B. Оставшееся меньшинство слов, объявленных простыми относительно B, передаётся на следующий уровень обслуживания, где делается всё то же самое, но с уменьшенным на единицу m. И так далее. В итоге количество слов, объявленных простыми относительно A, вычисляется как сумма убывающей вдвое геометрической прогрессии, и потому вдвое больше первоначально взятого, но это не страшно.

12.5. Отступление: on-line паросочетание

Немного модифицировав комбинаторное доказательство теоремы 202, можно получить доказательство более сильного (и более простого) комбинаторного утверждения.

Рассмотрим двудольный граф. Вершины левой доли образуют множество A, вершины правой доли — множество B, рёбра графа образуют множество $E \subset A \times B$. Задача о паросочетании обычно ставится так: имеется некоторое множество $A' \subset A$, надо для каждой вершины $a \in A'$ выбрать одного из её соседей в B (то есть вершину из B, соединённую с a ребром), причём выбранные соседи разных вершин должны быть различны.

Рассмотрим более сложный вариант этой задачи, который можно назвать «паросочетанием в режиме on-line»: нам по очереди указывают вершины в левой доле, и мы для них должны указать парную вершину в правой доле (и не имеем возможности впоследствии изменить этот выбор). Будем говорить, что граф $E \subset A \times B$ *допускает on-line паросочетание размера* k, если существует способ, гарантирующий спаривание любых k вершин левой доли (в том порядке, в котором их указывают). Другими словами, рассматривается игра на графе, в которой противник указывает нам по очереди k вершин из A, и после каждого его хода мы обязаны выбрать одного из соседей для только что указанной им вершины (и при этом не повторяться).

Замечание: это определение несимметрично (мы выбираем вершины справа, а противник слева).

[Замечание в сторону: из определения видно, что свойство «допускать on-line паросочетание данного размера» лежит в PSPACE. Есть ли какие-нибудь лучшие верхние оценки сложности?]

Theorem 204. [multi-muchnik-th1b] *Для некоторой константы c при любых a и m (где $a \geqslant m$) существует двудольный граф с 2^a вершин в левой доле и $a^c 2^m$ вершин в правой доле, в котором каждая вершина слева имеет не более a^c соседей справа и который допускает on-line паросочетание размера 2^m.*

Другими словами, существует граф с заданными размерами долей с небольшой (полиномиальной) степенью вершин левой доли, допускающий on-line паросочетание почти что максимально возможного размера (т. е. числа вершин справа) — максимального с точностью до полиномиального множителя.

Прежде чем доказывать теорему 204, объясним, как из неё вытекает утверждение теоремы 202. Как и раньше, начнём с того, что заменим слово A на его кратчайшее описание длины a. Положим m равным $K(A|B)+1$ и применим теорему 204; получится некоторый двудольный граф (2^a вершин слева и $a^c 2^m$ вершин справа, степень левых вершин не более a^c). Раз граф с указанными свойствами существует, то его можно найти перебором. Фиксируем этот граф и алгоритм построения on-line паросочетания. Далее для данного B можно перечислять слова, имеющие сложность меньше m относительно этого B. Их не больше 2^m и среди них есть слово A. По очереди подбираем для них пару в графе. Рассмотрим слово X, парное к слову A. Как и все правые соседи A (которых мало), оно имеет малую сложность относительно A. (Поскольку граф и алгоритм построения on-line паросочетания были найдены перебором, то они задаются параметрами a и m и имеют логарифмическую сложность.) С другой стороны, зная B и X (а также все числовые параметры), можно восстановить процесс построения паросочетания и дождаться, пока слово X будет соединено со некоторым словом, тем самым восстановив A.

$\triangleleft$ Перейдём теперь к доказательству теоремы 204. Заметим, что достаточно доказать более слабое утверждение, разрешив подыскивать пары не для всех 2^m элементов, указываемых в левой доле, а только для половины (по нашему выбору). В самом деле, если мы умеем это делать, то для необслуженных элементов можно параллельно запустить аналогичный процесс (с уменьшенным на 1 значением m и соответствующим графом, возможно, другим), число необслуженных уменьшится ещё вдвое и так далее. В итоге мы найдём паросочетание в графе, в котором левая доля общая, а правая является объединением правых долей всех использованных графов (для $m, m-1, m-2\ldots$ вплоть до нуля).

Остаётся заметить, что такому (ослабленному) требованию удовлетворяет граф со свойствами экспандера, о котором мы говорили. Алгоритм построения паросочетания будет самым простым: если у вершины есть ещё не использованный сосед, то его и выбираем, а если все соседи уже заняты, то сдаёмся и вершину не обслуживаем. Несложно сообразить, что в результате будет обслужена как минимум половина вершин. В самом деле, если обслужено меньше, то и справа использовано ровно столько же, то есть менее 2^{m-1} вершин. С другой стороны, для каждой необслуженной вершины, каковых существует более 2^{m-1}, все её соседи использованы. (Иначе почему мы её не обслужили?) А это противоречит свойству экспандера: согласно этому свойству множество из 2^{m-1} элементов (и потому любое большее) не может иметь меньше 2^{m-1} соседей. ▷

[Может быть, отсюда можно вывести и теорему Вольфа – Слепяна? Вроде бы нет, там важно, что хеш-функция только одна. Не написать ли где-нибудь доказательство (простейшего) варианта теоремы Вольфа – Слепяна?]

[Мы доказали существование графа с большим on-line паросочетанием, взяв вероятностное доказательство существования экспандеров и потом модифицируя граф-экспандер (дублируя его вершины и соединяя с графами меньшего размера). Было бы естественно ожидать прямого вероятностного доказательства существования графа с большим on-line паросочетанием.]

12.6. Относительное кодирование пары слов

[multi-bglvz]

В следующей задаче для данной пары слов A и B мы хотим передать по пунктирному каналу слово X длины не более k, которое позволяет получить A из B и одновременно позволяет получить B из A (рис. 35).

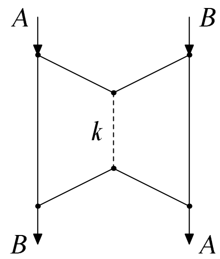

Рис. 35. Теорема Беннета – Гача – Ли – Витаньи – Цурека.
[multi-pic4]

Сразу же ясно, что это возможно лишь при $KS(A|B) \leqslant k$ и $KS(B|A) \leqslant k$. В самом деле, сложности слов, передаваемых по нижним наклонным линиям, не превосходят k (поскольку по определению эти слова просты относительно X), а вместе с тем эти слова позволяют получить B из A (формально говоря, из некоторого слова, простого относительно A) и наоборот.

Получаем необходимое условие разрешимости задачи:

$$\max(KS(A|B), KS(B|A)) \leqslant k$$

(как всегда, все неравенства понимаются с точностью до логарифмических слагаемых). Как показали Беннет, Гач, Ли, Витаньи и Цурек в [2], это необходимое условие является на самом деле и достаточным. Вот точная формулировка:

Theorem 205. `[multi-bglvz-th]` *Пусть A, B — произвольные слова, для которых $\mathit{KS}(A|B) < k$ и $\mathit{KS}(B|A) < k$. Тогда существует слово X длины k, для которого $\mathit{KS}(A|B,X) = O(\log k)$, $\mathit{KS}(B|A,X) = O(\log k)$ и $\mathit{KS}(X|A,B) = O(\log k)$.*

$\lhd$ Рассмотрим все пары слов $\langle A, B\rangle$, для которых $\mathit{KS}(A|B) < k$ и одновременно $\mathit{KS}(B|A) < k$. Получаем перечислимое бинарное отношение (на парах слов), все сечения которого (вертикальные, при фиксированном A, и горизонтальные, при фиксированном B) содержат не более 2^k элементов.

Рассматривая это отношение как (бесконечный) двудольный граф, можно сказать, что степень вершин этого графа (в каждой из долей) не превосходит 2^k.

Сейчас мы покажем, как разбить построенное множество пар на не более чем 2^{k+1} классов, каждый из которых является взаимно однозначным соответствием (не содержит двух пар на одной вертикали или горизонтали). В терминах графов: как раскрасить рёбра графа в 2^{k+1} цветов так, чтобы любые два ребра, имеющие общую вершину (слева или справа), имели разные цвета.

А именно, пронумеруем классы от 0 до $2^{k+1} - 1$. По мере появления новых пар в перечислении будем относить их к минимальному допустимому классу (в котором ещё нет пар с тем же первым или вторым элементом). (Другими словами, мы выбираем для вновь появившегося ребра первый цвет, не использованный для рёбер с общим концом.)

Ясно, что классов хватит, поскольку использованных классов меньше $2^k + 2^k$ (пар с тем же первым членом меньше 2^k, равно как и пар с тем же вторым членом). В терминах графа: новое ребро имеет общий левый конец менее чем с 2^k рёбрами и общий правый конец менее чем с 2^k, так что запрещены менее 2^{k+1} цветов.

Теперь в качестве X возьмём номер класса. (Он содержит $k+1$ битов, но отбрасывание одного бита меняет все условные сложности не более на $O(1)$.) Ясно, что зная A, число k и номер класса, можно порождать множество пар, классифицировать их описанным способом и подождать появления на A-вертикали пары из класса с данным номером X, поэтому $\mathit{KS}(A|B,X,k) = O(1)$ и $\mathit{KS}(A|B,X) = O(\log k)$. Аналогично и для $\mathit{KS}(B|A)$. Наконец, $\mathit{KS}(X|A,B,k) = O(1)$, поскольку зная A, B и k, можно дождаться классификации пары $\langle A, B\rangle$ и найти X. $\rhd$

$\boxed{\textbf{240}}$ Докажите более сильное утверждение о двудольных графах: если для (конечного) двудольного графа степень каждой вершины в левой и правой доле не превосходит N, то можно так раскрасить его рёбра в N цветов, чтобы рёбра одного цвета не имели общих концов. Объясните, почему в доказательстве теоремы 205 не удаётся сослаться на этот факт, а нужно доказывать его (ослабленный) вариант заново.

[Указание. Можно считать, что степень равна N, после чего применить теорему Форда – Фалкерсона или теорему Холла о паросочетаниях. Нам этого недостаточно, так как в нашем случае граф не задан целиком, а строится постепенно, и рёбра надо красить on-line.]

В терминах программ доказанную теорему можно сформулировать так: для любых слов A и B существует программа сложности $\max(\mathit{KS}(A|B), \mathit{KS}(B|A))$ (с точностью до логарифмического слагаемого), которая переводит A в B и B в A. В самом деле, эта программа состоит из слова X, программы получения A из пары $\langle B, X\rangle$ и программы получения B из пары $\langle A, X\rangle$, а также инструкций по различению слов A и B (для чего достаточно указать номер бита, который в этих словах отличается).

$\boxed{241}$ Пусть A и B — два независимых случайных слова длины n (то есть $KS(A) \approx n$, $KS(B) \approx n$ и $KS(\langle A, B\rangle) \approx 2n$. Укажите явно слово X, удовлетворяющее условиям теоремы. [Ответ: годится побитовая сумма слов A и B.]

Естественный вопрос, возникающий в связи с теоремой 205, таков: что можно сказать, если сложности $KS(A|B)$ и $KS(B|A)$ различны? Пусть, скажем, $KS(A|B)$ — большая из них (легче получить B из A, чем наоборот). Оказывается, тогда разделить слово X на две части: на информацию, необходимую для преобразования A в B (в лёгкую сторону, длиной $KS(B|A)$) и остаток (длины $KS(A|B) - KS(B|A)$), добавление которого позволяет сделать и обратное преобразование.

Формально говоря, верно такое утверждение:

Theorem 206. `[multi-bglvz-add]` *Пусть $KS(A|B) < k$, $KS(B|A) < l$ и $k > l$. Тогда можно найти такое слово X длины k, что $KS(X|A,B) = O(\log k)$, $KS(A|B,X) = O(\log k)$, а также $KS(B|A,X') = O(\log k)$, где X' — начало X длины l.*

$\lhd$ Рассуждаем аналогично доказательству теоремы 205, относя пары к 2^{l+1} классам и требуя, чтобы на одной вертикали было не более одной точки каждого класса, а на одной горизонтали — не более 2^{k-l} точек каждого класса. Тогда номер класса позволяет восстановить B по A, а для восстановления A по B нужно ещё $k - l$ битов информации (порядковый номер появления среди элементов данного класса). $\rhd$

$\boxed{242}$ Докажите утверждение теоремы 206 в форме, приведённой в [2]: в предположениях теоремы 206 найдутся слово Y сложности $k - l$ и слово X сложности l, для которых $KS(B,Y|A,X) = O(\log k)$ и $KS(A|B,Y,X) = O(\log k)$.

12.7. Кодирование при двух условиях

Рассмотрим теперь задачу передачи информации, которая в некотором смысле обобщает две предыдущие (рис. 36).

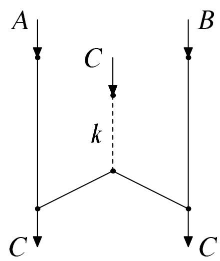

Рис. 36. Кодирование C при условиях A, B.

`[multi-pic5]`

Если в этой схеме положить $A = B$, то получится схема передачи информации раздела 12.3 (симметрично продублированная). Если же положить $C = \langle A, B\rangle$, то мы приходим к задаче раздела 12.6 (в каждом из нижних узлов одно из слов A и B известно, поэтому восстановить пару означает восстановить второе слово).

Необходимыми условиями разрешимости задачи являются неравенства

$$KS(C|A) \leqslant k, \quad KS(C|B) \leqslant k.$$

Оказывается, эти условия являются и достаточными. Вот точная формулировка:

Theorem 207. [multi-muchnik-th2] *Пусть A, B и C — произвольные слова сложности не более n, а k — натуральное число, причём $KS(C|A) \leqslant k$ и $KS(C|B) \leqslant k$. Тогда найдётся слово X длины не более $k + O(\log n)$, при котором $KS(X|C) = O(\log n)$ $KS(C|A, X) \leqslant O(\log n)$ и $KS(C|B, X) \leqslant O(\log n)$.*

В терминах программ эта теорема звучит так: для любых слов A, B, C длины не более n найдётся программа сложности $\max(KS(C|A), KS(C|B)) + O(\log n)$, имеющая логарифмическую сложность относительно C, переводящая любое из слов A и B в слово C. (Как и раньше, в эту программу надо включить информацию, позволяющую различить A и B; она имеет логарифмическую сложность.)

Заметим, что это утверждение остаётся содержательным, даже если не требовать простоты программы относительно C: никакого другого доказательства его в этом частном случае не известно. Другими словами, доказательство достаточности (пока?) не удаётся упростить, даже если добавить два новых ребра (рис. 37).

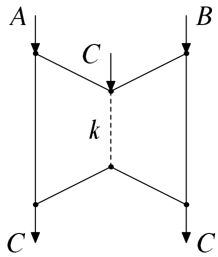

Рис. 37. Дополнительные рёбра.

[multi-pic6]

Можно также сформулировать следующее обобщение теоремы 207 на случай различных условных сложностей:

Theorem 208. [multi-muchnik-th2-add] *Пусть A, B и C — произвольные слова сложности не более n, а $k \geqslant l$ — натуральные числа, причём $KS(C|A) \leqslant k$ и $KS(C|B) \leqslant l$. Тогда найдётся слово X длины k, для которого $KS(X|C) = O(\log n)$ $KS(C|A, X) \leqslant O(\log n)$ и $KS(C|B, X') \leqslant O(\log n)$ для слова X' длины l, являющегося началом слова X.*

$\boxed{243}$ Сформулировать утверждение этой теоремы в терминах передачи информации по некоторой сети. [Указание: X передаётся по ребру пропускной способности k, из конца которого выходит ребро пропускной способности l.]

Все эти утверждения доказаны в той же работе Мучника [58]; приведём доказательство теоремы 208.

$\lhd$ Будем использовать тот же метод «отпечатков», что и раньше: слово X будет одним из небольшого числа отпечатков слова C.

Однако рассуждение требуем некоторых изменений. Если k и l отличаются, то отпечатки должны быть разных длин. Даже и в случае $k = l$ возникает очевидная проблема: среди отпечатков может найтись X, для которого $\mathit{KS}(C|A, X)$ мало, а также X', для которого $\mathit{KS}(C|B, X')$ мало, но нам ведь нужно одно и то же слово и для A, и для B.

Можно пытаться найти среди отпечатков «вдвойне хорошее» слово X, которое порождает мало коллизий и в S_A, и в S_B (через S_A и S_B мы обозначаем множества слов, простых относительно A и B соответственно). Такие слова существуют и их даже большинство (по тем же причинам, что и раньше: в $S_A \cup S_B$ лишь вдвое больше слов, чем в каждом из множеств S_A и S_B в отдельности), так что продолжая рассуждение с экспандером, можно заключить, что для большинства слов в S_A и для большинства слов в S_B найдётся вдвойне хороший правый сосед. Но дальше рассуждение не проходит: мы хотели бы сказать, что остальные слова имеют малую сложность, поскольку их мало и их можно порождать, но для их порождения надо знать сразу и A, и B, а у нас в условии лишь одно из слов A и B.

Что же делать? Будем рассматривать условия A и B независимо, но требовать, чтобы у слова C был не просто один хороший отпечаток, а чтобы больше половины отпечатков были хорошими. Если этого удаётся добиться для A и для B в отдельности, отсюда следует существование отпечатка, одновременно хорошего и для A, и для B.

Соответственно придётся изменить и свойство типа экспандера, которого мы требуем от двудольного графа $E \subset P \times Q$ (с левой долей P и правой долей Q). Теперь мы хотим, чтобы для любого множества $U \subset Q$ не слишком большого размера множество тех вершин $x \in P$, у которых половина или более соседей справа попадает в U, было бы малым. (Раньше мы считали плохими вершины, у которых *все* соседи попадали в U, а теперь достаточно половины.)

Более того, поскольку теперь нас интересуют разные условные сложности, мы будем рассматривать не только сами отпечатки, но и их начала (сразу всех длин, так проще). Утверждение о существовании графа с нужными свойствами теперь выглядит так (через $[u]_m$ обозначается начало слова u, имеющее длину m):

Лемма. Пусть даны натуральные числа n и N, а также положительное число ε, причём

$$n2^{N+2n+1}\varepsilon^{N/2} < 1.$$

Тогда существует семейство отображений

$$\chi_1, \ldots, \chi_N \colon \mathbb{B}^n \to \mathbb{B}^n$$

с таким свойством: для любого $m \in \{1, \ldots, n\}$ и для любого непустого подмножества $U \subset \mathbb{B}^m$, число элементов в котором не превосходит $\varepsilon 2^m$, количество тех $x \in \mathbb{B}^n$, для которых

$$[\chi_i(x)]_m \in U \quad \text{для половины или более значений } i \in \{1, \ldots, N\}$$

меньше $|U|$ (числа элементов в U).

[Пояснение: мы говорим не о графе со степенью вершин N в левой доле, а о семействе N отображений, поскольку допускаем кратные рёбра ($\chi_i(x) = \chi_j(x)$ при $i \neq j$), учитываемые с соответствующей кратностью.]

Доказательство леммы. Покажем, что для случайно выбранных функций $\chi_1, \ldots, \chi_N$ (все значения $\chi_i(x)$ при всех i и x независимы и равномерно распределены в $\mathbb{B}^n$) вероятность нарушения указанного свойства меньше единицы. Эту вероятность мы оценим сверху. Для каждого $m \leqslant n$, для каждого $t \leqslant \varepsilon 2^m$ и для любых множеств $T \subset \mathbb{B}^n$ и $U \subset \mathbb{B}^m$, содержащих по t элементов, оценим вероятность того, что для каждого элемента $x \in T$ не менее половины значений $[\chi_i(x)]_m$ (при $i = 1, \ldots, n$) попадает в U. Для фиксированного $x \in T$ вероятность того, что не менее половины его соседей попадает в U, не превосходит $2^N \varepsilon^{N/2}$, поскольку для каждого из не более чем 2^N подмножеств множества $\{1, \ldots, N\}$, содержащих более $N/2$ элементов, вероятность того, что все входящие в него значения i ведут внутрь U, не превосходит $\varepsilon^{N/2}$ (значения $[\chi_i(x)]_m$ при разных i независимы и равномерно распределены в $\mathbb{B}^m$, а доля U среди элементов $\mathbb{B}^m$ не превосходит ε). Такое событие должно произойти независимо для всех $x \in T$, так что полученную оценку надо возвести в степень t.

Таким образом, для интересующей нас вероятности (которая должна быть меньше единицы) мы получаем оценку

$$\sum_{m=1}^{n} \sum_{t=1}^{\varepsilon 2^m} \sum_{T \subset \mathbb{B}^n, |T|=t} \sum_{U \subset \mathbb{B}^m, |U|=t} (2^N \varepsilon^{N/2})^t$$

Число различных множеств T не превосходит 2^{nt} (столько есть последовательностей длины t, составленных из элементов $\mathbb{B}^n$); число различных множеств U не превосходит 2^{mt}. Учитывая это, получаем верхнюю оценку

$$\sum_{m=1}^{n} \sum_{t=1}^{\varepsilon 2^m} 2^{tn}\, 2^{tm}\, 2^t N\, \varepsilon^{Nt/2}$$

или

$$\sum_{m=1}^{n} \sum_{t=1}^{\varepsilon 2^m} \left(2^n\, 2^m\, 2^N\, \varepsilon^{N/2}\right)^t$$

Внутренняя сумма представляет собой геометрическую прогрессию. В условиях леммы знаменатель этой прогрессии меньше $1/2$, и потому сумма прогрессии не превосходит удвоенного первого члена, который не зависит от t. Учитывая это, получаем верхнюю оценку

$$2n \cdot (2^{N+n+m} \varepsilon^{N/2}) \leqslant n 2^{N+2n+1} \varepsilon^{N/2},$$

что меньше единицы по условию леммы.

Лемма доказана.

Мы будем использовать эту лемму при $\varepsilon = 1/n$. В этом случае можно переписать условие леммы как

$$n 2^{N+2n+1} < n^{N/2}$$

или

$$\log n + N + 2n + 1 < (N/2) \log n.$$

Видно, что можно положить $N = n$ или даже $N = \lceil cn/\log n \rceil$ при достаточно большом c, и условие леммы будет выполнено при всех достаточно бо́льших n.

Продолжим доказательство теоремы 208. Как и раньше, можно заменить слово C его кратчайшим (безусловным) описанием и считать его словом длины n, то есть элементом $\mathbb{B}^n$. (Сложность слов A и B роли не играет; в частности, она может быть и больше n.) Согласно лемме при $N = n$ и $\varepsilon = 1/n$ можно найти N отображений $\chi_1, \ldots, \chi_N \colon \mathbb{B}^n \to \mathbb{B}^n$ с указанными в лемме свойствами. При этом, как уже говорилось, взяв первый в каком-либо естественном порядке набор функций с такими свойствами, мы можем считать, что сложность этого набора есть $O(\log n)$, поскольку для его задания достаточно указать число n.

Пусть $\mathit{KS}(C|A) = k$ и $\mathit{KS}(C|B) = l$. (В условии теоремы были неравенства $\mathit{KS}(C|A) < k$ и $\mathit{KS}(C|A) < l$, но мы имеем право доказывать теорему для уменьшенных значений k и l, от этого утверждение становится только сильнее.) Оставляя от хеш-значений только k или l первых битов, мы получим N отображений $\mathbb{B}^n$ в $\mathbb{B}^k$ (соответственно в $\mathbb{B}^l$). Эти семейства задают двудольные графы на $\mathbb{B}^n \times \mathbb{B}^a$ и $\mathbb{B}^n \times \mathbb{B}^b$, в которых степень каждой вершины слева равна N (считая кратные рёбра). Нас интересуют ограничения этих графов на S_A и S_B, где S_A состоит из слов длины n, имеющих сложность не более k относительно A, а S_B состоит из слов длины n, имеющих сложность не более l относительно B. Мы выделяем в $\mathbb{B}^k$ плохие вершины, имеющие более n^c соседей в S_A; аналогичным образом плохими вершинами в $\mathbb{B}^l$ мы считаем те, у которых имеется более n^c соседей в S_B. (Точное значение достаточно большой константы c мы выберем позже.)

Число плохих вершин в обоих случаях не превосходит соответственно

$$2N \cdot 2^k/n^c \text{ и } 2N \cdot 2^l/n^c,$$

так как степень плохой вершины больше n^c, общее число рёбер в графе не больше $|S_A| \cdot N$ (соответственно $|S_B| \cdot N$), а $|S_A| < 2 \cdot 2^k$ и $|S_B| < 2 \cdot 2^l$.

Назовём плохими в S_A те вершины, у которых половина или более соседей в графе на $S_A \times \mathbb{B}^k$ (с учётом кратности) являются плохими. Лемма гарантирует, что число плохих вершин в S_A меньше

$$2N \cdot 2^k/n^c$$

(мы считаем, что c достаточно велико, поэтому оценка на число плохих вершин меньше $\varepsilon 2^k$, где $\varepsilon = 1/n$). Поскольку плохие вершины можно перечислять, зная n, k и A, сложность любой из них относительно A не превосходит

$$\log(2N \cdot 2^k/n^c) + O(\log n) \leqslant k - c\log n + O(\log n)$$

(напомним, что $N = n$). При достаточно большом c все плохие вершины имеют сложность (относительно A) меньше k, и слово C не попадает в их число. Это значит, что больше половины значений $[\chi_i(C)]_k$ (среди N вариантов $i = 1, 2, \ldots, N$) являются хорошими в $\mathbb{B}^k$.

Повторяя те же рассуждения для графа в $S_B \times \mathbb{B}^l$, мы получаем, что больше половины значений $[\chi_i(C)]_l$ хороши в S_B. Следовательно, найдётся значение i, которое даёт хороших соседей в обоих случаях. Тогда $X = [\chi_i(C)]_k$ и $X' = [\chi_i(C)]_l$ являются искомыми. $\triangleright$

244 Сформулируйте и докажите аналогичное утверждение для трёх условий (или для полиномиального их числа.

Может возникнуть желание ещё усилить результаты теорем 202 и 208. Нельзя ли ограничиться одним-единственным отпечатком? Скажем, нельзя ли для слова A длины n найти такое слово X длины $n/2$, что $\mathit{KS}(A|X, B) \approx 0$ для любого слова B, для которого $\mathit{KS}(A|B) \leqslant n/2$?

Легко понять, что этого всё-таки добиться нельзя. В самом деле, пусть такое слово X есть. Оно должно иметь сложность $n/2$ (иначе в нём было бы слишком мало информации). Если взять его в качестве B, то в условиях X и B информация будет дублироваться, и сложность $KS(A|X,B)$ будет примерно равна $n/2$ (а не нулю, как нам хотелось).

$\boxed{\mathbf{245}}$ Покажите, что двух (и любого фиксированного числа) отпечатков недостаточно. [Указание. В качестве B можно взять слово, которое получится соединением двух первых половин этих отпечатков. Если в итоге будет менее $n/2$ информации относительно A, то надо добавить подходящий кусок слова A. В результате оба отпечатка будут содержать дублирующую информацию и не подойдут. В [58] даны более точные количественные оценки.]

12.8. Поток информации через разрез

Мы рассмотрели несколько схем передачи информации; для каждой из них были указаны необходимые и достаточные условия разрешимости соответствующей задачи. Во всех приведённых примерах эти условия получаются по некоторой единой схеме, которую мы сейчас укажем явно.

Пусть задан граф передачи информации (ориентированный ациклический граф, на некоторых рёбрах которого указаны максимальные пропускные способности; для некоторых вершин заданы также входные и выходные слова). Мы хотим указать необходимые условия, то есть условия на эти слова, которые заведомо будут выполнены, если соответствующая задача разрешима.

Выделим произвольное множество I вершин графа (разрез) и будем изучать поток информации, проходящей через этот разрез (идущей извне множества I внутрь него). Рассмотрим пропускные способности всех рёбер, начало которых лежит вне I, а конец внутри I. Если среди них есть хоть одно ребро с неограниченной пропускной способностью, то никакого необходимого условия для такого I не получится. Пусть все пропускные способности этих рёбер ограничены и равны $u_1,\ldots,u_k$. Пусть $V_1,\ldots,V_l$ — все входные слова для вершин из I, а $W_1,\ldots,W_m$ — все выходные слова для вершин из I. Тогда можно записать такое необходимое условие:

$$KS(W_1,\ldots,W_m \mid V_1,\ldots,V_l) \leqslant u_1+\ldots+u_k.$$

(Как всегда, неравенства рассматриваются с точностью до логарифма суммарных длин входящих в них слов.) В самом деле, если мы знаем все слова $V_1,\ldots V_l$, а также все слова, написанные на ведущих внутрь I рёбрах, то можем восстановить (с логарифмической дополнительной информацией) все слова, выходящие из вершин множества I, включая $W_1,\ldots,W_m$. Это нужно делать, рассматривая вершины множества I по очереди (начало любого ребра должно предшествовать его концу; такой порядок существует, так как в графе нет циклов).

По существу это рассуждение оценивает «поток информации через разрез».

Покажем на примере, как получаются необходимые условия в рассмотренных нами случаях. Рассмотрим схему передачи информации раздела 12.6 и в качестве множества I возьмём множество из трёх вершин, показанное на рис. 38 внутри пунктирной линии.

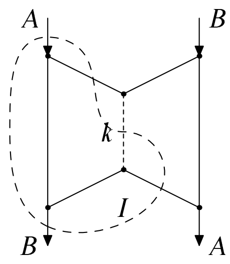

Рис. 38. Необходимое условие для теоремы Беннета – Гача – Ли – Витаньи – Цурека [multi-pic7]

В это множество входят слово A и слово длины k (по ребру ограниченной пропускной способности); два других ребра графа, пересекающие границу I, ведут изнутри наружу (напомним, что по нашему соглашению все рёбра идут сверху вниз). Получается условие $KS(B|A) \leqslant k$, которое мы и рассматривали.

246 Покажите, что все остальные необходимые условия, указанные нами для рассмотренных выше задач, также могут быть получены при подходящем выборе множества I на соответствующем графе.

12.9. Сети с одним источником

Возникает естественный вопрос: являются ли необходимые условия, указанные в предыдущем разделе (и взятые для всех множеств I, для которых они имеют смысл), одновременно и достаточными? В наших предыдущих примерах это оказывалось именно так. В общем случае, как мы увидим дальше, это неверно. Однако в случае, когда мы хотим передать какое-то одно слово A из одного источника в несколько мест назначения (другими словами, когда входное слово только одно, а все выходные слова равны входному), наши условия будут и достаточными.

Для классической теории информации аналогичная задача была рассмотрена в работах [1,39]; наше рассуждение следует использованной там схеме (с некоторыми изменениями, связанными с переходом от шенноновской энтропии к колмогоровской сложности).

Начнём с примера: пусть мы хотим передать слово A длины $2k$ в три пункта назначения, как показано на рис. 39; все каналы связи имеют неограниченную пропускную способность, кроме трёх первых, которые имеют пропускную способность k битов. Можно ли это сделать?

Легко понять, что в каждый из пунктов назначения *по отдельности* можно передать слово A. Например, чтобы передать его в левую из трёх вершин, надо разбить его на две половины, каждая по k битов, и передать эти половины по двум левым каналам связи пропускной способности k (третий канал пропускной способности k бесполезен, так как идущая по нему информация не может попасть в нужную нам вершину).

Аналогичным способом легко передать A в любую из трёх вершин, использовав два канала из трёх. Но передать слово A *одновременно* в три вершины уже не так просто: для этого его надо разрезать на «три половины», причём так, чтобы из любых двух половин его можно было составить обратно.

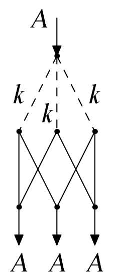

Рис. 39. Деление информации на части.

[multi-pic8]

Стандартный способ («разделение секрета с помощью линейных отображений») состоит в том, чтобы передать по трём каналам слова A_1, A_2 и $A_1 \oplus A_2$, где A_1 и A_2 — две половины слова A (каждая содержит k битов), а $A_1 \oplus A_2$ — побитовая сумма слов A_1 и A_2 по модулю 2. По любым двум из этих трёх k-битовых слов можно восстановить третье (оно равно их побитовой сумме), а потому можно восстановить и слово A.

Оказывается, аналогичный метод можно применить и в общем случае, и справедливо такое утверждение.

Theorem 209. *Рассмотрим схему передачи информации с единственным входным словом A длины n и выходными словами A, в которой заданы целые ограничения на пропускные способности некоторых рёбер.*

Пусть выполнены все необходимые условия описанного выше типа, то есть для любого множества вершин J, не содержащего вершину, в которую входит A и содержащего хотя бы одну вершину, из которой выходит A, сумма пропускных способностей рёбер с началами вне J и концами в J не меньше n (или есть хотя бы одно ведущее внутрь ребро с неограниченной пропускной способностью).

Тогда задача передачи информации по этой схеме разрешима с точностью до $O(\log n)$: можно написать на рёбрах такие слова, чтобы соответствующие условные сложности в каждой вершине не превосходили $O(\log n)$.

(Константа, подразумеваемая в $O(\log n)$, зависит от графа, но не зависит от числа n, пропускных способностей и слова A.)

$\triangleleft$ Рассмотрим сначала случай одного выходного слова. В этом случае требуется передать некоторый объём информации (а именно, n битов) из одной вершины (назовём её s) в какую-то другую вершину t (только одну!) по рёбрам.

Запакуем каждый бит в отдельный конверт; получится n конвертов, которые изначально находятся в вершине s. Организуем доставку этих конвертов в вершину t по рёбрам графа. При этом мы следим за пропускной способностью и требуем, чтобы по ребру с ограничением k было перевезено не более k конвертов.

Эта задача разрешима по теореме Форда – Фалкерсона о максимальном потоке и минимальном разрезе (см., например, [64]; поскольку все ограничения целочисленные, то и поток будет целочисленным).

Теперь на каждом ребре можно написать те биты, которые содержатся в перевозимых по нему конвертах. Точнее, после применения алгоритма для задачи Форда – Фалкерсона мы для каждого ребра имеем некоторый перечень битов (список их номеров в возрастающем порядке; общее число номеров не превосходит пропускной способности ребра), и записываем на этом ребре подпоследовательность, состоящую из битов с этими номерами.

Покажем, что выполнено ограничение на условную сложность. Рассмотрим вершину и слова на выходящих и входящих рёбрах этой вершины. Слова на выходящих рёбрах составлены путём перемешивания битов в словах на входящих рёбрах; схема этого перемешивания не зависит от слова A и алгоритмически вычисляется, если известно число n и пропускные способности рёбер. Без ограничения общности можно предполагать, что пропускные способности рёбер не больше n, поэтому схема перемешивания имеет сложность $O(\log n)$ (константа зависит от графа). А зная эту схему, можно получить выходящие слова по известным входящим.

Случай передачи информации в одну вершину разобран.

Идея доказательства теоремы в общем случае состоит в том, чтобы производить линейное кодирование. До сих пор мы лишь перекоммутировали биты (перекладывали конверты из входящих линий в выходящие). Более общий способ: в каждом узле применяется линейное отображение. Каждый выходящий бит будет линейной функцией от входящих битов. Для начала будем рассматривать биты как элементы поля $\mathbb{F}_2$ (поле из двух элементов: нуля и единицы, при этом $1 + 1 = 0$). Тогда l-битовые слова будут элементами l-мерного векторного пространства над этим полем.

Пусть в некоторую вершину входят рёбра с пропускной способностью $i_1, \ldots, i_p$, а выходящие рёбра имеют пропускные способности $j_1, \ldots, j_q$. (Мы будем считать, что все пропускные способности рёбер конечны, заменив бесконечность на n, число битов в передаваемом слове.) Тогда преобразование информации в этой вершине задаётся матрицей размера $(j_1 + \ldots + j_q) \times (i_1 + \ldots + i_p)$; умножая её на столбец входных битов (со всех входящих рёбер), получаем столбец выходных битов (для всех выходящих рёбер). Заметим, что по ребру пропускной способности k передаётся ровно k битов (независимо от того, насколько это ребро кажется полезным).

В приведённом выше примере (когда слово делилось на две части, после чего они складывались побитово) как раз и выполнялись такие линейные преобразования.

Пусть для каждой вершины такие линейные преобразования (матрицы) заданы. Тогда для каждого выхода возникает отображение входа в этот выход — некоторое линейное отображение n-мерных пространств над полем $\mathbb{F}_2$. Нам хотелось бы, чтобы все эти отображения были обратимы; в этом случае каждое выходное слово содержало бы полную (с точностью до линейного взаимно однозначного соответствия) информацию о входных битах, и задача передачи информации была бы разрешима для произвольного входного слова A (на каждом ребре графа нужно было бы написать передаваемый по нему набор битов для слова A).

В самом деле, сами линейные отображения можно выбрать небольшой сложности: если мы знаем, что существуют такие матрицы преобразования в вершинах, при которых все отображения входа в выходы одновременно обратимы, то первый (в каком-либо естественном порядке) набор таких матриц имеет логарифмическую сложность, поэтому для каждой вершины условная сложность выходящих слов при известных входящих логарифмическая.

Проблема в том, что не всегда можно добиться одновременной обратимости всех отображений вход $\to$ выход.

$\boxed{247}$ Рассмотрим граф рис. 40; входное слово имеет длину 2, пропускная способность всех рёбер равна 1. Покажите, что нельзя указать такие линейные преобразования в вершинах, чтобы все шесть отображений входа в выход были бы обратимыми. [Указание. Существует всего три ненулевых линейных функционала на $\mathbb{F}_2^2$, поэтому в двух промежуточных вершинах будет одинаковая информация.]

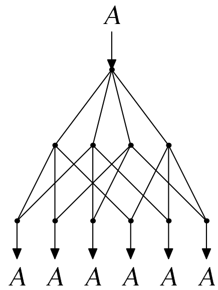

Рис. 40. Двух элементов в поле мало.

[multi-pic9]

Подчеркнём, что для каждого выхода по отдельности можно подобрать преобразования в вершинах, при которых соответствующее отображение обратимо, как видно из рассуждения для случая одного выхода. (При этом преобразования в вершинах будут перестановками битов, этого достаточно.) Проблема именно в том, чтобы сделать это одновременно для всех входов.

Временно изменим постановку задачи и вместо битов будем рассматривать элементы произвольного поля F. Входом тогда будет вектор из F^n, через ребро пропускной способности k будет проходить вектор из F^k, а преобразования в вершинах будут линейными над F, то есть будут задаваться матрицами с элементами из F.

Мы докажем, что если поле F достаточно велико, то можно выбрать преобразования в вершинах так, чтобы отображения входа во все выходы были бы одновременно обратимы.

Будем считать элементы матриц преобразования в вершинах переменными (принимающими значения в F). Тогда элементами матрицы преобразования входа в выход будут многочлены от этих переменных, и потому её определитель тоже будет многочленом. Степень этого многочлена, как легко проверить, не превосходит nE, где E — число рёбер графа (при движении от входа к выходу в каждом узле степень увеличивается на единицу, а степень определителя матрицы $n \times n$ в n раз больше степени матричных элементов).

Таким образом, для каждого выхода у нас есть многочлен не очень большой степени (определитель соответствующего преобразования), причём известно, что этот многочлен не обращается тождественно в нуль (ведь на этот выход в отдельности мы передавать информацию умеем). Теперь вспомним простой алгебраический факт:

Лемма Многочлен степени d от m переменных над полем F либо равен нулю тождественно, либо принимает нулевое значение с вероятностью не более $dm/|F|$ (все элементы F^m считаем равновероятными, $|F|$ — число элементов поля F).

Доказательство леммы проходит индукцией по m; запишем многочлен как многочлен от одной переменной, коэффициенты которого представляют собой многочлены от $m-1$

переменных. Этот многочлен одной переменной может быть нулевым или ненулевым в зависимости от значений остальных переменных; вероятность того, что он окажется нулевым, не больше $d(m-1)/|F|$ по предположению индукции (достаточно рассмотреть один его ненулевой коэффициент), а если многочлен ненулевой, то вероятность обратиться в нуль, попав в его корень, не больше $d/|F|$, так что всего получаем $(d + d(m-1))/|F|$. Лемма доказана.

Теперь заметим, что если вероятность обращения в нуль определителя для каждого выхода меньше единицы, делённой на число выходов, то найдутся значения переменных, при которых все определители ненулевые.

Подсчитаем: степень многочлена не больше nE, число переменных не больше n^2E (для каждого ребра есть матрица из переменных, которая выражает сигнал на этом ребре через сигналы, входящие в начальную вершину этого ребра), выходных рёбер тоже не больше E, так что если $n^3E^3 < |F|$, то наше рассуждение доказывает, что есть линейные преобразования в вершинах, делающие сразу все отображения входа в выходы обратимыми одновременно. (И такое преобразование можно найти перебором, так что среди них есть простые.)

[Верна ли оценка для числа переменных?! — В А в чём проблема? — Ш]

Как это можно применить в ситуации, когда имеется слово из n битов, а вовсе не элементов конечного поля? Как это принято в теории кодирования, будем разбивать слово на блоки некоторой длины k (в количестве n/k) и считать каждый блок элементом поля из 2^k элементов (такое поле, как известно из алгебры, существует).

Если оказалось, что число n, а также все пропускные способности делятся на k, и при этом $2^k > (n/k)^3E^3$, то всё хорошо.

Как быть в общем случае? Для начала нужно выбрать значение k, при котором $2^k > n^3E^3$ (мы берём k с запасом, пренебрегая делением на k в правой части). Отметим, что при этом $k = O(\log n)$. Затем надо округлить n и пропускные способности до целых кратных числа k; при этом n надо округлять с уменьшением, а пропускные способности — с увеличением, чтобы не нарушить условия на пропускные способности разрезов. При этом погрешность будет порядка $O(\log n)$, и остаётся только воспользоваться доказанным утверждением. $\triangleright$

$\boxed{\textbf{248}}$ Используя описанный метод, опишите вероятностный полиномиальный алгоритм отыскания величины максимального потока в ориентированном графе без циклов с целыми пропускными способностями.

Теперь мы переходим к примерам, где условия на поток информации через разрез оказываются необходимыми, но не достаточными.

12.10. Выделение общей информации

[multi-cominf]

Один пример такого рода (где необходимые условия на поток информации недостаточны) мы по существу уже рассматривали. Это задача о выделении общей информации, рассмотренная в главе 11. В разделе 11.2 для данных слов x, y, z и чисел α, β и γ мы интересовались, найдётся ли слово z, для которого

$$KS(z) < \alpha, \quad KS(x|z) < \beta, \quad KS(y|z) < \gamma.$$

Легко понять, что с точностью до логарифмических слагаемых эту задачу можно сформулировать как задачу передачи информации в графе рис. 41. В самом деле, если удаётся

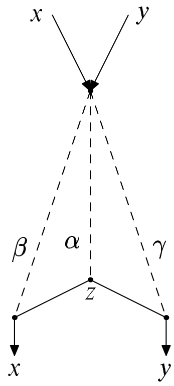

Рис. 41. Задача выделения общей информации.

[multi-pic10]

найти слово z, для которого выполнены указанные неравенства, то его можно передавать по среднему ребру, а по крайним рёбрам передать условные описания x и y при известном z. (И это слово, и условные описания можно найти перебором при известных x, y, z, так что в верхней вершине новой информации не возникает.) Напротив, если задача о передаче информации с указанными ограничениями разрешима, то слово z, передаваемое по среднему ребру, удовлетворяет неравенствам (с логарифмической точностью).

Видно, что условия на пропускные способности разрезов соответствуют неравенствам

$$KS(x) \leqslant \alpha + \beta, \quad KS(y) \leqslant \alpha + \gamma, \quad KS(x, y) \leqslant \alpha + \beta + \gamma,$$

и вся глава 11 была посвящена примерам ситуаций, когда эти условия оказываются недостаточными для существования слова z («общей информации»).

12.11. Упрощение программы

В предыдущем разделе приведён пример задачи передачи информации на графе, в которой необходимые условия на потоки не являются достаточными для её разрешимости. Представляют интерес другие примеры такого рода, по возможности попроще. Оказывается, что утверждение теоремы 202 лежит довольно близко к границе: если рассмотреть чуть более общую постановку задачи, то необходимое условие перестанет быть достаточным.

Рассмотрим задачу рис. 42 (её предложил М. Вьюгин). Разница с теоремой Мучника в том, что внизу требуется не восстановить одно из двух входных слов (P и A), а получить некоторое третье слово B). Оказывается, что и в этой задаче необходимые условия на поток информации $KS(B|A) \leqslant k$ и $KS(B|A, P) = 0$ (как всегда, рассматриваемые с логарифмической точностью) не являются достаточными.

(Три разных доказательства из статьи?)

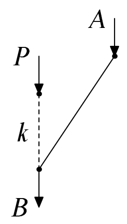

Рис. 42. Задача упрощения программы.
[multi-pic11]

12.12. Минимальная достаточная статистика

Другая задача, в которой необходимые условия также не совпадают с достаточными, показана на рис. 43. Здесь на выходе нужно получить одно из входных слов, но теперь

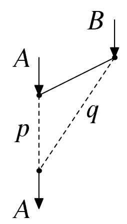

Рис. 43. Два канала ограниченной пропускной способности.
[multi-pic12]

уже пропускные способности обоих каналов ограничены. При этом мы дополнительно разрешаем использования информацию о слове B при формировании сообщения по левому каналу.

Необходимые условия, связанные с потоками информации через разрезы, таковы: $KS(A) \leqslant p+q$ и $KS(A|B) \leqslant p$. (Как всегда, мы опускаем оговорки о логарифмических поправках.)

249 Нарисуйте соответствующие разрезы.

Для некоторых пар слов эти необходимые условия являются также и достаточными.

250 [multi-max-profile] Проверьте, что если у слов A и B полностью выделяется общая информация (например, слова A и B являются перекрывающими кусками случайного слова), то эти необходимые условия являются также и достаточными. [Указание. Условие $KS(A|B) \leqslant p$ позволяет передать по левому каналу часть слова A, не попавшую в B, и ещё сколько-то, а остаток можно передать по правому каналу (условие $KS(A) \leqslant p+q$ гарантирует, что места в канале хватит).]

Однако в общем случае необходимые условия могут не быть достаточными. Для наглядности зафиксируем сложности и условные сложности слов A и B: пусть оба слова A и B имеют сложность $2n$, а пара $\langle A, B\rangle$ имеет сложность $3n$ (и потому условные сложности равны n). Необходимые условия $p+q \geqslant 2n$ и $p \geqslant n$ для этого случая показаны на рис. 44

Теперь попытаемся понять, при каких p и q задача заведомо разрешима (для любой пары слов A и B указанной сложности). Можно передавать слово A целиком по левому

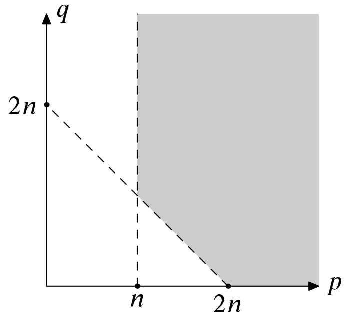

Рис. 44. Необходимые условия.

[multi-pic13]

каналу, поэтому задача разрешима при $p = 2n$, $q = 0$ (а также для всех бóльших p и q). Можно передавать слово B целиком по правому каналу, поэтому задача разрешима при $p = n$, $q = 2n$ (а также для всех бóльших p и q). Это соответствует двум квадрантам с вершинами в $(2n, 0)$ и $(n, 2n)$. Более того, если от B отрезать (скажем, последние) k битов, то условная сложность $KS(A|B)$ увеличится не более чем на k, поэтому задача разрешима при $q = 2n-k$, $p = n+k$. Таким образом, вся тёмно-серая область на рисунке 45 соответствует парам $\langle p, q\rangle$, для которых задача разрешима.

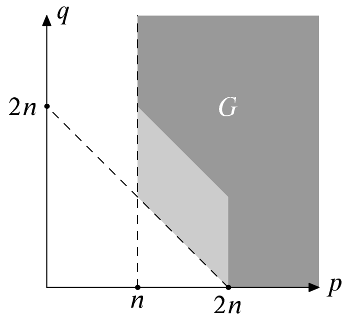

Рис. 45. Достаточные условия.

[multi-pic14]

Как и в случае с выделением общей информации (глава 11), можно сказать, что «профиль» пары слов A, B (множество пар $\langle p, q\rangle$, при которых задача разрешима) зависит от того, какую именно пару слов (с данной сложностью и условной сложностью) мы возьмём. Как мы видели в задаче 250, для некоторых пар слов этот профиль совпадает с максимально возможным. Следующая теорема утверждает, что бывает и наоборот — есть пара слов, при которых профиль совпадает с минимально возможным, то есть с тёмно-серой областью на рисунке, которую мы будем обозначать G.

Нужно только понять, как это правильно сформулировать. Хотелось бы сказать, что для всех B', простых относительно B, пара $\langle KS(A|B'), KS(B')\rangle$ находится в $O(\log n)$-окрестности множества G, причём константа в $O(\log n)$ не зависит ни от n, ни от B'. Однако выражение «простые относительно B» требует количественного уточнения: мы должны указать некоторую границу r и рассматривать слова B', для которых $KS(B'|B) < r$. Естественно ожидать, что чем больше r, тем дальше может отходить пара $\langle KS(A|B'), KS(B')\rangle$ от G. Как мы увидим, связь тут линейная.

Theorem 210. `[multi-statistic-1]` *Для каждого n можно указать такие слова A и B сложности $2n + O(\log n)$, что $KS(A, B) = 3n + O(\log n)$, и для любого B' пара $\langle KS(A|B'), KS(B')\rangle$ находится в $O(\log n) + O(KS(B'|B))$-окрестности множества G.*

$\triangleleft$ Утверждение этой теоремы, как часто бывает (см. выше в этой главе), соответствует некоторой игре. Мы опишем эту игру, укажем выигрышную стратегию и выведем отсюда утверждение теоремы.

Пусть фиксировано значение n; при каждом n будет своя игра. В этой игре мы имеем право для каждого B длины $2n$ указать до 2^n слов длины $2n$, условно именуя их «n-простыми относительно B».

Противник для каждых p, q и r (из некоторого множества допустимых троек натуральных чисел; это множество мы опишем позднее) имеет право:

• для каждого слова B длины $2n$ указать до 2^r слов длины q, которые условно именуются «r-простыми относительно B»;

•для каждого слова B' длины q указать до 2^p слов длины $2n$, которые условно именуются «p-простыми относительно B'».

Для каждой тройки $\langle p, q, r\rangle$ (из множества допустимых троек) это происходит независимо. Можно сказать, что мы играем против команды противника, в которой для каждой тройки $\langle p, q, r\rangle$ есть свой игрок, делающий свои объявления и подчиняющийся своим ограничениям, но цель игры у них общая.

Кроме этого, в команде противника есть ещё два дополнительных игрока. Один из них имеет право браковать слова длиной $2n$ (но не более 2^{2n-1} штук, то есть половины всех слов), называя их «плохими». Второй имеет право для каждого слова B длины $2n$ забраковать до 2^{n-2} слов длины $2n$ (свои для каждого B), объявляя их «плохими для данного B». (Рассматриваемый далее противник будет браковать слова длиной $2n$, имеющие сложность менее $2n - 1$, а также для каждого слова B браковать слова A, для которых $KS(A|B) < n - 2$, но в определение игры это не входит.) Мы используем границу $n - 2$ (а не, скажем, $n - 1$), поскольку нам потребуется некоторый запас, см. ниже.

Ходы делаются игроками постепенно: в любой момент каждый из игроков может объявить новое слово простым или плохим (не нарушая количественных ограничений). Поскольку общее число возможных ходов конечно, игра рано или поздно кончится, хотя внешний наблюдатель (не знающий используемых игроками стратегий) не сможет сказать, так ли это — игроки не объявляют о конце игры.

Остаётся определить, кто выигрывает в предельной позиции. Будем считать, что выиграл противник, если для каждой пары слов $\langle A, B\rangle$ длины $2n$ (каждое), , в которой слово A объявлено n-простым относительно слова B, причём A и B не забракованы (не объявлены плохими) и A не забраковано для данного B (не объявлено плохим для этого B), найдётся допустимая тройка $\langle p, q, r\rangle$ и слово B' длины q, для которой

• B' объявлено r-простым относительно B;

• A объявлено p-простым относительно B'.

Мы укажем простую стратегию, позволяющую выигрывать в этой игре (при некотором множестве допустимых троек $\langle p, q, r\rangle$), а затем выведем из этого утверждение теоремы.

Выигрывающая стратегия состоит в том, что на каждый ход противника (любого из участников противостоящей команды), состоящий в объявлении нового «простого» или «плохого» слова, мы отвечаем одним своим ходом. Этот ход состоит в добавлении (для некоторого слова B длины $2n$) одного n-простого слова A длины $2n$, которое помешало бы противнику выиграть (если он не сделает нового хода).

Для этого нужно, чтобы выполнялись следующие условия:

• выбранные слова A и B (ещё) не забракованы (не объявлены плохими);

• слово A (ещё) не объявлено плохим относительно B;

• ни для одной допустимой тройки $\langle p,q,r\rangle$ не найдётся слова B' длины q, которое было бы объявлено $\langle p,q,r\rangle$-игроком команды противника r-простым относительно B, и для которого A объявлено этим же игроком p-простым относительно B'.

Почему это возможно? На любой момент игры имеется не менее 2^{2n-1} незабракованных слов. Взяв в качестве B одно из них, мы имеем право объявить любое слово n-простым относительно B, если только уже не объявили 2^n слов простыми относительно B. Но если такое случилось для всех незабракованных B, это значит, что мы уже сделали $2^{2n-1}\cdot 2^n = 2^{3n-1}$ ходов, отвечая одним своим ходом на каждый ход противника, а противник так много ходов сделать не сможет (см. далее). Таким образом, слово B мы выбрать сможем.

Выбрав B, мы начинаем выбирать A. Фиксируем допустимую тройку $\langle p,q,r\rangle$ и подсчитаем, сколько слов не годятся из-за неё. Имеется не более 2^r слов B' длины q, объявленных r-простыми относительно B. Для каждого B' есть не более 2^p слов длины $2n$, объявленных p-простыми относительно этого B'. Таким образом, нам не годятся 2^{p+r} слов (для каждой тройки $\langle p,q,r\rangle$ из множества допустимых троек M), всего $2^{p+r}\cdot|M|$. Ещё нам не подходят слова, объявленные плохими (не более 2^{2n-1} штук), а также слова, объявленные плохими относительно выбранного B (их не более 2^{n-1}). Таким образом, искомый ход заведомо возможен, если

$$2^{p+r}\cdot|M| + 2^{2n-1} + 2^{n-1} < 2^{2n} \qquad (*)$$

Теперь надо подсчитать, сколько ходов (в течение одной партии) может сделать противник. Каждый из $|M|$ игроков, отвечающих за допустимые тройки $\langle p,q,r\rangle$, делает не более 2^{2n+r} ходов, объявляя слова r-простыми, а также не более 2^{q+p} ходов, объявляя слова p-простыми. Поэтому общее число ходов этих игроков не больше

$$|M|\cdot\left(2^{\max(2n+r)} + 2^{\max(q+p)}\right)$$

(где максимумы в показателе берутся по всем тройкам из M). Сюда ещё надо добавить 2^{2n-1} ходов при объявлении слов плохими и $2^{2n}\cdot 2^{n-2} = 2^{3n-2}$ ходов при объявлении слов плохими относительно других слов. Таким образом, обещанное условие на число ходов противника (меньше 2^{3n-1}) будет выполнено, если

$$|M|\cdot\left(2^{\max(2n+r)} + 2^{\max(q+p)}\right) + 2^{2n-1} + 2^{3n-2} < 2^{3n-1} \qquad (**)$$

Учитывая, что $2^k + 2^l$ близко к $2^{\max\{k,l\}}$, легко заметить, что условия $(*)$ и $(**)$ будут выполнены, если все тройки $\langle p,q,r\rangle \in M$ удовлетворяют неравенствам

$$\begin{aligned} p + r &< 2n - 3\log n - O(1);\\ 2n + r &< 3n - 3\log n - O(1);\\ p + q &< 3n - 3\log n - O(1). \end{aligned}$$

(поскольку таких троек не более $O(n^3)$, и мы как раз вычли $3\log n + O(1)$ как верхнюю оценку для $\log|M|$). Мы не указываем явно константу в $O(1)$ но, скажем, заведомо годится 10. Таким образом, мы можем включить в M все тройки, удовлетворяющие этим неравенствам, и в этом случае в описанной игре существует выигрышная стратегия.

Эту стратегию мы будем применять против стратегии противника, которая игнорирует делаемые нами ходы. Она объявляет плохими слова длины $2n$, имеющие сложность менее $2n-1$, объявляет плохими относительно слова B все слова, имеющие условную сложность менее $n-2$, а также объявляет u-простыми [относительно простыми] все слова, сложность [соответственно, относительная сложность] которых меньше u. (Точнее говоря, этим правилом руководствуется каждый $\langle p, q, r\rangle$-игрок для каждой тройки $\langle p, q, r\rangle \in M$.)

Запустим две описанные нами стратегии играть друг против друга. Это будет алгоритмический процесс, для задания которого достаточно знать число n. Момент его окончания алгоритмически найти нельзя, но он существует, и в этот момент найдётся пара слов $\langle A, B\rangle$, которая обеспечивает наш выигрыш. Покажем, что для них выполнено утверждение теоремы.

Длина слов A и B равна 2n; сложность равна $2n+O(1)$ (она не может быть меньше, так как иначе они были бы объявлены плохими). Поскольку в нашем процессе мы для каждого B указываем не более 2^n различных слов A, то $\mathit{KS}(A|B) \leqslant n + O(\log n)$; сложность $\mathit{KS}(A|B)$ не может быть меньше $n - O(1)$, иначе слово A было бы объявлено плохим относительно B.

Пусть дано произвольное слово B', для которого $\mathit{KS}(B'|B) < r$ и сложность B' равна q. Как обычно, поскольку мы допускаем $O(\log n)$-погрешность, можно заменить слово B' его кратчайшим описанием и без ограничения общности считать, что длина слова B' не превосходит q. Можно считать также, что r много меньше n, скажем, $r < n/2$ (иначе слагаемое $O(r)$ в правой части утверждения теоремы делает его тривиальным). Тогда второе неравенство из трёх выполнено. Значит, нарушается либо первое, либо третье неравенство (поскольку противник проигрывает), поэтому точка $\langle p, q\rangle$ после сдвига не более чем на r попадает вправо от одной из прямых рис. 46. Остаётся вспомнить о необходимом условии

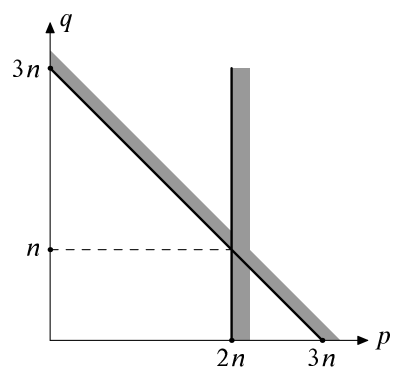

Рис. 46. Неравенства на p и q.

[multi-pic15]

рис. 44, и получится достаточное условие рис. 45.

Теорема 210 доказана. ▷

Приведём другое доказательство того же утверждения, использующее вероятностный метод. Оно во многом параллельно игровому, но есть важные различия.

Во-первых, вместо того, чтобы реагировать на ходы противника, мы сделаем все ходы сразу же и будем надеяться, что при любых действиях противника мы выиграем.

Во-вторых, мы теперь не указываем выигрышные ходы явно, а доказываем, что случайный ход с положительной вероятностью окажется выигрышным.

Техническое замечание: при таком подходе можно без ограничения общности считать, что оба игрока указывают максимальное разрешённое число слов.

Перейдём к формальному изложению, отмечая параллели с игровым доказательством в квадраных скобках. Пусть

$$U\colon \mathbb{B}^{2n} \times \mathbb{B}^{n} \mapsto \mathbb{B}^{2n}$$

[значения $U(B,X)$ при данном B и всевозможных X соответствуют n-простым относительно B словам в игровом доказательстве]. Пусть фиксировано некоторое конечное множество M троек натуральных чисел и для каждой тройки $\langle p,q,r\rangle \in M$ имеются два отображения

$$V_{p,q,r}\colon \mathbb{B}^{2n} \times \mathbb{B}^{r} \to \mathbb{B}^{q}$$

и

$$W_{p,q,r}\colon \mathbb{B}^{q} \times \mathbb{B}^{p} \to \mathbb{B}^{2n}.$$

Пусть, кроме того, имеется отображение

$$S\colon \mathbb{B}^{2n-2} \to \mathbb{B}^{2n},$$

а также отображение

$$T\colon \mathbb{B}^{2n} \times \mathbb{B}^{n-2\log n} \to \mathbb{B}^{2n}.$$

[Отображения $V_{p,q,r}$ и $W_{p,q,r}$ соответствуют ходам $\langle p,q,r\rangle$-игрока в команде противника. Именно, слова $V_{p,q,r}(B,X)$ при всевозможных X суть r-простые относительно B слова; слова $W_{p,q,r}(B',X)$ при всевозможных X суть p-простые относительно B' слова. Отображения S и T соответствуют ходам двух дополнительных игроков. Именно, $S(X)$ суть плохие слова в количестве не более 2^{2n-2} штук, а $T(B,X)$ суть плохие относительно B слова в количестве не более чем $2^{n-2\log n}$ штук. Границы для числа плохих слов уменьшены по сравнению с игровым доказательством: вместо $2n-1$ взято $2n-2$, а вместо $n-2$ взято $n-2\log n$. Это будет использовано в оценках.]

Будем говорить, что отображение U покрыто четвёркой V,W,S,T (состоящей из двух семейств отображений и двух отображений), если для всякого слова $B \in \mathbb{B}^{2n}$ и для всякого слова A, равного $U(B,X)$ при некотором $X \in \mathbb{B}^{n}$ [для любого слова B и любого слова A, объявленного нами n-простым относительно B], пара $\langle A,B\rangle$ удовлетворяет одному из четырёх условий:

(1) Слово B входит в область значений отображения S (то есть $B = S(Y)$ при некотором $Y \in \mathbb{B}^{2n-2}$). [Слово B противник объявил плохим.]

(2) Слово A входит в область значений отображения S (то есть $A = S(Y)$ при некотором $Y \in \mathbb{B}^{2n-2}$). [Слово A противник объявил плохим.]

(3) Слово A равно $T(B,Y)$ при некотором $Y \in \mathbb{B}^{n-2\log n}$. [Слово A противник объявил плохим для данного B.]

(4) Найдётся тройка $\langle p,q,r\rangle \in M$ и слово B' длины q, для которых

(а) $B' = V_{p,q,r}(B,Y)$ для некоторого $Y \in \mathbb{B}^r$ [$\langle p,q,r\rangle$-игрок команды противника объявил, что слово B' является r-простым относитльно B];

(б) $A = W_{p,q,r}(B',Z)$ для некоторого $Z \in \mathbb{B}^p$ [$\langle p,q,r\rangle$-игрок команды противника объявил, что слово A является p-простым относительно B'].

Мы докажем (при некотором условиях на множество M, указанных далее), что существует отображение U, не покрытое ни одной четвёркой V, W, S, T. Доказательство вероятностное: мы подсчитаем для данной четвёрки, сколько отображений U ими покрывается (т.е. вероятность для случайного отображения U оказаться покрытой), умножим эту вероятность на число четвёрок и установим, что произведение останется меньше 1.

Подсчёт для одной четвёрки. Пусть фиксированы V, W, S, T. Имеется не менее 2^{2n-1} слов B длины $2n$, нарушающих условие (1). Чтобы U было покрыто, необходимо, чтобы при любом таком B каждое из 2^n значений $A = U(B,X)$ для всех X длины n было покрыто одним из условий (2)–(4). Мы убедимся, что для данных B и X вероятность такого события не больше $1/2$. Из независимости следует, что вероятность для случайного U оказаться покрытым не больше

$$(1/2)^{2^{2n}\times 2^n} = (1/2)^{2^{3n}}.$$

Для данных B и X: не годятся слова A, покрытые условием (2) в количестве 2^{2n-2} штук, покрытые условием (3) в количестве $2^{n-2\log n}$ штук, а также покрытые условием (4) слова $W_{p,q,r}(V_{p,q,r}(B,Y),Z)$ в количестве $2^r \times 2^p$ штук для каждой тройки $\langle p,q,r\rangle \in M$. Всего получается не больше

$$2^{2n-2} + 2^{n-2\log n} + 2^{r+p} \cdot |M|,$$

что не больше 2^{2n-1} (половины всех слов), если

$$r + p + \log|M| < 2n - 3 \qquad (*)$$

при всех $\langle p,q,r\rangle \in M$ (это и будет первое из списка требований к M).

Осталось оценить количество всех четвёрок V, W, S, T. Для $V_{p,q,r}$ (при данных p, q, r) имеется не более

$$(2^q)^{2^{2n}\times 2^r} = 2^{q\cdot 2^{2n+r}}$$

вариантов, для $W_{p,q,r}$ (при данных p, q, r) имеется не более

$$(2^2 n)^{2^q\times 2^p} = 2^{2n\cdot 2^{q+p}}$$

вариантов, для S имеется не более

$$(2^{2n})^{2^{2n-2}} = 2^{2n\cdot 2^{2n-2}}$$

вариантов, для T имеется не более

$$(2^{2n})^{2^{2n}\times 2^{n-2\log n}} = 2^{(2^{3n}/n)+1}$$

вариантов. Первые две оценки возводим в степень $|M|$ и получаем оценку на число вариантов для V и W в целом, после чего полученные оценки перемножаем и заключаем, что общее количество четвёрок V, W, S, T не превосходит

$$2^{q\cdot 2^{2n+r}\times|M|} \cdot 2^{2n\cdot 2^{q+p}\times|M|} \cdot 2^{2n\cdot 2^{2n-2}} \cdot 2^{(2^{3n}/n)+1},$$

а двоичный логарифм этого числа не превосходит

$$q \cdot 2^{2n+r} \times |M| + 2n \cdot 2^{q+p} \times |M| + 2n \cdot 2^{2n-2} + (2^{3n}/n) + 1,$$

что будет меньше 2^{3n} (как требуется для завершения доказательства), если

$$2n + r + \log q + \log |M| < 3n - O(1) \qquad (**)$$

и

$$q + p + \log n + \log |M| < 3n - O(1). \qquad (***)$$

(Мы пользуемся тем, что $2^a + 2^b$ отличается не более чем в константу раз от $2^{\max(a,b)}$; два других условия $2n - 2 + \log 2n < 3n - O(1)$ и $2^{3n}/n + 1 < 3n - O(1)$ выполнены автоматически.)

Все три условия $(*)$–$(***)$ будут заведомо выполнены, если

$$p + r < 2n - 3\log n - O(1);$$
$$2n + r < 3n - 4\log n - O(1);$$
$$p + q < 3n - 4\log n - O(1)$$

для всех $\langle p, q, r\rangle$ из M, поскольку в таком случае $|M| = O(n^3)$.

Поэтому, взяв в качестве M множество троек, удовлетворяющих трём указанным только что неравенствам, мы заключаем, что существует отображение U, не покрытое никакой четвёркой V, W, S, T (при этом M).

При известном n такое отображение U можно найти перебором, поэтому (важный момент!) первое отображение U с такими свойствами будет иметь сложность не больше $\log n$.

Возьмём это U и конкретную четвёрку V, W, S, T, которой оно (по доказанному) не покрыто. А именно, пусть $\{S(\cdot)\}$ (множество значений отображения S) — все слова длины $2n$, имеющие сложность менее $2n - 2$. Пусть при любом $B \in \mathbb{B}^{2n}$ множество $\{T(B,\cdot)\}$ составляют все слова, имеющие условную сложность (относительно B) менее $n - 2\log n$. Пусть при данных p, q, r среди $V_{p,q,r}(B,\cdot)$ встречаются все слова условной сложности (относительно B) менее r, а среди $W_{p,q,r}(B',\cdot)$ встречаются все слова условной сложности (относительно B') менее p.

Раз U не покрыто, найдётся пара слов A и B, не обладающая ни одним из свойств (1)–(4). Отсюда следует, что $KS(A) = 2n + O(1)$ (слово имеет длину $2n$ и не может иметь сильно меньшую сложность, иначе его покрывает S). Аналогично $KS(B) = 2n + O(1)$. Условная сложность $KS(A|B)$ равна $n + O(\log n)$: она не может быть больше, так как $A = U(B, X)$ для некоторого слова X длины n, а сложность U есть $O(\log n)$; она не может быть меньше, так как иначе было бы выполнено свойство (3). Наконец, ни при каких p, q и r не найдётся слова B' длины q, при котором $KS(B'|B) < r$ и $KS(A|B') < p$, иначе было бы выполнено свойство (4).

Далее рассуждаем как раньше (в игровом доказательстве); на этом вероятностное доказательство теоремы 210 заканчивается.

Наконец, можно привести и «геометрическую» конструкцию, доказывающую то же самое утверждение. (В отличие от задачи выделения общей информации, геометрическое рассуждение даёт примерно те же оценки сложности.)

А именно, возьмём поле из 2^n элементов (или примерно такого количества, если мы хотим ограничиться вычетами по простому модулю) и рассмотрим двумерную плоскость над этим полем. Пусть $\langle A, B\rangle$ — случайная пара, состоящая из точки этой плоскости и проходящей через неё прямой. Тогда сложности как раз такие, как требуется в теореме 210.

Далее, пусть имеется слово B', для которого

$$KS(B'|B) \leqslant r, \quad KS(B') \leqslant q, \quad KS(A|B') \leqslant p. \qquad (*)$$

Мы должны показать, что пара $\langle p, q\rangle$ находится в $O(r)+O(\log n)$ окрестности множества G, установив, что в противном случае пара $\langle A, B\rangle$ имела бы меньшую сложность. В самом деле, оценим количество пар $\langle A, B\rangle$, для которых выполняются условия $(*)$ при некотором B'. Каждое из 2^q слов B' задаёт два множества:

- те слова A (длины $2n$), при которых $KS(A|B') \leqslant p$ (обозначим это множество $U_{B'}$);
- те слова B (длины $2n$), при которых $KS(B'|B) \leqslant r$ (обозначим его $V_{B'}$).

Множество $U_{B'}$ содержит 2^p элементов (точнее, $O(2^p)$, но для простоты записи мы ограниченные множители опускаем). Множество $V_{B'}$ может иметь разный размер (это зависит от B'), но известно, что эти множества покрывают множество из 2^{2n} слов длины $2n$ не более чем в 2^r слоёв (для каждого B есть не более 2^r слов B', простых относительно него).

Нам надо доказать, что объединение всех $U_{B'} \times V_{B'}$ покрывает лишь небольшую часть всех пар (так что случайная пара в него не попадёт). Для оценки количества покрытых пар применим уже известное свойство графа инцидентности (две прямые не могут проходить через две точки) и вытекающую из него оценку (лемма о четырёхугольниках, с. 321).

Нам будет удобно использовать утверждение этой леммы в такой форме (равносильной, как легко проверить): если в прямоугольной таблице $l \times L$ расставлены звёздочки, и никакие четыре не стоят на пересечении двух строк и двух столбцов, то число звёздочек не превосходит

- $O(L)$ при $l \leqslant \sqrt{L}$;
- $O(l\sqrt{L})$ при $l \geqslant \sqrt{L}$.

При этом следует отдельно рассматривать случай «больших» и «малых» $V_{B'}$. Начнём с первого. Если $V_{B'}$ содержит больше $\sqrt{|U_{B'}|}$, то есть больше $2^{p/2}$ элементов, то число покрытых пар не больше $2^{p/2}|V_{B'}|$ элементов для данного B'. Сумма по всех B' не больше $O(2^{p/2}2^{2n}2^r)$ (множество размера 2^{2n} покрыто не более чем в 2^r слоёв).

Теперь перейдём к малым $V_{B'}$. Для них $U_{B'} \times V_{B'}$ покрывает не более $O(2^p)$ элементов, и всего для 2^q различных B' получается $O(2^{p+q})$ слоёв.

Таким образом, если $p + q < 3n - O(\log n)$ и $(p/2) + 2n + r < 3n - O(\log n)$, то случайная пара $\langle A, B\rangle$ не будет обслужена ни одним из слов B'. (Тут ещё надо заметить, что множество обслуженных пар можно перечислять, зная n, p, q, r, то есть $O(\log n)$ битов информации.) Второе неравенство можно переписать как $p + 2r < 2n$; хотя это и немного хуже, чем неравенство $p+r < 2n$, которое было в первом доказательстве, но всё равно мы остаёмся в пределах оценки $O(r)$, упомянутой в теореме.

Третье доказательство теоремы 210 закончено.

Замечание. Это третье доказательство позволяет получить простое множество пар, большинство пар в котором удовлетворяют теореме (другими словами, позволяют получить стохастическую в смысле раздела 16.2 пару, удовлетворяющую условиям теоремы).

Того же самого, хотя и не в столь наглядной форме, можно добиться и модификацией второго (вероятностного) доказательства. Мы считали отображение U покрытым, если для всех пар определённого вида нечто верно; ослабим это условие и будем говорить, что U покрыто, если для половины пар нечто верно.

Для доказательства существования непокрытого U можно воспользоваться следующей (тривиальной) оценкой. Если вероятность каждого из независимых 2^k событий не больше $1/16$, то вероятность того, что произойдёт не менее 2^{k-1} событий, не больше $2^{2^k} \cdot (1/16)^{2^k/2} = 2^{-k}$.) Поэтому можно заменить вероятность $1/2$ на $1/16$, а всё остальное как раньше.

(Можно и иначе: не рассматривать множеств S и T и параллельно все допустимые тройки, а доказывать, что с близкой к единице вероятностью для случайно выбранного U доля пар, для которых выполнено условие (4), мала. Эти малые доли и малые отклонения от единицы затем складываются для всех троек из M.)

13. Информация и логика

[zur-deutung]

13.1. Задачи, логические операции, сложность

Под *алгоритмической задачей* будем понимать произвольное множество двоичных слов, а под *решением* задачи — любой его элемент.

Смысл этого определения можно пояснить следующим образом. Решение любой точно поставленной математической задачи можно записать на некотором подходящем формальном языке. Более того, такая запись должна быть конечной, чтобы её можно было использовать в математической практике. Поэтому можно считать, что любое решение любой задачи закодировано при помощи некоторого двоичного слова. Мы будем интересоваться лишь количеством информации в решениях задачи, а не содержательной её стороной, поэтому задачу естественно отождествлять с множеством её решений.

Определим сложность задачи как наименьшую сложность её решений:

$$KS(A) = \min\{KS(x) \mid x \in A\}.$$

(при этом минимум пустого множества полагаем равным $+\infty$).

Например, сложность синглетона $\{x\}$ совпадает со сложностью самого x. Другой пример: в разделе 1.2 мы сталкивались с задачей описания натурального числа не менее данного n. Ее сложность обозначалась через $KS_{\geqslant}(n)$. Используя новую терминологию, можно сказать, что $KS_{\geqslant}(n)$ есть сложность задачи, множество решений которой состоит из всех чисел, больших или равных n.

Если есть две задачи X и Y, то можно рассмотреть задачу «решить *обе* задачи X и Y», а также задачу «решить *хотя бы одну* из задач X и Y (указав при этом, какая именно из двух задач решена)». Решениями задачи «X и Y» должны быть пары, первая компонента которых есть решение задачи X, а вторая — задачи Y. А решениями задачи «X или Y» должны быть решения любой из задач с указанием, какая именно из двух задач решена. Таким образом, мы приходим к таким определениям логических операций над задачами:

$$X \wedge Y = \{[x, y] \mid x \in X, y \in Y\},$$
$$X \vee Y = \{[0, x] \mid x \in X\} \cup \{[1, y] \mid y \in Y\}.$$

Здесь $[x, y]$ обозначает код пары слов $\langle x, y\rangle$ при некотором вычислимом однозначном кодировании пар слов словами.

Например, сложность задачи $\{x\}\wedge\{y\}$ равна сложности пары $\langle x, y\rangle$, а сложность задачи $\{x\} \vee \{y\}$ равна минимуму из сложностей x и y (с точностью до $O(1)$). И вообще, для любых задач X, Y сложность задачи $X \vee Y$ равна наименьшей из сложностей задач X и Y (с точностью $O(1)$).

Задача $X\wedge Y$ называется *конъюнкцией* задач X и Y, а задача $X\vee Y$ — ***дизъюнкцией*** задач X и Y. Есть альтернативное естественное определение дизъюнкции задач, отражающее следующую идею. Можно разрешить решающему написать два слова, из которых первое должно быть решением первой задачи или второе — решением второй задачи, при этом решающий может и не знать, какой из двух случаев имеет место (что-то похожее происходит на письменных экзаменах по математике). Этой идее соответствует такое формальное определение «псевдодизъюнкции» задач:

$$X \tilde{\vee} Y = \{[x, y] \mid x \in X \text{ или } y \in Y\}.$$

Сложность псевдодизъюнкции отличается от сложности дизъюнкции всего лишь на $O(1)$, однако это принципиально разные задачи, о чем мы поговорим позже.

251 Докажите, что $KS(X \tilde{\vee} Y) = KS(X \vee Y) + O(1)$.

Условную сложность $KS(x|y)$ тоже можно понимать как сложность некоторой алгоритмической задачи, а именно задачи преобразования x в y. Эта задача получается из задач $\{x\}$ и $\{y\}$ помощью операции над задачами, называемой *импликацией*. Для того, чтобы её определить, зафиксируем некоторую главную универсальную вычислимую функцию $U : \Xi \times \Xi \to \Xi$. Это означает, что для любой вычислимой функции $V : \Xi \times \Xi \to \Xi$ найдется всюду определенная вычислимая функция $t : \Xi \to \Xi$ такая, что $U(t(p), x) = V(p, x)$ для всех двоичных слов p, x. Функция U не обязана быть оптимальной. Будем сокращать далее $U(p, x)$ как $[p](x)$.

Итак, положим

$$X \to Y = \{p \mid \forall x\, (x \in X \;\Rightarrow\; [p](x) \text{ определено и } [p](x) \in Y)\}.$$

Например, в разделе 6.4 мы изучали величину, $KS(x \mid\geqslant n)$, которая определялась, как наименьшая сложность программы, которая выдает x, получив на вход любое натуральное число, не меньшее x. Используя новые обозначения, можно написать $KS(x \mid\geqslant n) = KS(\{m \in \mathbb{N} \mid m \geqslant n\} \to \{x\})$. Другой пример: сложность задачи $\{x\} \to \{y\}$ равна условной сложности y при известном x (с точностью до $O(1)$).

Некоторые теоремы из главы о передачи информации по сетям естественным образом могут быть сформулированы как оценки сложности некоторых задач.

А именно, решениями задачи $(x \to y) \wedge (y \to x)$ (мы опускаем скобки при записи синглетонов) являются пары программ, из которых первая преобразует x в y, а вторая — y в x. Мы рассматривали задачу о минимальной возможной сложности такой программы и установили, что она равна максимальной из условных сложностей $KS(y|x)$ и $KS(x|y)$ (с точностью до $O(\log KS(x, y))$).

Другой пример. Решениями задачи $(x \to z) \wedge (y \to z)$ являются пары программ, из которых первая преобразует x в z, а вторая — y в z. Мы [увы, пока что не!] доказали, что минимально возможная сложность такой пары примерно равна максимуму условных сложностей $KS(z|x)$ и $KS(z|y)$ (с точностью до $O(\log KS(x, y, z))$).

252 Найдите сложность задачи $a \to (b \to c)$.

[Указание. Сложность этой задачи такая же, как и у задачи $(a \wedge b) \to c$.]

253 [pr3] Найдите сложность задачи $a \wedge (b \to c)$ (с логарифмической точностью). [Ответ: $KS(a) + KS(c|a, b)$.]

254 Докажите, что сложность задачи $(x \vee y) \to (x \tilde{\vee} y)$ ограничена константой, однако сложность обратной задачи $(x \tilde{\vee} y) \to (x \vee y)$ такая же, как и у задачи $x \vee y$ с точностью до $O(\log n)$, где n наибольшая из длин x, y. [Указание. Пусть p любое решение задачи

$(x\tilde{\vee}y) \to (x \vee y)$. Скажем, что пара слов (u, v) длины не больше n согласована с p, если для всех слов w длины не больше n программа p в применении к обоим словам $[u, w]$, $[w, v]$ дает некоторое решение задачи $u \vee v$. Тогда для любой согласованной с p пары $u = x$ или $v = y$.]

255 Докажите, что сложность задачи

$$((x\tilde{\vee}y) \to (x \vee y)) \to (x\tilde{\vee}y)$$

есть $O(\log n)$, где n наибольшая из длин x, y.

Две последних задачи объясняют разницу между обычной дизъюнкцией и псевдодизъюнкцией. В частности, из них следует, что задачи $x\tilde{\vee}x$ и x существенно различны, хотя они и имеют почти одинаковые сложности. А сложность задачи $(x\tilde{\vee}x) \to x$ оказывается примерно равной сложности самого x.

Импликация обладает следующим свойством: если задачи X и $X \to Y$ просты, то и задача Y проста:

$$KS(Y) \leqslant KS(X) + KS(X \to Y).$$

Это неравенство, верное с точностью $O(\log KS(X))$ (и с точностью $O(\log KS(X \to Y))$), является обобщением неравенства

$$KS(y) \leqslant KS(x) + KS(y|x),$$

верного для любых слов x, y также с логарифмической точностью. И так же, как неравенство для слов, может быть усилено:

$$KS(X \wedge Y) \leqslant KS(X) + KS(X \to Y).$$

Однако обратное неравенство уже неверно (в отличие от аналогичного неравенства для слов).

256 Приведите пример множеств X, Y, для которых $KS(X \wedge Y)$ значительно меньше, чем $KS(X)+KS(X \to Y)$. [Указание. В качестве X можно взять множество всех случайных слов длины n, а в качестве Y — множество всех случайных слов длины $2n$.]

Операции $\wedge, \vee, \to$, как операции над задачами, были определены Колмогоровым [21] и Клини [20] для интерпретации интуиционистского исчисления высказываний (IPC). Пусть $A(p, q, \dots)$ — пропозициональная формула со связками $\wedge, \vee, \to$. Для любых множеств слов $X, Y, \dots$ рассмотрим задачу $A(X, Y \dots)$, которая получится если подставить в эту формулу вместо переменных эти множества.

Теперь будет вполне уместным следующее замечание. В нашем определении операций над задачами имеется произвол: в выборе спаривающей функции $x, y \to [x, y]$, в выборе главной универсальной функции $U(p, x)$ и, наконец, в выборе слов $0, 1$ при определении дизъюнкции. Однако сложности любой задачи, посчитанные при двух различных выборах отличаются не более чем на константу. Точнее надо сказать так. Пусть $A(p, q, \dots)$ пропозициональная формула, а $X, Y, \dots$ произвольные множества. Пусть $A'(X, Y, \dots)$ и $A''(X, Y, \dots)$ две задачи, которые получились, если использовать разные спаривающие функции $[x, y]'$, $[x, y]''$, разные главные универсальные функции U' и U'', и, наконец, разные пары слов a', b' и a'', b'' в определении дизъюнкции. Тогда разность сложностей задач $A'(X, Y, \dots)$ и $A''(X, Y, \dots)$ ограничена константой по абсолютной величине. Доказать это можно по индукции. Точнее, по индукции для любой формулы можно построить две вычислимых функции: f_{12}^{A} преобразует любое решение задачи $A'(X, Y, \dots)$ в некоторое решение задачи $A''(X, Y, \dots)$, а f_{21}^{A} наоборот — любое решение задачи $A''(X, Y, \dots)$ — в некоторое решение задачи $A'(X, Y, \dots)$. Для пропозициональных переменных обе функции — это тождественная функция. Если формула A является конъюнкцией формул B и C, для которых обе функции уже определены, то f_{12}^{A} действует так: данное слово s представляем в виде $[u, v]'$, затем применяем к u и к v функции f_{12}^{A} и f_{12}^{B}, соответственно; полученные результаты спариваем с помощью второй спаривающей функции. Функция f_{21}^{A} действует аналогично. Если формула A есть дизъюнкция формул B и C, то построение функций для A из функций для B и C совершенно аналогично. Наконец, если формула A равна $B \to C$, то функция f_{12}^{A} определяется следующим образом. Рассмотрим вычислимую функцию $V(s, b) = f_{12}^{C}(U'(s, f_{21}^{B}(b)))$. Если s любое решение задачи $B'(X, Y, \dots) \to C'(X, Y, \dots)$, а b любое решение задачи $B''(X, Y, \dots)$, то $V(s, b)$ будет решением задачи $C''(X, Y, \dots)$. Поскольку U'' главная универсальная функция, существует всюду определенная вычислимая функция $t\colon \Xi \to \Xi$, для которой $U''(t(s), b) = V(s, b)$. Функция t и будет искомой функцией f_{12}^{A} для формулы A.

$\boxed{\textbf{257}}$ Провести это рассуждение подробно.

Если формула $A(p, q \dots)$ выводима в интуиционистском исчислении высказываний, то сложность задачи $A(X, Y, \dots)$ ограничена константой, не зависящей от $X, Y, \dots$. Более того, в этом случае существует строка s, которая является решением задачи $A(X, Y, \dots)$ для всех множеств $X, Y, \dots$. Это доказывается индукцией по длине вывода в IPC. Например, для формулы $p \to (q \to p)$ (одна из аксиом IPC) эта строка есть следующая программа — «преобразовать данное нам x в программу, которая, печатает x на любом входе».

$\boxed{\textbf{258}}$ Доказать, что для каждой выводимой в IPC формулы $A(p, q \dots)$ существует строка s, которая является решением задачи $A(X, Y, \dots)$ для всех множеств $X, Y, \dots$.

Верно и обратное: если формула $A(p, q, \dots)$ не выводится в IPC, то сложность задачи $A(X, Y, \dots)$ не ограничена константой. Более того, справедлива следующая теорема:

Theorem 211. `[th-int]` *Пусть пропозициональная формула $\Phi(t_1, \dots, t_k)$ со связками $\wedge, \vee, \to$ невыводима в IPC. Тогда найдется положительное ε и последовательность непустых конечных множеств $X_1^n, \dots, X_k^n$, состоящих из слов длины не более n, для которой сложность задачи $\Phi(X_1^n, \dots, X_k^n)$ не меньше εn при всех достаточно больших n.*

Это — главное утверждение настоящей главы. Мы его докажем, используя без доказательства некоторое чисто логическое утверждение о невыводимых в IPC формулах. В доказательстве этой теоремы важную роль играют оценки сложности задач, которые возникают при подстановке одноэлементных множеств в (невыводимые в IPC) пропозициональные формулы. Некоторые из таких задачи интересны сами по себе, как в уже приведенных примерах. Сейчас мы займемся анализом сложности тех из них, которые интересны с логической точки зрения.

Интересным примером с точки зрения сложности соответствующей задачи является формула

$$((p \to q) \to p) \to p.$$

Эта формула, называемая законом Пирса, является тавтологией (принимает значение ИСТИНА при любых значениях ИСТИНА/ЛОЖЬ для переменных), но невыводима в интуиционистском вычислении высказываний. При этом сложность задачи $((x \to y) \to x) \to x$ (как обычно, мы опускаем скобки при записи синглетонов) есть $O(\log n)$ для любых слов x, y длины не более n. Как мы уже говорили, подставляя вместо переменных в любую невыводимую в IPC формулу подходящие конечные непустые множества слов длины не более n, можно получить задачу сложности не меньше εn. В частности, это верно и для закона Пирса. То есть сложность задачи $((X \to Y) \to X) \to X$ мала по сравнению со сложностями X, Y, если X, Y — любые синглетоны, но может быть сравнимой со сложностями X, Y, если X, Y — произвольные непустые конечные множества.

Theorem 212. *Сложность задачи $((x \to y) \to x) \to x$ есть $O(\log n)$ для любых слов x, y длины не более n.*

$\lhd$ Нам достаточно указать алгоритм, который по n и любому решению задачи $(x \to y) \to x$ находит x. Этот алгоритм работает так. Пусть s — решение задачи $(x \to y) \to x$. Пусть $S = S_n$ обозначает множество всех слов длины не больше n. Для любой функции $\tau\colon S \to S$ зафиксируем некоторую программу l_τ, вычисляющую эту функцию. Назовем пару $(u, v) \in S \times S$ *согласованной* с s, если $[s](l_\tau) = u$ для всех $\tau\colon S \to S$ таких, что $\tau(u) = v$. Исходная пара (x, y) согласована с s. Имея s и n, мы начинаем перечислять согласованные с s пары и выдаем первую компоненту u первой появившейся пары (u, v). Докажем, что в самом деле $u = x$. Действительно, иначе существует функция τ, для которой $\tau(x) = y$ и $\tau(u) = v$, и программа s на входе l_τ выдает одновременно два различных слова u и x. $\rhd$

С формальной точки зрения в предыдущем доказательстве есть небольшой пробел. А именно, мы неявно использовали, что имеется алгоритм который находит программу l_τ, имея функцию τ (заданную словом, ей соответствующим при некотором вычислимом взаимно однозначном соответствии между Ξ и объединением по всем натуральным n множеств всех функций из S_n в S_n). Вычислимость функции $\tau \mapsto l_\tau$ (мы отождествляем функцию и соответствующее ей слово) доказывается так: рассмотрим вычислимую функцию $V(\tau, u) = \tau(u)$. Поскольку универсальная функция U является главной, существует всюду определенная вычислимая функция t из Ξ в Ξ такая, что $U(t(\tau), u) = V(\tau, u) = \tau(u)$. Если в качестве программы l_τ выбрать $t(\tau)$, то t и будет искомой вычислимой функцией.

$\boxed{\textbf{259}}$ Докажите, что в условии предыдущей теоремы можно заменить $O(\log n)$ на $O(\log k)$, где k — максимальная из сложностей x, y. [Указание. Надо перейти от слов x, y к их программам длины не более k относительно оптимального декомпрессора. В доказательстве следует заменить S на множество всех слов длины не больше k, а программа l_τ должна работать так: на входе u она находит первое описание слова u длины не более k, применяет к нему τ и применяет оптимальный декомпрессор к получившемуся слову.]

Из доказанного и неравенства $\mathit{KS}(Y) \leqslant \mathit{KS}(X) + \mathit{KS}(X \to Y) + O(\log \mathit{KS}(X))$ следует, что $K(x) \leqslant K((x \to y) \to x)) + O(\log n)$ (где n наибольшая из длин x, y). Интересно, что бывают формулы $A(p, q)$ и $B(p, q)$ такие, что для любых слов x, y сложность задачи $A(x, y)$ не превосходит сложности задачи $B(x, y)$, однако сложность импликации $A(x, y) \to B(x, y)$ велика по сравнению со сложностями x, y (при некоторых x, y). Вот пример таких формул: $(x \to y) \to y$ и $x \vee y$. Их сложности отличаются всего лишь на $O(\log n)$, что мы сейчас докажем, однако сложность импликации $((x \to y) \to y) \to (x \vee y)$ может быть примерно равна n, что мы тоже докажем.

Начнем с вычисления сложности задачи $(x \to y) \to y$.

Theorem 213. `[th-nested-impl]` *Сложность задачи $(x \to y) \to y$ равна сложности задачи $x \vee y$ (с точностью $O(\log n)$, где n наибольшая из длин x, y).*

$\lhd$ Нам достаточно предъявить алгоритм, который по любому решению задачи $x \vee y$ дает некоторое решение задачи $(x \to y) \to y$, и другой алгоритм, который по любому решению задачи $(x \to y) \to y$ и логарифмическому от n количеству дополнительной информации дает некоторое решение задачи $x \vee y$.

Первый алгоритм, получив на вход $[0, x]$ или $[1, y]$, должен найти некоторую программу, которая преобразует любое решение задачи $(x \to y)$ в y. Если нами получено слово $[1, y]$, то мы выдаем такую программу: не читая входа, печатать y. Если же нам дано слово слово $[0, x]$, то мы выдаем следующую программу: применяем данное нам решение задачи $(x \to y)$ к x, получая тем самым y, и печатаем его.

Теперь предъявим алгоритм, который по любому решению задачи $(x \to y) \to y$, числу n и еще одному дополнительному биту информации находит некоторое решение задачи $x \vee y$.

Пусть s любое решение задачи $(x \to y) \to y$. Обозначим через S множество всех слов длины не более n. Для любой функции $\tau\colon S \to S$ выберем программу l_τ, вычисляющую эту функцию, так, чтобы отображение $\tau \mapsto l_\tau$ было вычислимым. На этот раз назовем пару $(u, v) \in S \times S$ *согласованной с* s, если $[s](l_\tau) = v$ для всех τ таких, что $\tau(u) = v$.

По определению пара (x, y) согласована с s. Но могут существовать и другие пары, согласованные с s. Однако у любых двух согласованных с s пар (u', v') и (u'', v'') совпадают первые или вторые компоненты. Действительно, если первые компоненты различны, то существует функция τ, отображающая u' в v', а u'' в v''. Применение s к l_τ дает одновременно v' и v'', которые, следовательно должны совпадать.

Множество согласованных с s пар можно перечислять, зная s и n. Запустим процесс его перечисления. Пусть (u, v) — первая появившаяся пара. Если $u = x$, то мы знаем x и выдаём его. Иначе $v = y$ и мы выдаем y. Дополнительный бит информации нам нужен, чтобы понять, какой из двух случаев имеет место. $\rhd$

260 Вычислите сложность задачи $(x \to y) \to z$ (с точностью $O(\log n)$, где n наибольшая из длин x, y). [Указание. Она такая же, как и у задачи $z \vee (x \wedge (y \to z))$. Это можно доказать тем же способом, что и предыдущую теорему.]

[Решение. Нам достаточно предъявить алгоритм, который по любому решению вспомогательной задачи дает решение исходной задачи, и другой алгоритм, который по любому решению исходной задачи и логарифмическому от n количеству дополнительной информации дает решение вспомогательной задачи

Сначала предъявим первый алгоритм. По определению решением исходной задачи является любая программа, которая преобразует любое решение задачи $(x \to y)$ в z. Решения же вспомогательной задачи — это слово z и пары, состоящие из x и программы, переводящей y в z. Если данное нам решение вспомогательной задачи есть слово z, то мы выдаем следующее решение исходной задачи: не читая данной нам программы, печатать z. Если же данное нам решение вспомогательной задачи есть слово x и программа, преобразующая

y в z, то мы выдаем следующее решение исходной задачи: применяем данное нам решение задачи $(x \to y)$ к x, получая тем самым y. Затем применяем программу переводящую y в z к найденному y.

Теперь предъявим алгоритм, который по любому решению исходной задачи, числу n и еще одному дополнительному биту информации находит некоторое решение вспомогательной задачи.

Пусть s любое решение задачи $(x \to y) \to z$. Обозначим через S множество всех слов длины не более n. Для любой функции $\tau\colon S \to S$ зафиксируем некоторую программу l_τ, вычисляющую эту функцию. Назовем тройку $(u, v, w) \in S \times S \times S$ *согласованной с* s, если $[s](l_\tau) = w$ для всех τ таких, что $\tau(u) = v$.

По определению тройка (x, y, z) согласована с s. Множество согласованных с s троек можно перечислять, зная s и n. Запустим процесс его перечисления. Пусть (u, v, w) — первая появившаяся тройка. Рассмотрим два случая.

Первый случай: $w = z$. В этом случае мы знаем z и выдаём его.

Второй случай: $w \neq z$. Покажем, как в этом случае найти x и некоторое решение задачи $y \to z$. Слово x равно u. Действительно, если бы u не было равно x, то существовала бы функция τ, переводящая x в y, а u в v. Поскольку обе тройки (x, y, z) и (u, v, w) согласованы с s, применение s к l_τ должно было бы давать одновременно w и z. А это невозможно, поскольку $w \neq z$.

Нам осталось объяснить, как, имея y, найти z. Слово x мы уже нашли. Продолжаем перечислять согласованные тройки до тех пор, пока не найдем тройку, у которой первая компонента равна x, а вторая компонента равна y (такая тройка существует, например, исходная тройка (x, y, z)). Мы утверждаем, что найденная тройка и есть исходная. Действительно, если бы была согласованная с s тройка вида (x, y, w), отличная от исходной, то применение s к l_τ, где τ — любая функция, отображающая x в y должно было бы давать одновременно два различных слова z и w.

Дополнительный бит нам нужен, чтобы понять, какой из двух случаев имеет место.]

Theorem 214. *Сложность задачи* $((x \to y) \to y) \to (x \vee y)$ *равна* $\min(K(x|y), K(y|x))$ *(с точностью* $O(\log n)$*, где* n *— наибольшая из длин* x, y*.) В частности, если* x, y *случайные и независимые слова длины* n*, сложность этой задачи примерно равна* n*.*

$\lhd$ Легко убедиться, что сложность задачи $((x \to y) \to y) \to (x \vee y)$ не превосходит *KS*$(y|x)$ (с точностью $O(1)$): имея программу p, перерабатывающую x в y, а также любое решение s задачи $(x \to y) \to y$, можно найти y как $[s](p)$.

Докажем теперь, что сложность задачи $((x \to y) \to y) \to (x \vee y)$ не превосходит также $K(x|y)$ (с точностью $O(\log n)$). Для этого достаточно установить, что имея тройку, состоящую из n, любой программы p, преобразующей y в x, и любого решения s задачи $(x \to y) \to y$, можно найти x или y. Пусть S обозначает множество всех слов длины не больше n. Как и в предыдущем доказательстве, назовем пару $(u, v) \in S \times S$ *согласованной* с s, если $[s](l_\tau) = v$ для всех $\tau\colon S \to S$ таких, что $\tau(u) = v$.

Напомним, что у любых двух согласованных с s пар совпадают первые или вторые компоненты. Из этого следует, что либо первые компоненты всех согласованных пар равны x, либо вторые компоненты всех согласованных пар равны y (либо и то, и другое). Действительно, допустим имеются две согласованные с s пары (u', v'), (u'', v'') у первой из которых первая компонента u' отлична от x, а у второй — вторая компонента v'' отлична от y. Рассмотрим еще и пару (x, y), также согласованную с s. Тогда какие-то две из этих трех пар имеют одновременно разные первые и вторые компоненты: если $v' \neq y$, то такими парами будут (x, y) и (u', v'); если $u'' \neq x$, то ими будут (x, y) и (u'', v''); наконец, если $v' = y$ и $u'' = x$, то ими будут (u', v') и (u'', v'').

Имея n, p и s найдем первую пару (u, v), согласованную с s. Затем продолжим поиск других согласованных с s пар и одновременно запустим программу p на входе v. Как только мы найдем другую согласованную с s пару (u', v') или же обнаружим, что $[p](v) = u$, мы заканчиваем вычисление. В первом случае мы знаем x или y: если $v' \neq v$ то $x = u$, а если $u' \neq u$, то $y = v$. Во втором случае (когда $[p](v) = u$) мы знаем x, поскольку $x = u$. В самом деле, если бы слово x не совпадало с u, то тогда слово y должно было бы совпадать с v, а следовательно $u = [p](v) = [p](y) = x$. Заметим, что рано или поздно мы закончим вычисление: ведь если нет никаких других согласованных с s пар, кроме (u, v), то пара (u, v) совпадает с (x, y), а значит $[p](v) = u$.

Осталось доказать, что сложность задачи $((x \to y) \to y) \to (x \vee y)$ не может быть существенно меньше $\min\{K(y|x), K(x|y)\}$. Пусть нам дана программа p, решающая задачу $((x \to y) \to y) \to (x \vee y)$. Имея p мы построим пару программ (r_1, r_2), такую, что либо r_1 на входе x выдает y, либо r_2 на входе y выдает x (но при этом мы не будем знать, какой из случаев имеет место).

Мы знаем, что в применении к любому решению s задачи $(x \to y) \to y$ программа p выдает либо пару $[0, x]$, либо пару $[1, y]$. Будем применять программу p к программам следующего вида. Выберем любую пару A, B перечислимых неотделимых подмножеств $\mathbb{N}$. Для каждого натурального i и любых слов u, v определим следующую программу $q_i(u, v)$: получив на вход программу s запустить ее на входе u и параллельно перечислять A и B; если обнаружится, что программа s остановилась с результатом v, то выдать v и закончить вычисление; если обнаружится, что $i \in A$, то также выдать v и остановиться; наконец, если обнаружится, что $i \in B$ и вычисление s закончилось, то выдать $[s](u)$ и остановиться (даже если результат программы s не равен v). Первая и третья альтернатива могут обе осуществиться, наша программа в этом случае действует в зависимости от того, что обнаружится раньше. (Первая и вторая возможности тоже могут осуществиться обе, но выход все равно в них одинаков.) Если программа s является решением задачи $u \to v$, то во всех трех случаях программа выдаст v. Кроме того, в этом случае она обязательно остановится, поскольку применение s к u когда-нибудь закончится с результатом v. Поэтому для всех i, u, v программа $q_i(u, v)$ является решением задачи $(u \to v) \to v$. Пользуясь главностью универсальности функции U, можно считать, что отображение $i, u, v \mapsto q_i(u, v)$ вычислимо.

Полезные свойства программ вида $q_i(u, v)$ состоят в следующем: если i принадлежит A, то программа $q_i(u, y)$ является решением задачи $(x \to y) \to y$ при всех u. А если i принадлежит B, то наоборот программа $q_i(x, v)$ является решением задачи $(x \to y) \to y$ при всех v. Первое свойство дает нам возможность, имея p и y искать x следующим образом: перечисляем A и параллельно применяем программу p ко всевозможным программам вида $q_i(u, y)$, где $i \in A$, а u — любое слово. Применение p к любой такой программе должно остановиться с результатом $[0, x]$ или $[1, y]$. Если для хотя бы одной пары i, u нам повезет, и результатом окажется $[0, x]$, то мы узнаем x (о том, что нам повезло, мы поймем, увидев, что первая компонента результата равна 0).

Симметричным способом, используя второе свойство, можно, имея p и x, искать y: на этот раз поиск будет успешным, если для некоторого $i \in B$ и для некоторого слова v программа p на входе $q_i(x, v)$ выдает пару со первой компонентой 1.

Поэтому достаточно доказать, что хотя бы в одном из двух случаев нам повезет, то есть, существуют $i \in A$ и u, для которых программа p на входе $q_i(u, y)$ выдает $[0, x]$ или существуют $i \in B$ и v, для которых программа p на входе $q_i(x, v)$ выдает $[1, y]$. При этом в качестве u можно взять слово x, а в качестве v слово y. Действительно, для всех вообще натуральных чисел i программа p на входе $q_i(x, y)$ выдает $[0, x]$ или $[1, y]$, поскольку программа $q_i(x, y)$ является решением задачи $(x \to y) \to y$. Следовательно, множество

$$\{i \in \mathbb{N} \mid [p](q_i(x, y)) = [1, y]\}$$

разрешимо, а значит, не может отделять A от B. Это и означает, что выполнено одно хотя бы одно из двух условий нашего везения. $\rhd$

Заметим, что в доказательстве последнего пункта теоремы мы предъявили некоторую программу, которая решает задачу

$$(((x \to y) \to y) \to (x \vee y)) \to ((x \to y)\tilde{\vee}(y \to x))$$

для всех x, y длины не больше n.

В доказательстве теоремы 211 важную роль играют задачи, подобные задаче $((x \to y) \to y) \to (x \vee y)$.

Рассмотрим более общую задачу. Пусть даны $k \geqslant 2$ слов, которые мы обозначим $u_1, \ldots, u_k$. Пусть также даны два непустых непересекающихся подмножества I, J множества индексов $\{1, \ldots, k\}$. Рассмотрим задачу

$$((X \to Y) \to Y) \to Z,$$

где X есть конъюнкция синглетонов u_i, $i \in I$, (то есть, попросту кортеж состоящий из этих слов), Y — дизъюнкция синглетонов u_i, $j \in J$, а Z — дизъюнкция всех синглетонов $u_1, \ldots, u_k$. Например, при $k = 2$, $I = \{1\}$, $J = \{2\}$ мы получаем уже изученную задачу $((u_1 \to u_2) \to u_2) \to (u_1 \vee u_2)$.

Theorem 215. *Сложность задачи $((X \to Y) \to Y) \to Z$ не меньше минимальной из условных сложностей слов u_i при известных остальных словах из набора $u_1, \ldots, u_k$ (с точностью до $O(\log n)$, где n есть наибольшая из длин $u_1, \ldots, u_k$).*

$\lhd$ Нам достаточно построить k алгоритмов со следующим свойством. Для любой программы p, решающей задачу $((X \to Y) \to Y) \to Z$ найдется такое $i \leqslant k$, что i-ый алгоритм по p, кортежу слов $u_1, \ldots, u_k$ с пропущенным словом u_i и по n находит u_i. Нам будет удобно представлять множество решений задачи Z состоящим пар $\langle i, u_i \rangle$, где $i \leqslant k$, а множество решений задачи Y состоящим из таких же пар при $i \in J$.

Пусть дано решение p задачи $((X \to Y) \to Y) \to Z$. Мы знаем, что в применении к любому решению q задачи $(X \to Y) \to Y$ программа p выдает некоторую пару вида $\langle i, u_i \rangle$. Будем применять программу p к программам следующего вида. Фиксируем любую универсальную вычислимую функцию $T : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. Пусть j некоторое натуральное число и пусть $v = \langle v_1, \ldots, v_k \rangle$ некоторый кортеж слов, отличающийся от исходного кортежа

$u = \langle u_1, \ldots, u_k \rangle$ не более, чем в одной координате. По j и v можно найти некоторую программу $q_j(v)$, являющуюся решением задачи

$$(X \to Y) \to Y,$$

при том условии, что кортежи u и v отличаются в i-ой координате и $T(j,j) = i$ или вовсе не различаются (в последнем случае не важно, чему равно $T(j,j)$ и определено ли оно). Чтобы объяснить, зачем нам нужна такая странная программа, продолжим доказательство, предположив, что это возможно.

Алгоритм поиска u_i по данным p, n и кортежу u с пропущенным u_i действует так. Мы перебираем всевозможные кортежи v, отличающиеся от кортежа u не более, чем в i-ой координате, и перебираем все натуральные числа j. Если мы обнаружим, что $T(j,j) = i$ и первая компонента результата работы программы p на входе $q_j(v)$ также равна i, то мы узнае́м u_i. Действительно, по предположению программа $q_j(v)$ является решением задачи $(X \to Y) \to Y$, следовательно применение p к программе $q_j(v)$ обязательно закончится и результат принадлежит Y.

Мы утверждаем, что для хотя бы одного $i \leqslant k$ нам повезёт и это случится. То есть, существуют такие v, j, что $T(j,j) = i$ и первая компонента пары $[p](q_j(v))$ равна i. Действительно, в качестве v возьмём исходный кортеж u (перебор разных v нам всё равно необходим, поскольку кортежа u мы целиком не знаем). Первая компонента результата работы p на входе $q_j(u)$ при всех j находится в пределах $1, \ldots, k$. Рассмотрим вычислимую функцию, которая отображает j в первую компоненту пары $[p](q_j(u))$. Пусть l есть номер этой функции относительно универсальной функции T, то есть $T(l,j) =$ (первая компонента $[p](q_j(u)))$. Тогда при $j = l$ первая компонента пары $[p](q_j(u)$ будет равна $T(j,j)$.

Осталось объяснить, как же работает программа $q_j(v)$. Получив на вход программу s, запустим ее на входе V_I, состоящим из всех членов кортежа v с координатами из I, и параллельно вычисляем $T(j,j)$. В зависимости от того, какие вычисления и с каким результатом закончатся, мы делаем одно из трёх.

(1) Если обнаружится, что программа s остановилась с результатом вида $\langle i, v_i \rangle$, где $i \in J$, то выдаём $[i, v_i]$ и заканчиваем вычисление. Если s было решением задачи $X \to Y$, а кортежи u и v различаются не более, чем в одной координате, то то этот результат принадлежит Y. Действительно, если u и v совпадают, или различаются не в i-ой координате, то это очевидно. А иначе $V_I = X$ и, следовательно, результат работы s принадлежит Y.

(2) Если обнаружится, что вычисление $T(j,j)$ остановилось с результатом $i \in J$, а также вычисление s на входе V_I закончилось, то выдаем результат вычисления программы s и останавливаемся. Если s было решением задачи $X \to Y$, и кортежи u и v совпадают везде, кроме возможно, i-ой координаты, то этот результат принадлежит Y, поскольку $V_I = X$.

(3) Наконец, если вычисление $T(j,j)$ остановилось с результатом i, где $i \in I$, то выдаем пару любую пару $\langle l, v_l \rangle$, где $l \in J$ и останавливаемся. Выданный результат принадлежит Y, если кортежи совпадают везде кроме, возможно, i-ой координаты. $\triangleright$

Теперь ещё обобщим эту теорему. Пусть опять имеется кортеж слов $u = u_1, \ldots, u_k$. Пусть также имеется некоторое число N непересекающихся пар подмножеств множества индексов $\{1, \ldots, k\}$:

$$I_l \cap J_l = \varnothing, \qquad I_l, J_l \subset \{1, \ldots, k\}.$$

Для каждого $l \leqslant N$ определим задачи X_l и Y_l, как раньше, то есть, X_l есть конъюнкция синглетонов u_j для $j \in I_l$, а Y_l — дизъюнкция синглетонов u_l для $j \in J_l$. И пусть Z — дизъюнкция всех синглетонов $u_1, \dots, u_k$. Рассмотрим теперь задачу

$$\big(((X_1 \to Y_1) \to Y_1) \wedge \dots \wedge ((X_N \to Y_N) \to Y_N)\big) \to Z.$$

Theorem 216. [th-crit-impl] *Сложность этой задачи также не меньше минимальной из условных сложностей слов u_i при известных остальных словах из набора $u_1, \dots, u_k$ (с точностью до $O(\log n)$, где n есть наибольшая из длин $u_1, \dots, u_k$).*

$\lhd$ Доказательство этой теоремы в основном повторяет предыдущее доказательство. Для каждого $l \leqslant N$, каждого $j \in N$ и каждого кортежа $v = v_1, \dots, v_k$, отличающегося от кортежа u не более, чем в одной координате, мы, как и раньше, можем определить программу $q_{lj}(v)$, с таким свойством. Она является решением задачи

$$(X_l \to Y_l) \to Y_l,$$

если $T(j, j)$ определено и равно той координате, где кортежи u и v различаются, или если $v = u$ (в последнем случае неважно, определено ли и чему равно $T(j, j)$).

При любом $i \leqslant k$ можно опять рассмотреть следующий алгоритм поиска u_i, если дана любая программа p решения задачи

$$\big(((X_1 \to Y_1) \to Y_1) \wedge \dots \wedge ((X_N \to Y_N) \to Y_N)\big) \to Z,$$

а также даны n и все члены кортежа u, кроме i-ого. Применяем программу p к кортежу, состоящему из программ $q_{1j}(v)$, ..., $q_{Nj}(v)$. В качестве v пробуем все кортежи, отличающиеся от u не более, чем в i-ой координате, а в качестве j все такие натуральные числа, что $T(j, j) = i$. Если нам повезёт и для хотя бы для одной такой пары v, j программа p остановится и выдаст i, то мы найдём u_i. Так же, как и раньше, доказывается, что существует $i \leqslant k$, для которого такие v, j обязательно существуют, причем в качестве v можно взять сам кортеж u. $\rhd$

Пропозициональные формулы, соответствующие задачам, рассмотренным в этой теореме, называются *критическими импликациями*. А именно, так называются формулы вида

$$\big(((P_1 \to Q_1) \to Q_1) \wedge \dots \wedge ((P_N \to Q_N) \to Q_N)\big) \to R,$$

где R есть дизъюнкция переменных $s_1, \dots, s_k$, P_l — конъюнкции, а Q_l — дизъюнкции каких-то из этих переменных, причём P_l и Q_l не имеют общих переменных при каждом l. Критические импликации невыводимы в IPC (что следует из только что доказанной теоремы, а также может быть доказано с помощью моделей Крипке). Более того, они являются универсальными невыводимыми формулами в следующем смысле.

Theorem 217. [th-medvedev] *Для любой невыводимой в IPC пропозициональной формулы $A(t_1, \dots, t_m)$ со связками $\wedge, \vee, \to$ существуют число k, формулы $T_1, \dots, T_m$ с переменными $s_1, \dots, s_k$ и некоторая критическая импликация $J(s_1, \dots, s_k)$ такие, что формула*

$$A(T_1, \dots, T_m) \to J$$

выводима в IPC. При этом формулы $T_1, \dots, T_m$ содержат только связки $\wedge, \vee$.

Доказательство этой чисто логической теоремы выходит за рамки нашей книги. Интересующегося читателя мы отсылаем к работе [9], в которой приведено полное доказательство. Используя это утверждение, мы можем наконец доказать теорему 211.

$\lhd$ По предыдущей теореме существуют число k, формулы $T_1, \ldots, T_m$ с переменными $s_1, \ldots, s_k$ и связками $\wedge, \vee$ и некоторая критическая импликация $J(s_1, \ldots, s_k)$ такие, что формула

$$A(T_1, \ldots, T_m) \to J$$

выводима в IPC.

Возьмем случайные независимые слова $u_1, \ldots, u_k$ длины n/c (константу c подберем позже) и подставим одноэлементные множества, соответствующие этим словам, вместо переменных в формулы $T_1, \ldots, T_m$. Поскольку эти формулы не содержат импликации, мы получим конечные непустые множества слов $X_1, \ldots, X_m$. Формула $A(T_1, \ldots, T_m) \to J$ выводима в IPC, поэтому сложность задачи

$$A(X_1, \ldots, X_m) \to J(u_1, \ldots, u_k)$$

ограничена сверху константой. По теореме 216 сложность задачи $J(u_1, \ldots, u_k)$ не меньше $n/c - O(\log n)$. Поэтому и сложность задачи $A(X_1, \ldots, X_m)$ также не меньше $n/c - O(\log n)$, что больше $n/(2c)$ при всех достаточно больших n.

Осталось понять, почему длины всех слов в $X_1, \ldots, X_m$ ограничены сверху числом n при подходящем выборе константы c и при всех достаточно больших n. Любое слово из множеств $X_1, \ldots, X_m$ может быть получено из слов $u_1, \ldots, u_k$ с помощью фиксированного числа применений операции спаривания $v, w \mapsto [v, w]$ и операции спаривания с нулем или единицей. Если длина слова $[v, w]$ ограничена сверху линейной функцией от длин v, w, то утверждение очевидно. Ясно, что спаривающие функции с таким свойством существуют. Осталось заметить, что утверждение теоремы инвариантно относительно замены спаривающей функции на другую, поэтом мы можем предполагать, что спаривающая функция обладает этим свойством. $\rhd$

Доказательство этой теоремы опирается на недоказанную теорему 217. Существует доказательство немного более слабого утверждения, которое использует лишь существование контрмодели Крипке у невыводимых в IPC формул.

Theorem 218. *Пусть пропозициональная формула $\Phi(t_1, \ldots, t_k)$ со связками $\wedge, \vee, \to$ невыводима в IPC. Тогда найдется последовательность множеств $X_1^n, \ldots, X_k^n$, сложности не более $O(n)$, для которой сложность задачи $\Phi(X_1^n, \ldots, X_k^n)$ не меньше n при всех достаточно больших n.*

Разница с теоремой 211 в том, что мы теперь ограничиваем лишь сложность подставляемых множеств (а не длину слов в них). Кроме того мы не требуем их конечности.

$\lhd$ Пусть $\langle K, \leqslant \rangle$ — конечная модель Крипке, в корне которой ложна данная нам формула $\Phi(t_1, \ldots, t_k)$. Объясним, как по этой модели и произвольному числу n построить множества $X_1, \ldots, X_k$.

Отберем некоторое бесконечное множество длин, включающее нулевую длину, так чтобы любые две различные длины сильно различались (скажем не менее чем вдесятеро) и чтобы все ненулевые отобранные длины были значительно больше n. Будем называть такие длины *правильными*. Распределим правильные длины между мирами модели K так, чтобы каждому миру досталось бесконечное множество длин, причем нулевая длина досталась корню. И чтобы множества длин, сопоставленных каждому миру, было разрешимыми. Будем называть длины, отнесенные к миру u, *длинами из мира* u, слова этих длин, *словами из мира* u. Теперь мы можем объяснить как построить множество X_i, подставляемое вместо переменной t_i. Оно равно объединению следующих двух множеств слов. Первое состоит из множества всех случайных слов из миров, в которых истинна переменная t_i. Второе, обозначаемое C, не зависит от i и равно множеству всех пар вида $\langle x, y\rangle$, где x и y случайные слова из несравнимых миров. При этом слово считается случайным, если его сложность не меньше чем некоторая линейная функция его длины, скажем, не меньше десятой части длины. Случайным словом нулевой длины будем считать пустое.

Если соблюдать формальности, то в определении C вместо пары $\langle x, y\rangle$ следует брать ее код $[x, y]$, а при объединении множеств дописывать нуль или единицу в начало, чтобы элементы разного происхождения не смешивались:

$$C = \{[x, y] \mid x, y \text{ случайные слова из несравнимых миров}\},$$
$$X_i = \{0x \mid t_i \text{ истинна в мире } v, \text{ а } x \text{ случайное слово из мира } v\} \cup \{1y \mid y \in C\}.$$

Поскольку слова из разных миров имеют разные длины, по любому слову из X_i можно понять, из каких миров взялись составляющие его слова.

Мы докажем индукцией по построению формулы Ψ от переменных $t_1, \dots, t_k$, что множество решений задачи $\Psi(X_1, \dots, X_k)$ по существу совпадает с множеством X_Ψ, которое определяется по тому же правилу, что и значения переменных. По определению X_Ψ есть множество случайных слов из миров, в которых истинна формула Ψ, к которому еще добавлено C (соблюдая те же формальности, что и при определении X_i). Точнее мы докажем, что $\Psi(X_1, \dots, X_k)$ отличается от X_Ψ алгоритмическим преобразованием: для каждой формулы Ψ существуют две вычислимых функции f, g такие, что первая преобразует любое решение задачи $\Psi(X_1, \dots, X_k)$ в некоторый элемент X_Ψ, а вторая преобразует любой элемент множества X_Ψ в некоторое решение задачи $\Psi(X_1, \dots, X_k)$.

Если Ψ есть переменная, то множества $\Psi(X_1, \dots, X_k)$ и X_Ψ совпадают и обе функции тождественны. Поэтому достаточно доказать, что для множеств вида X_Ψ операции над задачами коммутируют с одноименными операциями над формулами:

(а) множество $X_\Psi \vee X_\Theta$ алгоритмически эквивалентно множеству $X_{\Psi\vee\Theta}$,

(б) множество $X_\Psi \wedge X_\Theta$ алгоритмически эквивалентно множеству $X_{\Psi\wedge\Theta}$,

(в) множество $X_\Psi \to X_\Theta$ алгоритмически эквивалентно множеству $X_{\Psi\to\Theta}$.

Кроме этого нам нужна еще устойчивость логических операций над задачами относительно алгоритмической эквивалентности, то есть, такое свойство: если множества U, V алгоритмически эквивалентны соответственно U', V', то множество $U \vee V$ алгоритмически эквивалентно множеству $U' \vee V'$ (и аналогичное свойство для других двух операций). Эти свойства очевидны, и мы переходим к доказательству (а), (б) и (в).

(а) По определению $X_{\Psi\vee\Theta}$ есть объединение множеств X_Ψ и X_Θ, поэтому по любому элементу $X_\Psi \vee X_\Theta$ легко указать некоторый элемент из $X_{\Psi\vee\Theta}$. В обратную сторону: если нам дан элемент объединения X_Ψ и X_Θ, то можно разобраться, какому из двух множеств он принадлежит (или обоим) и тем самым преобразовать его в некоторый элемент множества $X_\Psi \vee X_\Theta$.

(б) По определению $X_{\Psi\wedge\Theta}$ есть пересечение множеств X_Ψ и X_Θ, а $X_{\Psi\wedge\Theta}$ есть их декартово произведение. По любому элементу x из пересечения легко найти некоторый элемент декартова произведения, а именно $[x, x]$. В обратную сторону: пусть нам дан элемент $[x, y]$ из декартова произведения X_Ψ и X_Θ. Если хотя бы одно из слов x, y принадлежит C, то оно принадлежит и пересечению этих множеств. Пусть ни одно из них не принадлежит C, то есть x, y — это случайные слова из миров u, v, в которых истинны соответственно Ψ и Θ. Тогда эти миры мы можем найти.

Далее рассмотрим два случая.

(1) Эти миры несравнимы. Тогда мы просто выдаем пару $[x, y]$, которая по определению принадлежит C, а значит и пересечению X_Ψ и X_Θ.

(2) Миры u, v сравнимы, скажем, u предшествует v. Тогда в силу монотонности истинности, формула Ψ истинна также в мире v, а значит слово y принадлежит пересечению X_Ψ и X_Θ.

(в) Это наиболее интересный случай. Нам надо доказать, что по любому элементу из $X_{\Psi\to\Theta}$ можно найти некоторый элемент $X_\Psi \to X_\Theta$ и, обратно, по любому элементу из $X_\Psi \to X_\Theta$ можно найти некоторый элемент $X_{\Psi\to\Theta}$. Сначала докажем первое утверждение.

Пусть нам дано слово x из $X_{\Psi\to\Theta}$. Нам надо найти некоторое решение задачи $X_\Psi \to X_\Theta$. По другому можно сказать так: нам надо, имея x и любое слово y из X_Ψ, найти некоторое слово из X_Θ. Если хотя бы одно из слов x, y принадлежит C, мы выдаем это слово. Пусть ни одно из них не принадлежит C. Тогда x — случайное слово из некоторого мира u, в котором истинна формула $\Psi \to \Theta$, а слово y — случайное слово из некоторого мира v, в котором истинна формула Ψ (и эти миры мы можем найти, изучив длины x, y). Рассмотрим три случая.

(1) Если мир v предшествует миру u, то формула Ψ истинна и в мире u, следовательно и формула Θ истинна в мире u (напомним, что импликация $\Psi \to \Theta$ истинна в мире u). Значит слово x принадлежит множеству X_Θ.

(2) Если мир u предшествует миру v, то формула Θ истинна в мире v, поскольку в нем истинна формула Ψ и импликация $\Psi \to \Theta$ истинна в мире u. Значит слово y принадлежит множеству X_Θ.

(3) Если миры u и v несравнимы, то пара $[x, y]$ принадлежит C, а следовательно и X_Θ.

Осталось доказать, что по любому элементу из $X_\Psi \to X_\Theta$ можно найти некоторый элемент $X_{\Psi\to\Theta}$. Пусть нам дана программа r, преобразующая любое слово из X_Ψ в некоторое слово из X_Θ.

Если формула $\Psi \to \Theta$ истинна в корне, то пустое слово принадлежит $X_{\Psi\to\Theta}$ и мы выдаем его.

Пусть формула $\Psi \to \Theta$ ложна в корне. Тогда существует мир u, в котором истинна формула Ψ, но ложна Θ. Выберем правильную длину l из мира u, значительно большую длины программы r. Мы знаем, что в применении ко всем случайным словам длины l программа r остановится и выдаст слово из X_Θ. Но к сожалению мы не знаем, какие слова случайны, а какие нет. Поэтому будем применять параллельно программу r ко всем словам длины l. Если какое-то слово s получится более чем $2^{l/2}$ раз в качестве выхода r, то мы нашли нужное слово. Действительно, неслучайных слов длины l значительно меньше чем

$2^{l/2}$, а значит, хотя бы один раз слово s получено как результат работы на случайном слове. Следовательно, слово x принадлежит $X_\Theta \subset X_{\Psi\to\Theta}$.

Объясним, почему найдется слово, имеющее так много прообразов. Пусть x любое случайное слово длины l, а s результат работы программы r на входе x. Сложность s не превосходит суммы длин x и r, и поэтому значительно меньше $2l$. Кроме того, s принадлежит X_Θ, а, следовательно, имеет вид вид $0t$ или $1[y, z]$, где t, y, z случайные слова правильных длин. Поэтому длины t, y, z не могут быть больше l. Более того, длина t должна быть строго меньше l, так как слово t принадлежит миру, в котором истинна Θ, и этот мир не равен u (напомним, что Θ ложна в u). Значит длина t не больше $l/10$. Кроме того, слова y и z имеют разные длины, поскольку находятся в несравнимых мирах, следовательно одна из них должна быть не больше $l/10$. Если бы и другая длина тоже была всегда не больше $l/10$, то тогда множество результатов работы программы r на случайных входах имело бы мощность значительно меньше $2^{l/2}$. А следовательно, какое-то слово имело бы более $2^{l/2}$ прообразов.

Поскольку всё-таки длина одного из слов y, z может быть равна l, мы не можем гарантировать наличия слова с таким количеством прообразов. Тем не менее, наши рассуждения доказывают следующее. Либо некоторое слово s имеет более $2^{l/2}$ прообразов, либо верно следующее. Найдется слово y такое, что для более, чем $2^{l/2}$ слов x длины l программа r на входе x выдает слово вида $1[y, z]$ (или $1[z, y]$), где длина z равна l. В этом случае y обязано быть случайным словом из некоторого мира, не сравнимого с u. Действительно, хотя бы одно из слов x, на которых r выдает результат $1[y, z]$ (или $1[z, y]$) случайно, а значит результат работы r должен принадлежать X_Θ. А это может быть только когда y и z случайны и принадлежат несравнимым мирам. Поскольку, слово z принадлежит миру u (его длина равна l), слово y принадлежит миру, не сравнимому с u.

Применяя параллельно программу r ко всем словам длины l найдем такое слово s или такое y. Если реализовался первый случай и найдено слово s с большим числом прообразов, то оно принадлежит $X_\Theta \subset X_{\Psi\to\Theta}$.

Во втором случае поступим следующим образом. У нас есть случайное слово y из некоторого мира v, не сравнимого с u. Если в мире v истинна формула $\Psi \to \Theta$, то слово y и есть искомое. Иначе существует мир u_1 выше мира v, в котором истинна формула Ψ, но ложна Θ, и можно повторить все рассуждения для этого мира. В результате либо мы найдем некоторое слово из X_Θ, либо найдем случайное слово из некоторого мира v_1, не сравнимого с миром u_1. При этом мир v_1 не может быть ниже v (в этом случае он был бы и ниже u_1). Значит мир v_1 не сравним с миром v (в этом случае мы нашли пару случайных слов в несравнимых мирах и можем остановиться), либо строго выше него. Во втором случае (мир v_1 строго выше мира v), если формула $\Psi \to \Theta$ истинна в мире v_1, то можно остановиться. А иначе можно повторить все рассуждения для мира v_1. Рано или поздно мы остановимся, поскольку модель конечна и следовательно не содержит бесконечных возрастающих последовательностей миров.

Итак, мы доказали, что множество решений задачи $\Phi(X_1, \ldots, X_k)$ алгоритмически эквивалентно множеству X_Φ. По условию формула Φ ложна в корне, поэтому длины всех слов в X_Φ значительно больше n, значит сложность задачи $\Phi(X_1, \ldots, X_k)$ значительно больше n. Осталось понять, почему сложности всех множеств X_i меньше $O(n)$. Можно считать, что модель K содержит мир, в которой все формулы истинны (добавим новый мир, который больше всех остальных, и в котором все переменные истинны; на истинность формул в корне этот мир не повлияет). Также можно считать, что длина $2n$ правильная и принадлежит этому миру. Тогда множество случайных слов длины $2n$ принадлежит всем X_i. $\rhd$

Мы оценили сложности многих задач, полученных из одноэлементных множеств логическими операциями $\wedge, \vee, \to$. Могло создаться ошибочное впечатление, что сложность любой задачи, составленной из множеств $\{x\}, \{y\}, \ldots$ с помощью этих операций можно выразить через сложности слов $x, y, \ldots$, их пар, троек и т.д. Однако это не так. Задача $(x \to z) \wedge (y \to z)$ (сложность которой, напомним, примерно равна наибольшей из условных сложностей $\mathit{KS}(z|x)$, $\mathit{KS}(z|y)$) уже находится на грани той области, где это возможно. А именно, для задачи $(a \to c) \wedge (b \to d)$ это уже невозможно: существуют две четвёрки слов $\tilde{a}_n, \tilde{b}_n, \tilde{c}_n, \tilde{d}_n$ и $\bar{a}_n, \bar{b}_n, \bar{c}_n, \bar{d}_n$, для которых сложности этой задачи сильно отличаются, однако сложности $\tilde{a}_n, \tilde{b}_n, \tilde{c}_n, \tilde{d}_n$ близки соответственно к сложностям $\bar{a}_n, \bar{b}_n, \bar{c}_n, \bar{d}_n$, и то же самое верно для всех пар, составленных из a, b, c, d, троек и самой четвёрки.

$\boxed{\mathbf{261}}$ [pr1] Докажите, что с логарифмической точностью для любых a, b, c, d выполнены неравенства

$$\begin{aligned}
K((a \to c) \wedge (b \to d)) &\leqslant K(c|a) + K(d|b),\\
K((a \to c) \wedge (b \to d)) &\leqslant K(d|b, c) + K(c),\\
K((a \to c) \wedge (b \to d)) &\leqslant K(c|a, d) + K(d),\\
K((a \to c) \wedge (b \to d)) &\geqslant K(b, c, d|a) - K(b|a, c),\\
K((a \to c) \wedge (b \to d)) &\geqslant K(a, c, d|b) - K(a|b, d).
\end{aligned}$$

$\boxed{\mathbf{262}}$ Найдите последовательность четвёрок слов линейной от n длины, для которых наименьшая из трех верхних оценок сложности задачи $(a \to c) \wedge (b \to d)$ из задачи 261 превышает наибольшую из двух нижних оценок более, чем на $o(n)$. [Указание. Можно положить $a = d$ и $b = c$, где a, b — независимые случайные слова длины n.]

Theorem 219. *Существует положительное ε и последовательности четвёрок слов $\tilde{a}_n, \tilde{b}_n, \tilde{c}_n, \tilde{d}_n$ и $\bar{a}_n, \bar{b}_n, \bar{c}_n, \bar{d}_n$ линейной (от n) длины, для которых сложность задачи $(\tilde{a}_n \to \tilde{c}_n) \wedge (\tilde{b}_n \to \tilde{d}_n)$ более чем на εn превосходит сложность задачи $(\bar{a}_n \to \bar{c}_n) \wedge (\bar{b}_n \to \bar{d}_n)$. При этом сложности слов $\tilde{a}_n, \tilde{b}_n, \tilde{c}_n, \tilde{d}_n$ отличаются от сложностей $\bar{a}_n, \bar{b}_n, \bar{c}_n, \bar{d}_n$ не более, чем на $O(\log n)$ и то же самое верно для всех пар, троек и самих четвёрок.*

$\lhd$

Геометрическое доказательство.

Сначала сделаем простое замечание. Для четвёрки $\tilde{a}_n, \tilde{b}_n, \tilde{c}_n, \tilde{d}_n$ из условия теоремы оба неравенства из задачи 261, дающие нижнюю оценку сложности задачи $(a \to c) \wedge (b \to d)$, должны быть строгими — разница между левой и правой частями должна быть не меньше εn. Поэтому начнем с поиска такой четвёрки. А затем мы подберем другую четвёрку с теми же сложностями, для которой эта разница есть $o(n)$.

Будем использовать геометрическую конструкцию, которая нам помогла найти примеры слов, у которых не выделяется общая информация. Возьмем поле из 2^n элементов и будем рассматривать точки, прямые и плоскости в трёхмерном аффинном пространстве над этим полем. Всего имеется 2^{3n} точек. Количество прямых примерно равно 2^{4n} (прямая задается парой различных точек — 2^{6n} вариантов, причем каждая прямая таким образом может быть задана примерно 2^{2n} способами, так как содержит 2^n точек). Количество плоскостей примерно равно 2^{3n} (плоскость задается тремя точками — 2^{9n} вариантов, а поскольку каждая плоскость содержит 2^{2n} точек, она содержит примерно 2^{6n} троек точек). В качестве $\langle\tilde a, \tilde b\rangle$ мы возьмём случайную пару различных пересекающихся прямых, $\tilde c$ будет их точкой пересечения, а $\tilde d$ — плоскостью, в которой обе они лежат. Нетрудно посчитать, что для этой четвёрки обе нижние оценки из задачи 261 равны n (с точностью до $O(\log n)$).

Докажем, что

$$KS((\tilde a \to \tilde c) \wedge (\tilde b \to \tilde d)) \geqslant 1.5n$$

(с точностью до $O(\log n)$).

Пусть γ есть решение задачи $(\tilde a \to \tilde c) \wedge (\tilde b \to \tilde d)$, то есть γ — это пара программ $\langle\alpha, \beta\rangle$, первая из которых преобразует $\tilde a$ в $\tilde c$, а вторая $\tilde b$ в $\tilde d$. Рассмотрим множество S, состоящее из всех пар различных прямых a, b таких, что программа α преобразует a в общую точку a и b, а программа β преобразует b в плоскость, содержащую a и b. При известных α, β и n мы можем перечислять все элементы S. Так как пара $\langle\tilde a, \tilde b\rangle$ принадлежит S, мы можем заключить, что

$$7n \leqslant K(\tilde a, \tilde b) \leqslant K(\gamma) + \log|S|$$

(с точностью $O(\log\log|S|)$). Таким образом, нам достаточно доказать, что количество пар в S не превосходит $O(2^{5.5n})$. Это непосредственно следует из следующей комбинаторной леммы.

Лемма. Пусть f — некоторая функция, сопоставляющая каждой прямой некоторую точку на этой прямой, а g — некоторая функция, сопоставляющая каждой прямой некоторую плоскость, содержащую эту прямую. Пусть множество S состоит из всех пар прямых $\langle a, b\rangle$ таких, что точка $f(a)$ принадлежит b, а плоскость $g(b)$ содержит a. Тогда S содержит не более $O(2^{5.5n})$ пар.

$\lhd$ Сначала посмотрим, какая оценка получится, если не прибегать к хитростям. Для каждой прямой b существует примерно 2^{2n} прямых a в плоскости $g(b)$, поэтому мощность S не более, чем в 2^{2n} раз, превосходит количество прямых 2^{4n}, что даёт верхнюю оценку $|S| \leqslant 2^{6n}$. Такую же оценку мы получим, если для каждой прямой a подсчитаем количество прямых b, проходящих через $f(a)$. Заметим, что в первом подсчёте мы не использовали того, что прямая b должна содержать точку $f(a)$, а во втором — того, что прямая a должна лежать в плоскости $g(b)$. Наш план таков: мы модифицируем первое из рассуждений, показав, что в S в среднем на каждую прямую b приходится не более $2^{1.5n}$ прямых a. При этом мы уже будем учитывать условие $f(a) \in b$. (Можно действовать и симметричным образом — показать, что в S в среднем на каждую прямую a приходится не более $2^{1.5n}$ прямых b.)

Разобьём S на слои, поместив в один слой пары $\langle a, b\rangle$ с одинаковым значением $g(b)$. Все пары данного слоя лежат в одной плоскости, а сам слой однозначно задается этой плоскостью. Мы ограничим сверху количество пар в каждом слое, а затем просуммируем полученные оценки. Итак, фиксируем плоскость d и оценим сверху количество пар в S, лежащих в слое, соответствующем d.

Для этого рассмотрим произвольную точку c на плоскости d и обозначим через A_c множество прямых a в плоскости d с $f(a) = c$, а через B_c множество прямых b, содержащих точку c, для которых $g(b) = d$ (из условия следует, что прямая a содержит точку c, а прямая b содержится в плоскости d). Ясно, что размер слоя не превосходит

$$\sum_c |A_c||B_c| \leqslant \sqrt{\sum_c |A_c|^2 \sum_c |B_c|^2}.$$

Обе суммы под корнем легко ограничить сверху, поскольку они имеют ясный смысл. А именно, $\sum_c |A_c|^2$ пропорционально вероятности того, что для двух равномерно и независимо выбранных прямых a', a'' в плоскости d будет выполнено $f(a') = f(a'')$. Действительно, вероятность этого события есть сумма по всем c вероятности пересечения независимых событий $f(a') = c$ и $f(a'') = c$. Вероятности этих событий равны между собой и равны отношению мощности A_c к общему количеству прямых в плоскости d, примерно равному 2^{2n}. С другой стороны вероятность события $f(a') = f(a'')$ не превосходит примерно 2^{-n} (при любом фиксированном a' вероятность события $f(a') = f(a'')$ не превосходит вероятности того, что прямая a'' проходит через точку $f(a')$, что примерно равно 2^{-n}). Поэтому $\sum_c |A_c|^2$ не превосходит примерно

$$(2^{2n})^2 \cdot 2^{-n} = 2^{3n}.$$

Все эти подсчеты верны с точностью до мультипликативной константы.

Вторая сумма $\sum_c |B_c|^2$ связана со средним количеством общих точек у двух равномерно и независимо выбранных прямых b', b'' в множестве M_d, состоящем из всех прямых b, для которых $g(b) = d$ (все они лежат в плоскости d). Действительно, среднее количество общих точек b' и b'' равно сумме по всем c вероятности того, что точка c принадлежит одновременно b' и b''. События $c \in b'$ и $c \in b''$ независимы и вероятности их равны между собой и равны отношению мощности B_c к общему количеству прямых в M_d. Поэтому сумма $\sum_c |B_c|^2$ равна отношению среднего количества общих точек у b' и b'' к квадрату мощности M_d. С другой стороны, поскольку любые две прямые имеют не более одной общей точки, или совпадают (что бывает с вероятностью $1/|M_d|$), среднее количество общих точек не превосходит $1 + 2^n/|M_d|$. Поэтому

$$\sum_c |B_c|^2 \leqslant |M_d|^2(1 + 2^n/|M_d|) = |M_d|^2 + |M_d|2^n.$$

Напомним, что количество пар в слое, задаваемом плоскостью d, не превосходит $\sqrt{\sum_c |A_c|^2 \sum_c |B_c|^2}$, поэтому оно не больше чем

$$\sqrt{2^{3n}(|M_d|^2 + |M_d|2^n)}$$

(с точностью до умножения на постоянный множитель). Простым возведением в квадрат устанавливается, что эта величина не превосходит

$$2^{1.5n}(|M_d| + 2^n)$$

Осталось просуммировать полученные оценки по всем d:

$$|S| \leqslant 2^{1.5n} \sum_d (|M_d| + 2^n) = 2^{1.5n}(\sum_d |M_d| + \sum_d 2^n).$$

Семейства прямых M_d образуют разбиение множества всех прямых, следовательно сумма мощностей M_d равна общему количеству прямых 2^{4n}. Кроме того, количество различных плоскостей d примерно равно 2^{3n}, поэтому вторая сумма также примерно равна 2^{4n}. Лемма, а вместе с ней и нижняя оценка сложности задачи $(\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d})$, доказана. $\rhd$

Для завершения доказательства теоремы нам достаточно построить другую четвёрку $\langle \bar{a}, \bar{b}, \bar{c}, \bar{d} \rangle$ с теми же сложностями, что и у $\langle \tilde{a}, \tilde{b}, \tilde{c}, \tilde{d} \rangle$, для которой сложность задачи $(\bar{a} \to \bar{c}) \wedge (\bar{b} \to \bar{d})$ примерно равна n.

Для этого возьмём случайное слово длины $7n$ и разрежем его на 7 блоков u, v, w, p, q, r, s длины n. Положим $\bar{a} = uvws$, $\bar{b} = pqrs$, $\bar{c} = ups$, $\bar{d} = vqs$. Нетрудно проверить, что сложности слов обеих четвёрок, их пар и т. д. одинаковы и равны:

$$\begin{aligned}
&K(a) = K(b) = 4n, \quad K(c) = K(d) = 3n,\\
&K(a,b) = 7n, \quad K(a,c) = K(a,d) = K(b,c) = K(b,d) = K(c,d) = 5n,\\
&K(a,c,d) = K(b,c,d) = 6n, \quad K(a,b,c) = K(a,b,d) = K(a,b,c,d) = 7n.
\end{aligned}$$

При этом сложность задачи $(\bar{a} \to \bar{c}) \wedge (\bar{b} \to \bar{d})$ примерно равна n, поскольку, зная побитовую сумму слов p и v (по модулю 2), можно $\bar{a}$ преобразовать в $\bar{c}$, а $\bar{b}$ в $\bar{d}$. $\rhd$

$\lhd$

Вероятностное доказательство.

Фиксируем n. Мы определим четвёрку слов $\langle \tilde{a}, \tilde{b}, \tilde{c}, \tilde{d} \rangle$ длины n, сложности которых примерно равны n, сложности все пар, составленных из них, — примерно $2n$, сложности всех троек, а также сложность самой четвёрки — примерно $3n$.

Для этого будем рассматривать всевозможные множества четвёрок слов длины n. Назовём такое множество Q *равномерным*, если для каждой тройки $\langle a, b, c \rangle$ слов длины n существует ровно одно d, для которого четвёрка $\langle a, b, c, d \rangle$ принадлежит Q. То есть, оно задает функцию $\langle a, b, c \rangle \mapsto d$. Мы определим некоторое алгоритмически проверяемое свойство множеств, которое гарантирует, что для большинства четвёрок $\langle a, b, c, d \rangle$ в множестве Q сложность задачи $(a \to c) \wedge (b \to d)$ не меньше $2n - O(\log n)$. Затем, подсчетом вероятности того, что случайное равномерное множество обладает этим свойством, мы докажем непустоту этого свойства. Простым перебором по n можно будет найти множество с таким свойством, поэтому первое найденное множество будет иметь сложность порядка $O(\log n)$. В качестве $\langle \tilde{a}, \tilde{b}, \tilde{c}, \tilde{d} \rangle$ мы возьмём любую случайную четвёрку в этом множестве. Ее сложность будет равна примерно $3n$, как мы и хотели. Сложность тройки $\langle \tilde{a}, \tilde{b}, \tilde{c} \rangle$ также равна примерно $3n$, то есть она случайна. Поэтому случайны и все пары, составленные из $\tilde{a}, \tilde{b}, \tilde{c}$, и сами $\tilde{a}, \tilde{b}, \tilde{c}$.

Сложность задачи $(\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d})$ будет не меньше $2n - O(\log n)$. Действительно, множество всех четвёрок из Q, для которых она меньше, перечислимо по данному n. По условию, это его мощность значительно меньше Q. Следовательно, все они не случайны.

Кроме того, из сказанного следует, что и тройка $\langle \tilde{b}, \tilde{c}, \tilde{d} \rangle$ случайна. Действительно, сложность задачи $(\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d})$ не меньше $2n$ и ограничена сверху суммой $KS(\tilde{c}) + KS(\tilde{d}|\tilde{b}, \tilde{c}) \leqslant 2n$. Следовательно, оба слагаемых в этой сумме примерно равны n, то есть слово $\tilde{d}$ независимо от пары $\langle \tilde{b}, \tilde{c} \rangle$. А поскольку сама пара $\langle \tilde{b}, \tilde{c} \rangle$ случайна, то случайна и тройка $\langle \tilde{b}, \tilde{c}, \tilde{d} \rangle$. Единственное, что нам нужно, но не следует из сказанного, это случайность двух оставшихся троек $\langle \tilde{a}, \tilde{b}, \tilde{d} \rangle$ и $\langle \tilde{a}, \tilde{c}, \tilde{d} \rangle$. Как добиться этого, мы обсудим позднее.

Пусть S обозначает множество всех слов длины n, и пусть M некоторое множество всюду определенных функций из S в S. Скажем, что множество M *обслуживает* четвёрку $\langle a,b,c,d\rangle \in S^4$, если $f(a)=c$ и $g(b)=d$ для некоторой пары $\langle f,g\rangle \in M$. Упомянутое свойство множества Q состоит в следующем: *любое множество функций M из S в S мощности менее 2^k обслуживает менее $\varepsilon|Q|$ четвёрок из Q*. Числа k,ε мы определим позднее. Число k будет чуть меньше $2n$, а число ε будет иметь порядок $O(1/n)$. Это свойство гарантирует, что для большинства четвёрок $\langle a,b,c,d\rangle$ в Q сложность задачи $(a\to c)\wedge(b\to d)$ больше k. Действительно, решениями задачи $(a\to c)\wedge(b\to d)$ являются пары программ $\langle p,q\rangle$, для которых $[p](a)=c$, $[q](b)=d$. Для каждой пары программ $\langle p,q\rangle$ сложности меньше k доопределим любым способом функции $a\mapsto[p](a)$ и $b\mapsto[q](b)$ на всё множество S. Рассмотрим получившееся множество M пар функций. Его мощность меньше 2^k. Следовательно, M обслуживает не более, чем $\varepsilon|Q|$ четвёрок из Q. А по построению M обслуживает все четвёрки $\langle a,b,c,d\rangle$, для которых сложность задачи $(a\to c)\wedge(b\to d)$ меньше k.

Теперь подберём такие $\varepsilon=O(1/n)$ и $k=n-O(\log n)$, что наше свойство непусто. Для этого оценим вероятность, что случайное равномерное множество Q не обладает свойством. Случайный выбор равномерного множества Q заключается в следующем: для каждой тройки $\langle a,b,c\rangle$ мы независимо выбираем с равной вероятностью некоторое слово d длины n и помещаем получившуюся четвёрку в Q.

Сначала фиксируем множество M из 2^k функций и оценим сверху вероятность того, что оно обслуживает более, чем $\varepsilon|Q|$ четвёрок. Для этого назовём тройку $\langle a,b,c\rangle$ *плохой*, если при случайном выборе пары $\langle f,g\rangle$ из M вероятность события $f(a)=c$ больше $n\cdot 2^{-n}$ (это свойство не зависит от b). В противном случае назовём тройку *хорошей*. При случайном выборе тройки $\langle a,b,c\rangle$ среднее значение вероятности события $f(a)=c$ (вероятность здесь берется по случайному выбору пары функций из M) равно 2^{-n}. По неравенству Чебышёва доля троек, для которых эта вероятность больше $n\cdot 2^{-n}$, меньше $1/n$. Следовательно, плохих троек менее $(1/n)2^{3^n}$, а хороших более $(1-1/n)2^{3n}$.

Для любой хорошей тройки $\langle a,b,c\rangle$ вероятность того, что для случайно выбранного d множество M обслуживает четвёрку $\langle a,b,c,d\rangle$ не превосходит $1/n$. Действительно, если M обслуживает четвёрку $\langle a,b,c,d\rangle$, то d принадлежит множеству, состоящему из всех таких $g(b)$, что для некоторой функции f пара $\langle f,g\rangle$ принадлежит M и $f(a)=c$. Это множество содержит не более $n2^{-n}\cdot|M|=2^{k-n+\log n}$ элементов. Следовательно, если положить $k=2n-2\log n$, это множество содержит менее, чем $(1/n)$-ую часть всех слов длины n.

Теперь воспользуемся оценкой Чернова: если проводить N независимых испытаний, вероятность успеха в каждом из которых равна p, то для любого $0<\varepsilon\leqslant p(1-p)$ доля успешных испытаний будет меньше $p+\varepsilon$ с вероятностью не меньше $1-2^{-\varepsilon^2N/(2p)}$. Мы будем применять следующее его следствие, получаемое при $\varepsilon=p/2$: если вероятность успеха в отдельном испытании не превосходит p, то вероятность того, что доля успешных испытаний больше $3p/2$, меньше $2^{-pN/8}$. В нашем случае испытания нумеруются хорошими тройками $\langle a,b,c\rangle$, так что $2^{3n-1}\leqslant N\leqslant 2^{3n}$. Испытание с номером $\langle a,b,c\rangle$ заключается в случайном выборе d, а успехом является то, что четвёрка $\langle a,b,c,d\rangle$ обслуживается множеством M. Вероятность успеха не больше $p=1/n$. Поэтому при случайном выборе множества Q с вероятностью не менее $1-2^{-\Omega(N/n)}$ для не более чем $3/(2n)$-ой части хороших троек соответствующая четвёрка из Q обслужена M. Вспомнив еще, что в Q может быть не более $|Q|/n$ четвёрок, с плохой тройкой $\langle a, b, c\rangle$ мы получим, что с вероятностью не более $2^{-\Omega(N/n)}$ доля обслуженных четвёрок в Q больше $5/(2n)$.

Осталось убедиться, что эта вероятность даже при умножении на количество различных множеств M остается малой. Количество разных M не превосходит квадрата количества разных функций, отображающих слова длины n в слова длины n (оно равно 2^{n2^n}), возведенному в степень 2^k (такова мощность M). Получается формула

$$2^{2n2^n2^k} = 2^{2^{n+k+\log n+1}}.$$

Сравним эту формулу с формулой для вероятности того, что доля обслуженных четвёрок больше $5/(2n)$. При перемножении этих формул показатели экспонент первого этажа складываются и нам нужно чтобы отрицательное слагаемое $-\Omega(N/n)$ было больше положительного $2^{n+k+\log n+1}$ по абсолютной величине. Поскольку $N \geqslant 2^{3n-1}$, для этого достаточно положить $k = n - 2\log n - c$, где c достаточно большая константа.

Осталось объяснить, какими свойствами Q будет обеспечена случайность двух оставшихся троек $\langle\tilde a, \tilde b, \tilde d\rangle$, $\langle\tilde a, \tilde c, \tilde d\rangle$ (для любой случайной четвёрки $\langle\tilde a, \tilde b, \tilde c, \tilde d\rangle$ из Q). Самое простое — это действовать так же, как в доказательстве случайности тройки $\langle\tilde b, \tilde c, \tilde d\rangle$. Для этого нам нужно, чтобы для большинства четвёрок из Q сложной была не только задача $(a \to c)\wedge(b \to d)$, но также и симметричные ей задачи $(c \to b)\wedge(a \to d)$ и $(b \to a)\wedge(c \to d)$. А для этого надо усилить свойство, с помощью которого определялось Q, потребовав, чтобы были выполнены еще и два симметричных условия: при любом M количество четвёрок из Q, для которых $f(c) = b$, $g(a) = d$ для некоторой пары $\langle f, g\rangle$ из M не должно превосходить $\varepsilon|Q|$, и аналогичное условие должно быть выполнено условие для пары равенств $f(b) = a$, $g(c) = d$. Проведя симметричные рассуждения, можно убедиться, что при случайном выборе равномерного множества, вероятность нарушения каждого из двух новых условий также ничтожно мала. Следовательно и вероятность нарушения хотя бы одного из них ничтожно мала.

Для завершения второго доказательства теоремы осталось предъявить другую четвёрку с теми же сложностями, что и у построенной, но для которой сложность задачи $K((a \to c) \wedge (b \to d))$ существенно меньше, чем $2n$. Для этого возьмём случайное слово длины $3n$, разрежем его на три части длины n. Это будут a, b и c; затем положим $d = a \oplus b \oplus c$. Зная $a \oplus c$, мы можем по a найти c, а по b найти d, следовательно $K((a \to c) \wedge (b \to d)) = n$ (с точностью до $O(\log n)$). $\triangleright$

263 Докажите, что сложность задачи $(p \vee q) \to (r \vee s)$ также не выражается через сложности p, q, r, s, их пар, троек, и четвёрки. [Можно положить $p = a$, $q = b$, $r = ac$, $s = bd$, где a, b, c, d одна из двух четвёрок, использованных в доказательстве теоремы (скажем, в первом). Сложность полученной задачи зависит от того, какую из двух четвёрок взять, а сложности p, q, r, s, их пар, троек, и четвёрки не зависят от этого.]

Полезно сравнить вероятностное доказательство с геометрическим. Геометрическое доказательство более конструктивно — в нем первая четвёрка задана более явно, чем в вероятностном доказательстве. Зато в вероятностном доказательстве для первой четвёрки сложность задачи $(a \to c) \wedge (b \to d)$ равна верхним оценкам из задачи 261, которые все равны $2n$. Для четвёрок из геометрического доказательства эти верхние оценки также все равны $2n$, однако для первой четвёрки мы получили лишь нижнюю оценку $1.5n$ для сложности задачи $(a \to c) \wedge (b \to d)$. Можно ли получить лучшую оценку, неизвестно.

14. Алгоритмические свойства

Полнота и неполнота функции сложности (Мучник и предшественники) частоты в вычислимых последовательностях и $0'$-полумера?

15. Сложности, меры, философия

случайность по нескольким мерам, дефект бернуллиевости

штраф по Вовку

последовательность, случайная по нескольким мерам, даёт одинаковые предсказания

стохастические объекты (Вьюгин, Гач) последовательности, случайные по вычислимым мерам (Звонкин-Левин? связь со стохастичностью?)

16. Алгоритмическая статистика

[statist]

16.1. Постановка задачи. Дефект случайности.

[finite-random-deficiency]

Общими словами задачу математической статистики можно описать так: имеются результаты наблюдений, нужно предложить правдоподобную вероятностную гипотезу, их объясняющую (согласованную с ними).

Пусть, скажем, имеется некоторое устройство («чёрный ящик»). Мы включили его, и оно выдало последовательность из миллиона битов (=число от 0 до $2^{1\,000\,000} - 1$). Что можно сказать о внутреннем строении «чёрного ящика», зная эту последовательность?

Классическая математическая статистика, увы, про это ничего не говорит. Вот если бы мы имели данные от большого числа одинаковых независимых устройств, или могли включать наше устройство много раз подряд (и предположить, что результаты последовательных включений независимы), тогда другое дело. (Заметим в скобках, что ситуации однократного и невоспроизводимого эксперимента не так уж и редки на практике.) Или, скажем, если бы мы заранее имели некоторое параметрическое семейство гипотез, — тогда можно было бы выбирать максимально правдоподобное значение параметра. А так ничего не скажешь — ведь все $2^{1\,000\,000}$ возможных исходов с точки зрения математической статистики совершенно равноправны (на множестве возможных исходов нет никакой структуры).

Тем не менее здравый смысл иногда позволяет выдвинуть кажущиеся разумными предположения. Например, если наше устройство выдало миллион нулей, скорее всего оно только и умеет что выдавать нули. А если оно выдало последовательность нулей и единиц безо всяких закономерностей, то похоже, что это результат миллиона независимых бросаний монеты. И хорошо бы как-нибудь уточнить механизм таких догадок.

В первом примере (слово из одних нулей) мы предлагали в качестве объяснения гипотезу, согласно которой только это слово и могло появиться. Формально говоря, наша гипотеза указывала в качестве множества возможных результатов эксперимента одноэлементное множество, состоящее из слова $00\ldots0$. Примерно то же естественно сказать и для любого двоичного слова x малой сложности: можно предположить, что чёрный ящик как раз и предназначен для порождения такого слова (имеет множество возможностей $\{x\}$).

Второй пример (слово без закономерностей, у которого сложность примерно равна длине) указывает другой крайний случай, когда мы говорим, что чёрный ящик мог выдать любое слово, то есть указываем в качестве множества возможностей множество всех слов длины $1\,000\,000$.

Можно привести и промежуточный пример. Пусть оказалось, что первые $500\,000$ битов нашего слова равны нулю, а дальше идут $500\,000$ «случайных битов» (то есть некоторое слово длины и сложности $500\,000$). Видимо, разумная гипотеза в этом случае такова: чёрный ящик устроен так, что он сначала выдаёт $500\,000$ нулей, а потом уже $500\,000$ случайно выбранных символов. Соответствующее множество возможностей состоит из всех двоичных слов длины $1\,000\,000$, у которых первая половина состоит только из нулей (таких слов, очевидно, $2^{500\,000}$).

Общая схема всех наших примеров такова. Нам дают некоторое слово x длины n. Мы предлагаем в качестве «объяснения» этого слова некоторое множество A, содержащее x. При этом мы хотим, чтобы:

- множество A было простым (колмогоровская сложность A была бы мала);
- слово x было бы «типичным представителем» множества A.

Как уточнить эти пожелания? Чтобы говорить о колмогоровской сложности конечного множества, запишем элементы этого множества в список (заранее договорившись о порядке — скажем, в лексикографическом порядке), этот список закодируем двоичным словом и возьмём сложность этого слова. (Разные естественные способы кодирования дадут варианты сложности, отличающиеся не более чем на константу, поскольку от одного из них можно алгоритмически переходить к другому).

Понятие «типичного представителя» также можно уточнить в терминах колмогоровской сложности. Вспомним, что если множество A содержит N элементов, то условная сложность $\mathit{KS}(x|A)$ любого его элемента при известном A не превосходит $\log_2 N + O(1)$, поскольку каждый элемент может быть задан своим порядковым номером в множестве A. Большинство элементов множества A имеют условную сложность, близкую к $\log_2 N$ (поскольку элементов меньшей сложности ма́ло по сравнению с N), и типичными представителями множества A можно назвать как раз входящих в это большинство.

Другими словами это можно объяснить так. Назовём (неотрицательную с точностью до $O(1)$) величину
$$d(x|A) = \log_2 |A| - \mathit{KS}(x|A)$$
дефектом случайности слова x в множестве A (мы будем использовать это равенство лишь в случае, когда $x \in A$, хотя правая часть имеет смысл всегда; при $x \notin A$ естественно считать, что $d(x|A) = +\infty$, поскольку в этом случае гипотеза A «максимально непригодна» для объяснения слова x).

«Типичными представителями» множества A мы считаем те $x \in A$, для которых величина $d(x|A)$ мала.

264 Докажите, что вероятность того, что случайно взятый элемент данного множества A имеет дефект больше k, не превосходит 2^{-k}.

(Вероятность понимается здесь просто как доля элементов с большим дефектом в множестве A. Чтобы это утверждение было формально верным, надо брать целую часть сверху от $\log_2 |A|$ в определении дефекта, но поскольку сложность по существу определена с точностью до константы, этим можно пренебречь.)

Заметим ещё, что функция d (как функция двух аргументов) перечислима снизу (мы можем постепенно получать всё более точные нижние оценки для неё, хотя и не можем сказать, когда мы дошли до предела). Это непосредственно следует из того, что функция KS перечислима сверху.

265 Пусть $\delta(x|A)$ определено для любого элемента x любого конечного множества A, причём: (а) функция δ перечислима снизу; (б) для любого множества A и любого числа k доля элементов A, для которых $\delta(x|A) > k$, не превосходит 2^{-k}. Тогда $\delta(x|A) \leqslant d(x|A) + O(1)$.

Смысл этой задачи (которая является простым следствием аналогичного утверждения для условной сложности, см. теорему 19 на с. 41) можно объяснить так. Могут быть разные мнения о том, какие элементы в каких множествах «нетипичны» (разные способы измерения дефекта). Если мы все их отнормируем по доле нетипичных элементов, потребовав, чтобы доля k-нетипичных элементов не превосходила 2^{-k}, а также будем требовать, чтобы нетипичность можно было рано или поздно алгоритмически обнаруживать, то среди всех способов измерения дефекта существует наилучший (с точностью до константы он обнаруживает не меньше дефектов, чем любой другой).

266 Покажите, что *префиксный дефект случайности* элемента x конечного множества A, определяемый как $d_P(x|A) = \log_2 |A| - \mathit{KP}(x|A)$, является максимальной перечислимой снизу функцией δ двух аргументов (x и A), для которой среднее значение $(1/|A|)\sum_{x\in A} 2^{\delta(x|A}$ не превосходит единицы для любого конечного множества A. [Указание: воспользуйтесь описанием условной префиксной сложности как логарифма априорной вероятности.]

Таким образом, задачу поиска хорошего объяснения для данного слова x можно сформулировать так: надо найти *простое множество A, для которого дефект $d(x|A)$ мал.*

Всегда ли (для любого ли x) это возможно? Этот вопрос мы рассмотрим в следующем разделе.

Отметим ещё, что знатоки теории вероятностей могут законно упрекнуть нас в том, что мы рассматриваем лишь равномерные распределения в качестве вероятностных гипотез. Вместо этого можно было бы говорить о произвольных распределениях вероятностей (для простоты — с конечным носителем и рациональными значениями вероятностей) и определять дефект слова x относительно распределения P как $-\log_2 P(x) - \mathit{KS}(x|P)$ (считая его бесконечным для тех x, для которых $P(x) = 0$: для таких x объяснение P совсем непригодно).

Для равномерных распределений (все элементы конечного множества имеют вероятность $1/|A|$) это определение совпадает с прежним. Можно заметить, что и в общем случае отличие не слишком велико, как показывает следующая задача:

267 `[measure-stochastic]` Пусть для слова x сложности n найдено вероятностное распределение сложности не больше k, при котором дефект x не больше l. Тогда существует множество A сложности не больше $k + O(\log(l + n))$, содержащее x, относительно которого дефект x не превосходит $l + O(\log(l + n))$. [Указание. Пусть p — вероятность слова x относительно распределения P, округлённая вниз до ближайшей степени двойки. Положим $A = \{y \mid P(y) \geqslant p\}$.]

Поэтому в дальнейшем мы в основном ограничиваемся конечными множествами (и равномерными распределениями вероятностей на них) в качестве описаний. При этом важно подчеркнуть, что говоря о сложности конечного множества, мы имеем в виду его сложность как конструктивного объекта (списка элементов), а не, скажем, сложность его перечисления (сложность программы, порождающей все его элементы). Если допустить последнее, то определение стохастичности станет неинтересным: любой объект x сложности n содержится в множестве всех объектов сложности не больше n, которое имеет $O(2^n)$ элементов и которое можно порождать программой сложности $O(\log n)$, но вряд ли стоит считать S_n хорошим «объяснением» для x.

Аналогичным образом (в определении с мерами) рассматривать сложность меры как конечного объекта (если её значения рациональны) или даже сложность мер с вычислимыми значениями, имея в виду сложность соответствующей программы, но нельзя рассматривать перечислимые снизу меры и сложность программ, приближающих их снизу — ибо в этом случае в качестве простого объяснения для произвольного слова можно будет взять максимальную перечислимую снизу полумеру.

16.2. Стохастические объекты

[statist-stochastic] Будем говорить, что слово x является (α, β)-стохастическим, если для него существует «хорошее объяснение»: существует конечное множество A, содержащее слово x, для которого $\mathit{KS}(A) \leqslant \alpha$ и $d(x|A) \leqslant \beta$.

Возникает такой естественный вопрос: будем рассматривать все слова длины n и положим α и β равными $O(\log n)$ или $o(n)$ (тем самым сложность гипотезы, как и полагается, мала по сравнению с длиной объясняемых слов). Будут ли при таких α и β существовать нестохастические («необъяснимые») объекты? Оказывается, что да.

Theorem 220. [nonstochastic-existence] *При $2\alpha + \beta < n - O(\log n)$ существуют слова длины n, не являющиеся (α, β)-стохастическими.*

(Точнее: существует такая константа c, что для всех достаточно больши́х n и для всех α и β с $2\alpha + \beta < n - c\log n$ существуют слова длины n, не являющиеся (α, β)-стохастическими.)

$\triangleleft$ Рассмотрим список всех конечных множеств сложности не больше α. Сложность такого списка не больше $\alpha + O(\log \alpha) = \alpha + O(\log n)$ (см. с. 32). Для краткости мы будем игнорировать члены порядка $O(\log n)$, надеясь, что читатель к этому уже привык и легко исправит рассуждение; в частности, мы будем считать, что список имеет сложность α.

Отберём в этом списке множества, имеющие не более $2^{\alpha+\beta}$ элементов. Результирующий список также имеет сложность α и содержит не более 2^{α} множеств размера $2^{\alpha+\beta}$. Всего в них $2^{2\alpha+\beta} < 2^n$ элементов, и поэтому есть слова длины n, не попадающие в эти множества. Рассмотрим первое такое слово (скажем, в алфавитном порядке). Оно имеет сложность α, поскольку для его задания (помимо числа n) достаточно знать этот список.

Покажем, что это слово (обозначим его t) не может быть (α, β)-стохастическим. Если t содержится в некотором множестве A сложности не больше α, то размер этого множества больше $2^{\alpha+\beta}$, поскольку все меньшие множества мы исключили по построению. Тогда $d(t|A) = \log|A| - \mathit{KS}(t|A) \geqslant (\alpha + \beta) - \mathit{KS}(t) \geqslant (\alpha + \beta) - \alpha \geqslant \beta$.

(Во всех этих построениях нужно, конечно, оставить запас размера $c\log n$ при достаточном c, чтобы компенсировать добавки порядка $O(\log n)$.) $\triangleright$

С другой стороны имеется следующая очевидная оценка:

Theorem 221. [nonstochastic-nonexistence] *Если $\alpha + \beta > n + O(\log n)$, то все слова длины n являются (α, β)-стохастическими.*

$\triangleleft$ Достаточно разбить все слова длины n на 2^{α} множеств размера 2^{β}. $\triangleright$

Как мы увидим, реальная ситуация ближе к этой нижней оценке, чем к предыдущей верхней (см. задачу 282 на с. 404).[future-stochastic-improvement]

Естественно поинтересоваться, насколько часто встречаются нестохастические объекты. Например, можно спросить, какую долю они составляют среди слов длины n. Заранее ясно, что эта доля не превосходит $2^{-\beta}$, поскольку в множестве A, состоящем из всех слов длины n, слова с дефектом β составляют долю не более $2^{-\beta}$.

С другой стороны, если $2\alpha+\beta$ много меньше n, можно продолжить доказательство теоремы 220, используя имеющийся в нём запас. А именно, для некоторого h можно взять все множества сложности не больше α с числом элементов не больше $2^{\alpha+\beta+h}$ и затем взять первые 2^h элементов, не покрытых этими множествами; такие элементы найдутся, если $2\alpha+\beta+h<n$. Сложность этих элементов будет не больше $\alpha+h$ и потому их дефект в любом множестве из более чем $2^{\alpha+\beta+h}$ элементов будет больше β. Проведя эти рассуждения аккуратно (не забывая о членах порядка $O(\log n)$), получаем такой результат:

Theorem 222. *Среди слов длины n доля не-(α,β)-стохастических слов не меньше $2^{-2\alpha-\beta-O(\log n)}$.*

Можно также интересоваться не просто долей нестохастических слов (другими словами, вероятностью получения такого слова при бросании монеты), а их суммарной априорной вероятностью (вероятностью их получения на выходе оптимальной вероятностной машины). Формально говоря, пусть $m(x)$ — априорная вероятность слова x (в смысле главы 4, равная $2^{-\mathit{KP}(x)+O(1)}$). Рассмотрим сумму $m(x)$ по всем словам длины n, не являющимся (α,β)-стохастическими.

Theorem 223. *Если $2\alpha+\beta<n-O(\log n)$ и $\alpha<\beta-O(\log n)$, то указанная сумма есть $2^{-\alpha+O(\log n)}$.*

$\triangleleft$ Нам нужно указать верхнюю и нижнюю оценку для этой суммы. Нижняя оценка получается из доказательства теоремы 220; в самом деле, построенное там нестохастическое слово имеет сложность α и потому уже для него одного априорная вероятность равна $2^{-\alpha}$ (как всегда, мы опускаем слагаемые $O(\log n)$).

Для получения верхней оценки рассмотрим сумму $m(x)$ по *всем* словам длины n. Получится некоторое действительное число ω, не превосходящее единицы. Пусть w — его приближение снизу с точностью $2^{-\alpha}$ (первые α битов оставляем, остальное выбрасываем). Все слагаемые в сумме $\omega=\sum\{m(x)\mid l(x)=n\}$ перечислимы снизу; будем строить их рациональные приближения и остановимся, как только суммы этих приближений превысят w.

В результате мы получим некоторую меру P с рациональными значениями на множестве всех слов длины n, имеющую сложность α (поскольку она вычислимо строится по слову w длины α). Мера P отличается от априорной вероятности не более чем на $2^{-\alpha}$ (суммарно по всем словам).

Задача 267 позволяет использовать в определении стохастичности произвольные меры (а не только равномерные) с погрешностью $O(\log n)$ в параметрах. Остаётся показать, что суммарная априорная вероятность всех слов, для которых дефект относительно P больше β, не превосходит $2^{-\alpha}$. В самом деле, для таких слов x мы имеем

$$-\log P(x)-\mathit{KS}(x|P)>\beta.$$

Сложность P не больше α и потому $\mathit{KS}(x)$ превосходит $\mathit{KS}(x|P)$ не более чем на α, откуда

$$-\log P(x) - \mathit{KS}(x) > \beta - \alpha.$$

Поскольку члены порядка $O(\log n)$ мы игнорируем, можно заменить обычную сложность на префиксную:

$$-\log P(x) - \mathit{KP}(x) > \beta - \alpha.$$

Вспоминая, что префиксная сложность связана с априорной вероятностью, получаем, что

$$\log(m(x)/P(x)) > \beta - \alpha.$$

для любого слова x с дефектом (относительно P) больше β. По условию $\alpha < \beta$ с запасом, достаточным для компенсации всех допущенных нами ошибок, так что можно считать, что для всех таких слов выполнено неравенство $P(x) < m(x)/2$, или $(m(x) - P(x)) > m(x)/2$. Вспоминая, что сумма $m(x) - P(x)$ по всем вообще x не превосходит $2^{-\alpha}$ по построению меры P, заключаем, что и сумма всех $m(x)$ по словам дефекта больше β относительно P не превосходит $2^{-\alpha+O(\log n)}$, что и требовалось. $\triangleright$

Понятие стохастического объекта можно рассматривать как конечный аналог понятия последовательности, случайной по некоторой вычислимой мере. Тут есть не только аналогия, но и прямая связь, как показывает следующая задача.

268 `[infinite-finite-stochastic]` Покажите, что если последовательность ω случайна в смысле Мартин-Лёфа относительно некоторой вычислимой меры, то её начальные отрезки длины n являются $(O(\log n), O(\log n))$-стохастическими словами. [Указание: можно воспользоваться определением стохастичности с мерами.]

Выведите отсюда, что существуют последовательности, не случайные в смысле Мартин-Лёфа ни по какой вычислимой мере. [Указание: добавление начального отрезка малой длины не сильно нарушает стохастичность.]

16.3. Двухчастные описания

`[two-part-descriptions]`

Можно оценивать качество статистических гипотез и в немного другой системе координат. Начнём с такого замечания. Если слово x является элементом конечного множества A, то можно задать это слово, указав:

- множество A;
- порядковый номер слова x в этом множестве (относительно какого-либо естественного порядка на словах, скажем, алфавитного).

Из этого следует, что $\mathit{KS}(x) \leqslant K(A) + \log|A|$ для любого элемента x любого конечного множества A (опять же с логарифмической точностью).

Одно и то же слово x может иметь много таких «двухчастных» описаний. Какие из них лучше, какие хуже? Ясно, что вообще-то мы хотели бы сделать обе части описания как можно меньше: если удалось найти более простое множество того же размера или найти меньшее множество той же сложности (среди множеств, содержащих x), то это хорошо. Но как сравнивать простое, но большое множество с меньшим, но более сложным? Одна из возможностей — сравнить длины соответствующих двухчастных описаний и предпочесть более короткое. (По-английски этот принцип сравнения называют Minimal Description Length principle, или MDL.)

Следующее простое наблюдение показывает, что мы можем перераспределять информацию между частями двухчастного описания, перемещая её из второй части в первую (уменьшая размер множества за счёт увеличения сложности).

Theorem 224. [description-shift] *Пусть конечное множество A содержит слово x и $k < \log|A|$. Тогда можно найти конечное множество A', содержащее x, для которого $|A'| \leqslant |A|/2^k$ и $KS(A') \leqslant KS(A) + k + O(\log k)$.*

$\triangleleft$ Запишем элементы A в каком-то фиксированном порядке (например, алфавитном) и разобьём их на 2^k частей (первые $|A|/2^k$ элементов, следующие и т.д.; мы опускаем очевидные оговорки на случай, когда нацело не делится). В качестве A' возьмём часть, в которую попало слово x. Чтобы задать её, достаточно указать A, число k и номер части, на который нужно не более k битов. $\triangleright$

Это утверждение удобно иллюстрировать картинкой. Для данного слова x рассмотрим множество P_x всех пар $\langle k, l\rangle$, при которых существует конечное множество $A \ni x$ с $KS(A) \leqslant k$ и $\log|A| \leqslant l$. Из определения ясно, что вместе с каждой точкой в это множество входят все точки справа от неё (с бóльшими k) и все точки сверху от неё (с бóльшими l). Только что доказанная теорема показывает, что можно сдвигаться вправо-вниз на вектор $\langle k, -k\rangle$ (с логарифмической точностью).

Как мы увидим, обратное движение возможно далеко не всегда, так что среди двухчастных описаний одной и той же (суммарной) длины наиболее ценным является описание с более простым множеством большего размера (из него можно получить остальные, но не наоборот).

Схематически строение множества P_x можно описать так (все утверждения с логарифмической точностью). В него входит точка $\langle 0, l(x)\rangle$, соответствующая множеству A всех слов той же длины, что и x. Кроме того, в него входит точка $\langle 0, KS(x)\rangle$, соответствующая $A = \{x\}$. Множество P_x ограничено кривой, идущей от первой точки до второй, причём эта кривая не заходит в треугольник $k + l \leqslant KS(x)$ (пунктирная линия) и всюду идёт вниз под углом 45° или более, согласно теореме 224 (рис. 47).

При взгляде на рисунок возникает естественный вопрос: какие граничные кривые возможны? Может ли, например, граница области P_x пройти по пунктирной линии? Может, для этого достаточно в качестве x взять случайное слово, к которому приписать недостающие (до желательной длины) нули; тогда в качестве A можно взять множество всех слов, на конце которых стоят эти самые нули. Не столь очевиден ответ на другой вопрос: может ли кривая идти под углом 45° до самого конца, где круто спускаться из точки $\langle KS(x), l(x) - KS(x)\rangle$ к оси абсцисс (в точке $\langle KS(x), 0\rangle$). Неформально говоря, это означает, что слово x допускает по существу лишь два вида статистических гипотез: множество всех слов длины $l(x)$ (и его части, получаемые по теореме 224), и одноэлементное множество $\{x\}$.

$\boxed{\textbf{269}}$ Покажите, что в этом случае слово x не будет (α, β)-стохастическим при α, β, малых по сравнению с k и $n-k$, поскольку для стохастических слов граница множества P_x близка к пунктирной линии.

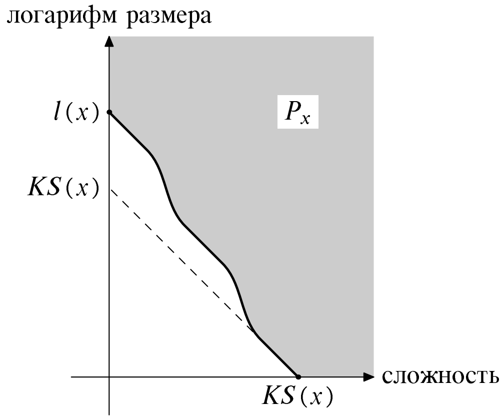

Рис. 47. Множество P_x.

[mdl.1.eps]

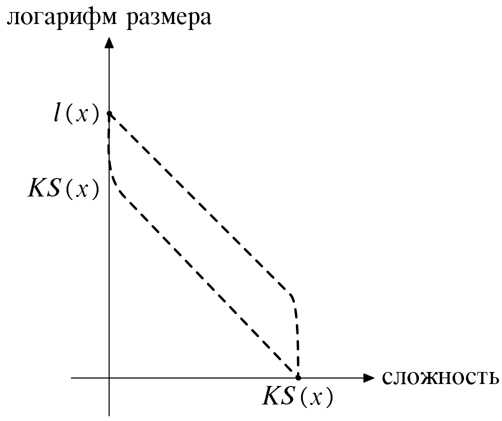

Рис. 48. Две граничные кривые.

[mdl.2.eps]

Оказывается, что возможны не только два крайних случая (они изображены на рис. 48), но и все промежуточные: любая (достаточно простая) кривая с указанными выше свойствами может быть границей множества P_x при некотором x (с логарифмической точностью). Более точно, имеет место следующий результат:

Theorem 225. [stat-any-curve] *Пусть* $k \leqslant n$ *и пусть* $n = t_0 > t_1 > \ldots > t_k = 0$ — *последовательность натуральных чисел, имеющая сложность* m. *Тогда существует слово* x *сложности* $k + O(\log n) + O(m)$ *и длины* $n + O(\log n) + O(m)$, *для которого множество* P_x *отстоит от множества* $T = \{\langle i, j\rangle \mid (i < k) \Rightarrow (j > t_i)\}$ *не более чем на* $O(\log n) + O(m)$.

(Говорят, что множество P отстоит от множества Q не более чем на ε, если P содержится в ε-окрестности Q и наоборот.)

$\triangleleft$ Мы уже говорили, что из этой теоремы следует существование нестохастических слов. Поэтому не удивительно, что её доказательство использует тот же приём, что и доказательство теоремы 220.

Для каждого i от 0 до k составим список всех множеств сложности не больше i и размером не больше 2^{t_i}; объединив все такие множества (при данном i), получим некоторое множество S_i, состоящее не более чем из 2^{i+t_i} элементов. (Здесь и далее при оценке числа элементов мы опускаем постоянные и полиномиальные по n множители, поскольку они соответствуют слагаемым $O(\log n)$ в оценках длин и сложностей.) Поскольку t_i убывают (что соответствует наклону 45° на рисунке), то $i + t_i$ не возрастают с ростом i, и потому любое из S_i содержит не более $2^{t_0} = 2^n$ элементов. Объединение всех S_i также содержит не более чем 2^n элементов (с точностью до полиномиального множителя), и потому среди слов длины n (точнее, $n + O(\log n)$) есть слово, не входящее ни в одно из множеств S_i. Первое (в каком-либо, например, алфавитном порядке) такое слово и будет искомым словом x.

По построению P_x лежит выше кривой t_i (содержится в множестве T). Остаётся оценить сложность слова x и показать, что множество P_x прилегает к кривой (то есть что T содержится в окрестности P_x).

Верхняя оценка сложности слова x следует из того, что список всех объектов сложности не больше k с указанием их сложности имеет сложность $k + O(\log k)$ (достаточно знать k и самую долгоиграющую программу длины не больше k, см. с. 32); помимо этого списка нам надо знать последовательность $t_0, t_1, \ldots, t_k$, имеющую сложность m.

Нижняя оценка: меньше k сложность слова x быть не может, так как все одноэлементные множества такой сложности выброшены.

Остаётся показать, что при любом $i \leqslant k$ слово x можно поместить в множество A сложности i (или чуть больше) и размера 2^{t_i} (или чуть больше). Искомое множество A строится «с нескольких попыток».

Вначале возьмём первые 2^{t_i} слов нужной длины в качестве A и параллельно развернём процесс порождения множеств сложности не больше j и размера не больше 2^{t_j} при всех $j = 0, \ldots, k$. Для каждого j получится список множеств, который мы будем называть j-списком. В ходе этого процесса будут обнаруживаться подлежащие выбрасыванию слова — элементы всех множеств во всех списках; появление нового множества приводит к одномоментному выбрасыванию всех входящих в него слов. Пока не все слова из A оказались выброшенными (пока объединение порождённых множеств не покроет целиком множество A), мы можем не беспокоиться, поскольку первое по порядку невыброшенное слово принадлежит A. Как только все слова из A будут выброшены, придётся заменить A, взяв в качестве нового кандидата первые 2^{t_i} невыброшенных (на данный момент) слов нужной длины. После этого мы спокойно ждём до тех пор, пока все слова нового множества A не будут выброшены (не попадут в объединение S_i); когда и если это случится, мы заменим A, взяв первые 2^{t_i} пока ещё не выброшенных слов, и так далее.

Рано или поздно этот процесс выбрасывания остановится, и текущее значение множества A на этот момент будет содержать наше слово x. Таким образом, мы нашли множество правильного размера, содержащее x, вопрос только в сложности этого множества.

Чтобы описать процесс порождения выбрасываемых множеств, достаточно знать последовательность $t_0, \ldots, t_k$ (и длину слов, равную примерно n), то есть всего $m + O(\log n)$ битов информации. Поэтому с интересующей нас точностью сложность каждого из кандидатов на роль множества A оценивается логарифмом его порядкового номера (сколько кандидатов перед ним пришлось отбросить). Таким образом, нам осталось доказать, что число отброшенных кандидатов ненамного больше 2^i; тогда сложность любого из кандидатов (в том числе и последнего, правильного) будет ненамного больше i.

Итак, сколько раз мы могли заменять множество A в связи с появлением новых множеств в списках? Отдельно посмотрим на появление множеств небольшой сложности и большого размера, а именно, входящих в j-списки с $j \leqslant i$. Такое событие могло случиться не более $O(2^i)$ раз, так как в j-списке не более $O(2^j)$ множеств. Поэтому мы можем считать лишь интервалы между двумя соседними изменениями множества A, в течение которых таких событий не случалось. Это значит, что все элементы предыдущего A были покрыты множествами из j-списков при $j > i$. Во всех этих множествах не так много элементов: общее количество слов во всех множествах j-списка не больше $2^j \cdot 2^{t_j}$; поскольку $t_j + j \leqslant t_i + i$ и количество разных значений j невелико, получается всего примерно $O(2^{t_i+i})$ элементов, и после деления на $|A| = 2^{t_i}$ остаётся 2^i интервалов. $\triangleright$

Эта теорема показывает, что, помимо сложности $\mathit{KS}(x)$, слово x может иметь другие «сложностные характеристики»; такой характеристикой можно считать границу множества P_x (или, если угодно, само множество). Эта характеристика, если можно так выразиться, «бесконечномерна» (представляет собой не одно число и не набор чисел фиксированной длины, а целую кривую).

Классификацию слов по сложности можно представлять себе как вложенную цепочку множеств $S_0 \subset S_1 \subset S_2 \subset \ldots$, где S_i — множество слов сложности меньше i. В этой цепочке множества S_i перечислимы равномерно по i и содержат $O(2^i)$ элементов.

Теперь вместо этой классификации у нас есть двумерная классификация множеств $S_{i,j}$, где множество $S_{i,j}$ состоит из всех слов x, входящих в конечные множества A с $\mathit{KS}(A) < i$ и $\log|A| < j$. (Такое множество A мы будем называть $(i * j)$-*описанием* любого его элемента x.) Мы получаем двумерную таблицу из множеств $S_{i,j}$, монотонную по обоим направлениям ($S_{i,j}$ растёт с ростом i и с ростом j). Кроме того, теорема 224 показывает, что эта таблица монотонна по диагонали: $S_{i,j} \subset S_{i+k,j-k}$. (Как всегда, мы пренебрегаем логарифмическими поправками: точнее следовало бы написать $S_{i,j} \subset S_{i+k+O(\log k),j-k}$.)

Чтобы лучше понять смысл этой двумерной «стратификации», полезно посмотреть на эквивалентные определения множеств $S_{i,j}$. При этом мы будем, как обычно, пренебрегать логарифмическими слагаемыми и считать два варианта определения (обозначаемые S и S') эквивалентными, если $S_{i,j} \subset S'_{i+O(\log n),j+O(\log n)}$ и наоборот (здесь и далее $n = i + j$).

Говоря о «перечне» в формулировке следующей теоремы, мы подразумеваем алгоритм, который работает и время от времени выдаёт на выход некоторые двоичные слова (возможно, с повторениями); длиной перечня мы будем называть общее число выданных слов (с учётом кратностей). В условии (в) мы считаем, что алгоритм может выдавать на выход слова не поодиночке, а «порциями» произвольного размера (который может меняться от порции к порции).

Theorem 226. [two-part-versions] *Следующие свойства слова x эквивалентны в описанном смысле (из одного следует другое с логарифмической добавкой к параметрам):*

(а) слово x принадлежит $S_{i,j}$ (имеет $(i * j)$-описание);

(б) существует простой (сложности $O(\log n)$) перечень длины не более 2^{i+j}, в котором слово x в первый раз появляется за 2^j или более шагов от конца;

(в) существует простой (сложности $O(\log n)$) перечень длины не более 2^{i+j}, включающий в себя x, в котором слова выдаются не более чем в 2^i порций;

(г) в любом перечне, включающем все слова сложности не более 2^{i+j} (и только их) по одному разу, слово x встречается за 2^j или более шагов от конца.

$\triangleleft$ Пусть выполнено (а). Будем перечислять все множества сложности не более 2^i и размера не более 2^j. Когда очередное множество появляется, добавляем в перечень все его элементы. При этом получится не более чем 2^i порций с 2^j (или меньше) элементов в каждой, так что всего будет не более чем 2^{i+j} элементов, как и требуется в условии (в). При этом сложность перечня логарифмическая, так как нужно знать лишь i и j. Итак, (а)$\Rightarrow$(в).

Чтобы получить (б), надо немного модифицировать конструкцию и добавлять после каждой порции 2^j произвольных ещё не порождённых элементов; общее число элементов увеличится на 2^{i+j}, что в пределах допустимого. Таким образом, (а)$\Rightarrow$(б).

Напротив, из (б) легко следует (а), достаточно лишь нарезать элементы перечня на части размера 2^j; при этом получится не более чем 2^i частей и останутся ненарезанными не более чем 2^j элементов. Следовательно, x войдёт в одну из частей. Каждая часть задаётся своим порядковым номером и потому имеет сложность i (плюс логарифмическая добавка, включающая в себя сложность перечня).

Чтобы получить (а) из (в), при появлении каждой новой порции элементов перечня мы разбиваем её на части из 2^j элементов или меньше (для последней, неполной, части в каждой порции). При этом для каждой порции пропадает не более 2^j мест, так что общее число потерянных мест не больше 2^{i+j} и число частей есть 2^i, а сложность каждой не больше i.

Итак, свойства (а)–(в) равносильны (с логарифмической точностью), и осталось показать, что они равносильны (г). Очевидно, что (г) является усилением (б), так что достаточно проверить, что из (а) следует (г).

Итак, пусть слово x содержится в некотором конечном множестве A сложности не более i, причём $\log|A| \leqslant j$. Все элементы множества A, в том числе и x, имеют сложность не более $i+j+O(\log(i+j))$; пренебрегая логарифмическими поправками, мы будем опускать последнее слагаемое (надеясь, что привыкший к этому читатель легко внесёт необходимые исправления).

Пусть также дан некоторый простой перечень, включающий в себя по одному разу все слова сложности не более $i+j$ (и только их). Мы хотим доказать, что слово x появится в этом перечне не слишком поздно (после него будут ещё как минимум 2^j слов). Зная множество A, мы можем порождать перечень, пока в нём не обнаружатся все элементы A. Порождённая к этому моменту часть перечня представляет собой конечное множество B сложности не более i (поскольку для задания B достаточно знать A и сам перечень, который прост). Рассмотрим первые (по порядку) 2^j слов вне B: они определяются множеством B (i битов) и своим порядковым номером (j битов) и потому имеют сложность не более $i+j$; это и будут искомые 2^j элементов перечня, стоящие в нём после x. $\triangleright$

$\boxed{\mathbf{270}}$ Докажите, что если есть два простых (имеющих сложность $O(\log n)$) перечня слов сложности меньше n, то для любого слова x сложности меньше x логарифмы количества слов после x в обоих перечнях отличаются не более чем на $O(\log n)$.

Можно сказать, что мы ввели дополнительную классификацию слов сложности не больше n, измеряя логарифм расстояния до конца списка таких слов. С точки зрения двумерной таблицы $S_{i,j}$ это соответствует возрастающей последовательности множеств $S_{i,j}$ на диагонали $i+j=n$ (строго говоря, возрастание множеств вдоль этой диагонали — при увеличении i и соответствующем уменьшении j — имеет место лишь с точностью до логарифмических поправок). Случайные слова длины n (то есть слова сложности n и длины n) стоят в начале этой классификации, имея $(n * 0)$-описания; в конце её стоят (немногочисленные) слова, имеющие лишь $(0 * n)$-описания.

$\boxed{271}$ Покажите, что все слова сложности n, стоящие в самом конце списка (с логарифмической точностью, то есть слова, за которыми стоит $\mathrm{poly}(n)$ слов), «почти одинаковы»: они имеют условную сложность $O(\log n)$ друг относительно друга.

Покажите, что среди них находятся $B(n+O(\log n))$ и $BB(n+O(\log n))$, рассмотренные в разделе 1.2 (с. 32).

Если угодно, можно считать n минус логарифм расстояния до конца списка мерой «извращённости» слова: случайные слова длины n при этом имеют нулевую извращённость, а слова у конца списка (типа $B(n)$) — максимальную (близкую к n). Но надо иметь в виду два важных обстоятельства.

• «извращённость» слов x и y может сильно отличаться, если даже $KS(x|y) \approx 0$ и $KS(y|x) \approx 0$. (В самом деле, кратчайшее описание y любого слова x является случайным и находится в самом начале соответствующего списка.)

Однако если слова x и y переходят друг в друга при простой вычислимой биекции, это уже не так (см. следующую задачу).

• «извращённость» данного слова x сложности n (определяемая его местом в перечне всех слов сложности не выше n) может сильно упасть, если мы рассмотрим то же слово x в перечне слов сложности не больше n' при $n' > n$. Так что фактически характеристикой слова x является функция $n' \mapsto$ расстояние от x до конца перечня слов сложности не выше n'. По существу это та же самая граничная кривая множества P_x, но только в других координатах (теперь мы смотрим на точки, в которых диагонали $i + j = n'$ заходят в множество P_x).

$\boxed{272}$ Пусть слова x и y соответствуют друг другу при вычислимой биекции сложности t. Докажите, что если $x \in S_{i,j}$, то $y \in S_{i+O(\log(i+j)+t),j+O(\log(i+j)+t)}$.

$\boxed{273}$ Обозначим через $KS'(x|y)$ минимальную сложность программы для *всюду определённой* функции, переводящей y в x. Покажите, что $KS(x|y) \leqslant KS'(x|y)$ с логарифмической точностью, но существуют слова x и y длины n, для которых $KS(x|y) \approx 0$ и $K'(x|y) \approx n$.

$\boxed{274}$ Покажите, что если $KS'(x|y) \leqslant n$ и $KS'(y|x) \leqslant t$, то существует вычислимая биекция сложности меньше $2t + O(\log t)$, переводящая x в y. [Указание: строим искомую биекцию по частям с двух сторон.]

Возвращаясь к теореме 15 (с. 32), можно сказать, что эта теорема выделяла среди слов сложности n некоторые особые слова (все имеющие малую сложность относительно друг друга); теперь мы поняли «философский смысл» этих слов — это «наименее стохастические» (=стоящие у конца списка) и эквивалентные им слова. Среди эквивалентных, отметим, есть и стохастические, длина которых близка к сложности. Такова, например, двоичная запись числа k_n слов сложности не больше n (пункт (б) в теореме 15) или числа m_n слов длины не более n, на которых определён оптимальный декомпрессор.

Кстати, использованное при доказательстве теоремы 226 рассуждение позволяет установить связь между k_n (или m_n) при разных n, как показывает следующая задача:

$\boxed{275}$ [omega-words-connections] Докажите, что при $n' < n$ слово $k_{n'}$ (точнее следовало бы сказать «слово, являющееся двоичной записью числа $k_{n'}$) эквивалентно (с точностью до условной сложности $O(\log n)$) первым n' битам слова k_n. Докажите аналогичное утверждение для слов m_n и $m_{n'}$. [Указание. Для k_n по существу надо доказать, что указание списка слов сложности не больше n с не более чем 2^s пропусками равносильно указанию числа, большего $B(n-s)$. В одну сторону: зная такой список, мы дожидаемся T шагов, пока в перечислении слов сложности не больше n появятся все элементы этого списка. Любое число $t > T$ должно иметь сложность не меньше $n-s$, иначе по этому числу можно было построить список сложности меньше $n-s$, содержащий все слова сложности n, кроме 2^s, и первые 2^s слов вне этого списка приводили бы к противоречию. В другую сторону: будем перечислять слова сложности не больше n, выдавая их порциями по 2^s штук; зная $B(n-s)$, можно найти число полных порций (шаг, на котором появляется последняя полная порция, не превосходит $B(n-s)$, поскольку определяется числом полных порций и имеет сложность не больше $n-s$) и тем самым получить искомый список с точностью до последней неполной порции. Аналогичное рассуждение годится и для m_n.]

Следующий результат обобщает утверждение задачи 29 (с. 44). Там по существу утверждалось, что если слово x имеет много описаний данной длины, то оно имеет и описания меньшей длины. Оказывается, что аналогичное утверждение верно и для $(i*j)$-описаний.

Theorem 227. [improving-descriptions-1] *Пусть для слова x существует не менее 2^k множеств, являющихся его $(i*j)$-описаниями. Тогда оно имеет $(i*(j-k))$-описание и даже $((i-k)*j)$-описание.*

В формулировке этой теоремы мы опускаем, как обычно, слагаемые вида $O(\log(i+j+k))$, которые разрешается прибавлять к параметрам описаний. Слово «даже» напоминает о теореме 224, которая позволяет от $(i-k)*j$ перейти к $i*(j-k)$.

$\triangleleft$ Первое (более простое) утверждение теоремы легко следует из рассуждений, использованных в доказательстве теоремы 226. Будем перечислять все множества A сложности не более i и размера не более 2^j и смотреть, какие элементы x покрыты ими с кратностью 2^k. Таких элементов будет не более $2^{i+j}/2^k = 2^{i+j-k}$, и они выдаются не более чем в 2^i порций (каждое новое множество A соответствует одной порции). Остаётся вспомнить свойство (в) теоремы 226.

Для доказательства второго утверждения нам понадобится уменьшить число порций до 2^{i-k}. Вот как это делается. По-прежнему перечисляя множества сложности не более i и размера не более 2^j, мы теперь обращаем внимание не только на «полноценные» элементы, покрытые с кратностью 2^k, но и на «кандидатов» — элементы, покрытые с вдвое меньшей кратностью (2^{k-1}). Как только появляется полноценный элемент, не вошедший в перечень, мы включаем в этот перечень не только его, но заодно уж и всех обнаруженных к этому моменту кандидатов (кроме уже включённых в перечень). От этого общее число включённых в перечень элементов увеличится не более чем вдвое (что несущественно при нашей логарифмической точности). Зато число порций сильно уменьшится: ведь каждая из них включает в перечень всех кандидатов, и чтобы появился непокрытый полноценный элемент, нужно не менее 2^{k-1} новых множеств (кратность должна возрасти по меньшей мере от 2^{k-1} до 2^k). Это и даёт необходимое уменьшение числа порций. $\triangleright$

Только что доказанную теорему можно переформулировать следующим образом:

Theorem 228. [improving-descriptions-2] *Если слово x имеет $(i * j)$-описание A, для которого $KS(A|x) \geqslant k$, то оно имеет и $(i * (j - k))$-описание и даже $((i - k) * j)$-описание.*

(Как и раньше, мы опускаем логарифмические слагаемые, необходимые в точной формулировке.)

$\lhd$ В самом деле, зная слово x, а также значения параметров i и j (последнее требует логарифмического числа битов), мы можем перечислять все $(i * j)$-описания слова x. Поэтому сложность таких описаний (с точностью до логарифмических слагаемых) не превосходит их числа, и если есть описание A с большим $KS(A|x)$, то это гарантирует, что описаний много и можно применить предыдущую теорему. $\rhd$

Эта теорема гарантирует, что если мы берём описания с предельно возможными параметрами (на границе множества P_x для данного x), то все эти описания будут просты относительно x. Что, с точки зрения здравого смысла, довольно естественно: если в описании есть «лишняя» (не входящая в x) информация, то вряд ли оно будет оптимальным...

16.4. Ограниченные классы гипотез

В этом параграфе мы будем предполагать, что класс статистических гипотез ограничен некоторым семейством $\mathcal{A}$, состоящим из конечных подмножеств множества двоичных слов.

Неформально говоря, ограничение класса гипотез означает, что у нас заранее имеется некоторая информация об источнике происхождения данного слова x. А именно, нам известно, что слово x было получено случайным выбором в одном из множеств из $\mathcal{A}$ (но мы не знаем в каком). Поэтому, подыскивая статистическую гипотезу для x, мы ограничиваемся множествами, принадлежащими $\mathcal{A}$.

Оказывается, основные результаты предыдущих параграфов обобщаются (с некоторым ухудшением оценок) на случай любых семейств $\mathcal{A}$, удовлетворяет следующим трем условиям.

(1) Семейство $\mathcal{A}$ перечислимо. Это означает, что имеется алгоритм, печатающий в некотором порядке списки элементов всех множеств из $\mathcal{A}$; список очередного множества можно начинать только тогда, когда уже закончен список предыдущего.

(2) Для всех n семейство $\mathcal{A}$ содержит множество, состоящее из всех слов длины n.

(3) Для некоторого полинома p для каждого множества $A \in \mathcal{A}$ для любых натуральных чисел n и $c < |A|$ выполнено следующее. Существует покрытие множества всех слов длины n из A не более, чем $p(n)|A|/c$ множествами из $\mathcal{A}$, причем мощность каждого из покрывающих множеств не превосходит c.

Для любого слова x обозначим через $P_x^{\mathcal{A}}$ множество пар $\langle i, j\rangle$, для которых x имеет $(i*j)$-описание, *принадлежащее* семейству $\mathcal{A}$. Множество $P_x^{\mathcal{A}}$ всегда включено в множество P_x; чем больше семейство $\mathcal{A}$, тем больше множество $P_x^{\mathcal{A}}$. Для семейства $\mathcal{A}$, состоящего из всех конечных подмножеств, выполнено $P_x^{\mathcal{A}} = P_x$.

Пусть семейство $\mathcal{A}$ обладает свойствами (1)–(3). Тогда для любого слова x множество $P_x^{\mathcal{A}}$ обладает теми же свойствами, что и множество P_x. А именно, для любого слова длины n выполнено следующее.

- Множество $P_x^{\mathcal{A}}$ содержит пару, отстоящую от пары $\langle 0, n\rangle$ не более, чем на $O(\log n)$. Действительно, по свойству (2) семейство $\mathcal{A}$ содержит множество всех слов длины n, а значит любое слово длины n имеет $O(\log n) * n$-описание в $\mathcal{A}$.

- Множество $P_x^{\mathcal{A}}$ содержит пару, отстоящую от пары $\langle \mathit{KS}(x), 0\rangle$ не более чем на константу. Действительно, из условия (3), примененного к $c = 1$ и множеству всех слов длины n в качестве A, следует, что $\mathcal{A}$ содержит все синглетоны, поэтому каждое слово имеет $(\mathit{KS}(x) + O(1)) * 0$-описание.

- Имеет место аналог теоремы 224: $\langle i, j\rangle \in P_x^{\mathcal{A}} \Rightarrow \langle i + k + O(\log n), j - k\rangle \in P_x^{\mathcal{A}}$ (для всех $k \leqslant j$). Действительно, пусть x имеет $i * j$-описание $A \in \mathcal{A}$. Пусть нам дано n, множество A и k. Будем порождать множества из семейства $\mathcal{A}$ до тех пор, пока не найдем покрытие множества всех слов длины n из A не более чем $p(n)2^k$ множествами мощности $|A|2^{-k}$ или менее (здесь p — полином из условия (3)). Сложность множества, покрывающего x, не превосходит $i + k + O(\log n + \log k)$, поскольку его можно найти, зная A, n, k и его порядковый номер среди множеств, покрывающих A. Без ограничения общности мы можем считать, что $k \leqslant n$, поскольку иначе утверждение очевидно: множество $\{x\}$ является $i + k + O(\log n) * (j - k)$-описанием x. Поэтому слагаемое $O(\log k)$ можно опустить.

Пример. Пусть семейство $\mathcal{A}$ состоит из всех *шаров Хэмминга* (то есть множеств вида $\{x \mid l(x) = l(y),\ d(x, y) \leqslant r\}$, где y — некоторое слово (называемое центром шара), r — некоторое натуральное число, называемое радиусом шара, а $d(x, y)$ обозначает расстояние Хэмминга (количество позиций, в которых x и y различны). Неформально говоря, ограничение класса гипотез шарами Хэмминга означает, что нам известно следующее: для некоторого (неизвестного) слова y и некоторого (неизвестного) натурального d слово x было получено из y изменением не более чем d битов, при этом номера измененных битов выбирались случайно. (Слово y было передано по ненадежному каналу передачи, уровень ненадежности которого неизвестен.)

$\boxed{276}$ Докажите, что для любого $r \leqslant n$ множество всех слов длины n можно покрыть $\mathrm{poly}(n)2^n/V$ шарами Хэмминга радиуса r, где V обозначает мощность шара радиуса r. [Указание. Выберем случайным образом N шаров радиуса r. Для фиксированного слова x вероятность того, что оно не будет покрыто ни одним из выбранных шаров равна $(1 - V2^{-n})^N < e^{-V2^{-n}N}$. Если положить $N = n \ln 2 \cdot 2^n/V$, то эта верхняя оценка будет равна 2^{-n}. Значит при таком N вероятность того, что хотя бы одно слово длины n не будет покрыто, меньше 1.]

$\boxed{277}$ Докажите, что семейство всех шаров Хэмминга удовлетворяет условиям (1)–(3). [Указание. Пусть дан шар A радиуса a и число $c < |A|$. Нам нужно покрыть A шарами мощности c или меньше. Поскольку множество всех слов длины n можно покрыть двумя шарами радиуса $n/2$, без ограничения общности мы можем считать, что $a \leqslant n/2$. Выберем наибольшее $b \leqslant n/2$ такое, что мощность шара радиуса b не превосходит c, и будем искать покрытие A шарами радиуса b. Мощности шаров Хэмминга с радиусами, различающимися на 1, отличаются не более, чем в $n + 1$ раз, поэтому $V \geqslant c/(n + 1)$ и нам достаточно покрыть A не более чем $\mathrm{poly}(n)|A|/V$ шарами радиуса b.

Покроем все точки на расстоянии не более b от центра шара A одним шаром радиуса b (с тем же центром). Остаток шара разобьем на концентрические сферы: каждая сфера состоит из всех точек на расстоянии ровно d от центра шара A, где $d \in \{b+1, b+2, \ldots, a\}$ Количество сфер не превосходит n, поэтому достаточно научиться покрывать сферу S радиуса $d \in (b, n/2]$ не более, чем $\mathrm{poly}(n)|S|/V$ шарами радиуса b.

Для этого рассмотрим число f, являющееся решением уравнения $b + f(1 - 2b/n) = d$, округленным до ближайшего целого числа. Выберем случайным образом шар B радиуса b с центром на расстоянии f от центра S. Не менее $1/\operatorname{poly}(n)$-ой части точек любого такого шара B принадлежат S. Действительно, обозначим через x и y центры S и B, соответственно. Выберем в слове y некоторую b/n-ую часть битов, в которых x совпадает с y, и некоторую b/n-ую часть битов, в которых x отличается от y, и изменим все выбранные биты. Любое слово, полученное таким образом, расположено на расстоянии b от y и на расстоянии $f - (b/n)f + (n - f)(b/n) = d$ от x. Общее количество таких слов равно $\binom{f}{f(b/n)}\binom{n-f}{(n-f)(b/n)}$, что с точностью до умножения на полином от n равно $2^{fh(b/n,1-b/n)+(n-f)h(b/n,1-b/n)} = 2^{nh(b/n,1-b/n)}$, то есть, равно V (с той же точностью). Итак, каждый шар B радиуса b с центром на расстоянии f от x прокрывает не менее $V/\operatorname{poly}(n)$ точек из S. При случайном выборе шара B каждое $z \in S$ имеет равные шансы быть покрытым. Поэтому вероятность того, что любое фиксированное $z \in S$ покрывается случайным шаром B, не меньше $V/(|S|\operatorname{poly}(n))$. Следовательно, для подходящего полинома с большой вероятностью случайно выбранные $\operatorname{poly}(n)|S|/V$ шаров радиуса b с центром на расстоянии f от x покроют все точки S.]

278 Пусть семейство $\mathcal{A}$ состоит из всех шаров Хэмминга. Докажите, что существуют такие слова x, для которых множество $P_x^{\mathcal{A}}$ значительно меньше множества P_x (точная формулировка: для некоторого положительного ε для всех достаточно больших n существует слово x длины n, для которого $P_x^{\mathcal{A}}$ отстоит от P_x более чем на εn). [Указание. Фиксируем достаточно маленькое положительное α и обозначим через V мощность шара радиуса αn. Для каждого $n > 3$ существует множество E, состоящее из $N = 2^n/V$ слов длины n такое, что любой шар Хэмминга радиуса αn содержит не более n слов из E. Действительно, рассмотрим случайно и независимо выбранные N слов $y_1, \ldots, y_N$. Вероятность того, что случайно выбранное y_i попадает в фиксированный шар A радиуса αn, есть $V2^{-n}$. Значит вероятность того, что не менее n разных y_i попали в A, не превосходит количества n-элементных подмножеств I множества индексов $\{1, \ldots, N\}$, помноженного на вероятность $(V2^{-n})^n$ попадания в шар A всех y_i для i из данного множества индексов I. Ограничивая сверху биномиальный коэффициент $\binom{N}{n}$ величиной $N^n/n! < (N/2)^n$, получаем верхнюю оценку $(N/2)^n(V2^{-n})^n = 2^{-n}$. Количество шаров радиуса αn равно 2^n, значит с положительной вероятностью случайно выбранные слова $y_1, \ldots, y_N$ образуют нужное множество E. Некоторое множество E с таким свойством может быть найдено перебором по данному n, а значит имеет сложность $O(\log n)$. Возьмем в качестве x случайный элемент E (то есть любое слово x, для которого $\mathit{KS}(x) \geqslant \log|E|$). Сложность любого шара A радиуса αn, содержащего x, будет не меньше $\mathit{KS}(x) - O(\log n)$, поскольку, зная шар A и номер x в $A \cap E$, можно найти x. Поэтому x не имеет $i * \log V$-описаний в $\mathcal{A}$ для всех i меньших $\log|E| - O(\log n)$. С другой стороны, x имеет $O(\log n) * \log|E|$-описание (само множество E). Если α достаточно мало, то $V \leqslant |E|$ при всех достаточно больших n. А значит, P_x содержит некоторую точку вблизи точки $\langle \log|E| - \log V, \log V\rangle$, а все точки P_x^A удалены от точки $\langle \log|E| - \log V, \log V\rangle$ не менее, чем на $(\log V)/3$.]

279 Опишите множество $P_x^{\mathcal{A}}$ для x и $\mathcal{A}$ из предыдущей задачи. [Указание. Граница множества $P_x^{\mathcal{A}}$ состоит из вертикального отрезка $\mathit{KS}(A) = n - \log V$, $\log|A| \leqslant \log V$ и прямолинейного отрезка $\mathit{KS}(A) + \log|A| = n$, $\log V \leqslant \log|A| \leqslant n$, наклоненного под углом в 45° к осям координат.]

Пусть семейство $\mathcal{A}$ обладает свойствами (1)–(3). Тогда выполнен аналог теоремы 225 с немного худшей точностью: $O(\log n)$ надо заменить на $O(\sqrt{n\log n})$.

Theorem 229. *Пусть $k \leqslant n$ и пусть $n = t_0 > t_1 > \ldots > t_k = 0$ — последовательность натуральных чисел, имеющая сложность m. Тогда существует слово x сложности $k + O(\sqrt{n\log n}) + O(m)$ и длины n, для которого множество $P_x^{\mathcal{A}}$ отстоит от множества $T = \{\langle i,j\rangle \mid (i<k) \Rightarrow (j>t_i)\}$ не более чем на $O(\sqrt{n\log n}) + O(m)$.*

$\triangleleft$ В отличие от доказательства теоремы 225, мы не сможем сначала выбрать x, а потом для каждой пары $\langle i,j\rangle$ на границе T указать его $(i+O(\sqrt{n\log n}))*j$-описание. Нам придется строить x вместе с его $(i+O(\sqrt{n\log n}))*j$-описаниями.

Сначала уменьшим множество пар $\langle i,j\rangle$, для которых нам нужно искать описания. Выберем на границе множества T пары $\langle i_l, j_l\rangle$, идущие через равные промежутки длины $\sqrt{n\log n}$ по второй координате. А именно для каждого $l \in \{0,1,\ldots,N = \sqrt{n/\log n}\}$, определим $j_l = l\sqrt{n\log n}$ и положим i_l равным наименьшему i, для которого $\langle i, j_l\rangle \in T$. Докажем, что достаточно построить слово x длины n, не имеющее $(i_l - O(\sqrt{n\log n}))*j_l$-описаний из $\mathcal{A}$ ни для одного $l = 0,1,\ldots,N$, и имеющее $(i_l + O(\sqrt{n\log n}))*j_l$-описание для каждого $l = 0,1,\ldots,N$. Действительно, для всех j, не имеющих вида j_l, рассмотрим l, для которого $j \in [j_l, j_{l+1}]$. Граничная точка множества $P_x^{\mathcal{A}}$ в j-ой горизонтали удалена не более, чем на $O(\sqrt{n\log n})$ от его граничной точки в j_l-ой горизонтали. По построению, последняя удалена не более, чем на $O(\sqrt{n\log n})$ от точки $\langle i_l, j_l\rangle$, а та не более, чем на $O(\sqrt{n\log n})$ от граничной точки T в j-ой горизонтали. Из близости границ множеств $P_x^{\mathcal{A}}$ и T следует близость и самих множеств.

Выберем достаточно большое $\delta = O(\sqrt{n\log n})$ и для каждого $l = 0,1,\ldots,N$ будем перечислять все $(i_l-\delta)*j_l$-описания из $\mathcal{A}$. В каждый момент перечисления будем обозначать через G множество слов, не покрытых ни одним из перечисленных к этому моменту описаний. Мы можем выбрать δ столь большим, чтобы G всегда составляло не менее половины всех слов длины n. Наша цель в том, чтобы после каждого изменения G были определены множества $A_0,\ldots,A_N$ из $\mathcal{A}$ мощностей не более $2^{j_0},\ldots,2^{j_N}$, соответственно. При этом мы хотим, чтобы: (1) после каждого изменения G и последующего обновления $A_0,\ldots,A_N$ множество $G\cap A_0\cap\cdots\cap A_N$ было не пусто, и (2) общее число изменений A_l не превосходило $2^{i_l+O(\sqrt{n\log n})}$ (для всех l). Тогда в качестве x можно будет взять любое слово из множества $G\cap A_0\cap\cdots\cap A_N$ после последнего изменения G. По построению x не будет иметь $(i_l-\delta)*j_l$-описаний из $\mathcal{A}$. Для каждого l последнюю версию множества A_l можно найти, зная n, l и общее количество изменений A_l, а значит его сложность будет ограничена $i_l + O(\sqrt{n\log n})$.

Для того, чтобы поддерживать свойство (1) нам придется следить, чтобы множество $G\cap A_0\cap\cdots\cap A_N$ было не только не пустым, но и, более того, чтобы все его части вида $G\cap A_l\cap\cdots\cap A_N$ были достаточно большими. Порог ν_l, ниже которого не должна опускаться мощность множества $G\cap A_l\cap\cdots\cap A_N$, мы подберем позднее. Сначала посмотрим, какой порог можно гарантировать. В качестве ν_N мы возьмем 1/2, а в качестве A_N множество всех слов длины n, и A_N никогда менять не будем. Как только при появлении нового $(i_l-\delta)*j_l$-описания для хотя бы одного s мощность множества $G\cap A_s\cap\cdots\cap A_N$ станет меньше ν_s, мы рассмотрим наибольшее s, для которого произошло нарушение, и обновим множества $A_1,\ldots,A_s$ следующим образом.

Зафиксируем полином $p(n)$ из условия (3) и положим $\alpha = p(n)$. Покроем множество A_{s+1} не более чем $\alpha 2^{j_{s+1}-j_s}$ множествами из $\mathcal{A}$ мощности не более 2^{j_s} и выберем среди них то, которое покрывает наибольшее количество слов из $G \cap A_{s+1} \cap \dots \cap A_N$. Это и будет новое множество A_s. Аналогичным образом определим A_{s-1} как то множество из покрытия множества A_s, которое покрывает наибольшее количество элементов из $G \cap A_s \cap \dots \cap A_N$, и т.д. Построение гарантирует нам, что при всех $t \leqslant s$ мощность $G \cap A_t \cap \dots \cap A_N$ не более, чем в $\alpha^{s+1-t} 2^{j_{s+1}-j_t}$ меньше мощности $G \cap A_{s+1} \cap \dots \cap A_N$. Таким образом, можно положить $\nu_l = \alpha^{l-N} 2^{j_l-1}$. Тогда после обновления мощность множества $G \cap A_l \cap \dots \cap A_N$ будет не меньше ν_l для всех $l = 0, 1, \dots, N$. Следовательно, множество $G \cap A_0 \cap \dots \cap A_N$ никогда не станет пустым.

Однако выбранное значение порога ν_l не гарантирует того, что A_l изменяется не больше $2^{i_l + O(\sqrt{n \log n})}$ раз. В худшем случае может оказаться, что оно меняется после каждого изменения G. Однако, если уменьшить порог ν_l до величины $\nu_l = (2\alpha)^{l-N} 2^{j_l-1}$, то после каждого обновления A_l мощность множества $G \cap A_l \cap \dots \cap A_N$ будет вдвое превосходить ν_l. Поэтому, в следующий раз порог ν_l будет перейден только после того, как мощность G уменьшится не менее чем на ν_l.

Поскольку обновление A_l может быть также вызвано также тем, что перейден порог ν_s при некотором $s > l$, нам нужно оценивать сверху общее число перехода порогов ν_s при $s \geqslant l$. Между любыми двумя последовательными переходами порога ν_s мощность множества G уменьшалась не менее, чем на ν_s, что не меньше, чем ν_l. Уменьшение мощности G на ν_l может быть вызвано порождением нового “большого” множества из $\mathcal{A}$ (большими мы называем множества мощности больше 2^{j_l}), либо порождением достаточного количества мелких множеств (мощности не больше 2^{j_l}). Общее количество больших множеств не превосходит $\sum_{s>l} 2^{i_s-\delta}$, а общее количество слов в мелких множествах не превосходит $\sum_{s \leqslant l} 2^{i_s+j_s-\delta}$. Поэтому общее количество перехода порога ν_s (для любого $s \geqslant l$) ограничено сверху величиной

$$\sum_{s>l} 2^{i_s-\delta} + \sum_{s \leqslant l} 2^{i_s+j_s-\delta}/\nu_l \leqslant N 2^{i_l-\delta} + N 2^{i_l+j_l-\delta}/\nu_l$$
$$= N 2^{i_l-\delta} + N 2^{i_l+j_l-\delta+N\log(2\alpha)-j_l+1} = N 2^{i_l-\delta} + N 2^{i_l-\delta+\sqrt{n/\log n}\cdot O(\log n)}.$$

Общее количество обновлений A_l превосходит эту верхнюю границу не более чем в N раз, и подходящим выбором δ может быть сделано меньше 2^{i_l}. $\triangleright$

$\boxed{\mathbf{280}}$ (1) Пусть x слово длины n, а r натуральное число, не превосходящее $n/2$. Обозначим через $\mathit{KS}_r(x)$ наименьшую (простую) сложность слова y той же длины, что x, и отличающегося от x не более, чем в r позициях. Докажите, что $\mathit{KS}_r(x)$ с точностью до $O(\log n)$ равно наименьшему i, для которого x имеет $i * \log V(r)$-описание, являющееся шаром Хэмминга, где $V(r)$ обозначает мощность шара Хэмминга радиуса r.

(2) Опишите все возможные формы функции $\mathit{KS}_r(x)$, как функции r, с точностью до слагаемого $O(\sqrt{n \log n})$. [Указание. Для всех слов x длины n выполнено $\mathit{KS}_0(x) = \mathit{KS}(x)$, $\mathit{KS}_n(x) = O(\log n)$ и

$$0 \leqslant \mathit{KS}_a(x) - \mathit{KS}_b(x) \leqslant \log(V(b)/V(a)) + O(\log n)$$

для всех $a < b \leqslant n/2$. Обратно, для любых $k \leqslant n$ и любой функции $t : \{0, 1, \dots, n\} \to \mathbb{N}$ сложности m такой, что $t(0) = k$, $t(n) = 0$ и $0 \leqslant t(a) - t(b) \leqslant \log(V(b)/\log V(a))$ для всех

$a < b \leqslant n/2$, существует слово x длины n и сложности $k + O(\sqrt{n \log n}) + O(m)$ такое, что $\mathit{KS}_a(x) = t(a) + O(\sqrt{n \log n}) + O(m)$.]

Теорема 226 давала критерий того, что данное слово имеет $i*j$-описание (без ограничения класса описаний). Неясно, имеет ли она аналог для произвольного семейства множеств $\mathcal{A}$. А следующая за ней теорема 227 переносится почти без изменений на случай произвольного перечислимого семейства множеств.

Theorem 230. [improving-descriptions-1-gen] *Пусть семейство $\mathcal{A}$ перечислимо. Пусть слово x длины n имеет не менее 2^k множеств из $\mathcal{A}$, являющихся его $(i*j)$-описаниями. Тогда оно имеет $(i-k)*j$-описание в $\mathcal{A}$ (и, следовательно, $i*(j-k)$-описание, если семейство $\mathcal{A}$ удовлетворяет требованию (3)).*

В формулировке этой теоремы мы опускаем, как обычно, слагаемые вида $O(\log(n+i+j+k))$, которые разрешается прибавлять к параметрам описаний.

$\lhd$ Будем порождать все $(i*j)$-описания из $\mathcal{A}$, то есть множества из $\mathcal{A}$ мощности не более 2^j и сложности не более i. Фиксируем n и будем отбирать некоторые из порожденных описаний, стремясь обеспечить следующее: в любой момент каждое слово x длины n, которое имеет не менее 2^k описаний (среди уже порожденных) принадлежит хотя бы одному из отобранных множеств, а общее количество отобранных множеств не превосходит $2^{i-k}p(n,k,i,j)$, где p — некоторый полином. Нам достаточно доказать, что для некоторого полинома p существует стратегия отбора, гарантирующая выполнение этих двух требований. (Поскольку стратегия может быть найдена перебором по данным n,i,j,k, ее колмогоровская сложность ограничена $O(\log(n+k+i+j))$.)

Существование такой стратегии можно доказать конструктивно и с помощью вероятностного рассуждения.

Вероятностное доказательство. Рассмотрим игру двух противников, в которой игроки ходят по очереди, и каждый делает 2^i ходов. Первый игрок в свою очередь указывает некоторое множество слов длины n, а второй на следующем за этим ходу сообщает, отбирает он это множество или нет. Второй игрок проигрывает, если после некоторого его хода количество отобранных множеств превысит $2^{i-k+1}(n+1)\ln 2$ или найдется слово x, принадлежащее не менее 2^k множествам первого игрока, но не принадлежащее ни одному из отобранных множеств.

В этой игре один из игроков имеет выигрышную стратегию. Мы утверждаем, что этим игроком является второй. Рассуждая от противного, допустим, что выигрышную стратегию имеет первый и будем в дальнейшем предполагать, что первый игрок придерживается некоторой фиксированной выигрышной стратегии.

Рассмотрим следующую вероятностную стратегию для второго игрока: после каждого хода A первого игрока отбираем множество A с вероятностью $p = 2^{-k}(n+1)\ln 2$. Чтобы получить противоречие, нам достаточно доказать, что эта стратегия обыгрывает с положительной вероятностью зафиксированную выигрышную стратегию первого игрока.

Среднее количество отобранных множеств равно $p2^i = 2^{i-k}(n+1)\ln 2$. По неравеству Чебышева вероятность того, что количество отобранных множеств вдвое превысит свое среднее значение, меньше 1/2. Поэтому достаточно доказать, что второе условие выигрыша первого игрока (после некоторого хода найдется x, принадлежащее не менее 2^k множествам первого игрока, но не принадлежит ни одному из отобранных множеств) выполнено с вероятностью не более $1/2$.

Нам достаточно доказать, что для любого фиксированного x это условие выполняется с вероятностью не более 2^{-n-1}. Докажем индукцией по t, что вероятность события «после некоторого хода второго игрока слово x принадлежит не менее чем t множествам первого игрока, но не принадлежит ни одному из отобранных множеств" не превосходит $(1-p)^t$ (это событие мы обозначим через R_t). Для $t = 0$ утверждение очевидно. Чтобы сделать индуктивный переход, нам достаточно показать, что вероятность R_{t+1} при условии R_t не больше $1-p$.

Пусть $z = (z_1, z_2, \ldots, z_s)$ произвольная последовательность первых s ходов второго игрока ($z_i = 1$, если i-ое множество первого игрока отбирается, и $z_i = 0$, иначе). Будем называть последовательность z неудачной, если после s-ого хода второго игрока слово x впервые оказалось покрыто t множествами первого игрока и не покрыто ни одним из отобранных множеств (из этого следует, что s-ое множество первого игрока содержит x и $z_s = 0$). Неудачные последовательности попарно несогласованы и каждая последовательность ходов второго игрока, принадлежащая событию R_t начинается с некоторой неудачной последовательности. Поэтому события "стратегия второго игрока сделала ходы $z = (z_1, z_2, \ldots, z_s)$" для неудачных z образуют разбиение события R_t. Значит нам достаточно доказать, что для любой неудачной последовательности z вероятность события R_{t+1} при условии, что второй игрок сделал ходы z" не больше $1-p$. Эта условная вероятность есть вероятность того, что выигрышная стратегия первого игрока в ответ на ходы $z_1, \ldots, z_s$ и случайные последующие ходы второго игрока в некоторый момент выдаст множество, содержащее x, а второй игрок не отберет его. Поскольку решение не отбирать это множество принимается с вероятностью $1-p$ независимо от предыстории игры, оцениваемая вероятость равна произведению $1-p$ и вероятности того, что первый игрок в ответ на ходы $z_1, \ldots, z_s$ и случайные последующие ходы второго игрока в некоторый момент выдаст множество, содержащее x. Следовательно, она не превосходит $1-p$.

При $t = 2^k$ мы получаем

$$(1-p)^t = (1 - 2^{-k}(n+1)\ln 2)^{2^k} < e^{-2^{-k}(n+1)\ln 2 \cdot 2^k} = 2^{-n-1}.$$

Коструктивное доказательство. Рассмотрим ту же самую игру, только верхнюю границу количества отобранных множеств $2^{i-k+1}(n+1)\ln 2$ заменим на $2^{i-k}i^2 n \ln 2$ и разрешим второму игроку на каждом ходу отбирать несколько множеств (из множеств, указанных первым игроком на предыдущих ходах). Укажем явно выигрышную стратегию второго игрока в этой игре. Она состоит из независимого применения i стратегий, занумерованных числами $1, 2, \ldots, i$.

Стратегия номер s просыпается после ходов первого игрока, номер которых кратен 2^s. Проснувшись, она формирует семейство S, состоящее из последних 2^s множеств первого игрока и множество T, состоящее из всех слов, которые покрыты не менее $2^k/i$ множествами из S. Затем она отбирает некоторые множества из S так, чтобы все слова из T оказались покрыты отобранными множествами. Применяя жадный алгоритм (берем множество из S, покрывающее наибольшее количество слов из T, затем множество, которое покрывает наибольшую часть остатка и т.д.), мы можем сделать это, отобрав не более $in2^{s-k}\ln 2$ множеств. Действительно, каждое слово из T покрыто не менее, чем $2^k/i$ множествами из S. Поэтому некоторое множество из S покрывает не менее $|T|2^{k-s}/i$ слов из

T (общее число пар $\langle x, A\rangle$, где $x \in A \in S$ и $x \in T$, не меньше $|T|2^k/i$, поэтому хотя бы одно множество $A \in S$ встречается не менее, чем в $|T|2^{k-s}/i$ парах). Значит, и жадный алгоритм выберет множество, покрывающее не менее $|T|2^{k-s}/i$ слов из T. Аналогичные рассуждения применимы и к остатку T: второе множество, выбранное жадным алгоритмом, покрывает не менее $2^{k-s}/i$-ой доли остатка T. Поэтому количество элементов T, оставшихся непокрытыми после выбора $in2^{s-k}\ln 2$ множеств не превосходит

$$|T|(1-2^{k-s}/i)^{in2^{s-k}\ln 2} < 2^n e^{(-2^{k-s}/i)in2^{s-k}\ln 2} = 1,$$

то есть непокрытых элементов не останется. (Можно было бы применить и вероятностное рассуждение, показав, что случайно выбранные $in2^{s-k}\ln 2$ множеств из S с положительной вероятностью покроют T. Но в этом случае стратегия стала бы менее явной.)

Таким образом, общее количество множеств, отобранных стратегией с номером s, не превосходит $in2^{i-s}2^{s-k}\ln 2 = in2^{i-k}\ln 2$, что дает в целом не более $i^2n2^{i-k}\ln 2$ отобранных множеств.

Осталось доказать, что после каждого нашего хода любое слово принадлежащее не менее чем 2^k множествам первого игрока, покрыто некоторым отобраным множеством. Рассмотрим ход номер t и любое непокрытое после этого хода слово x. Разложим t в двоичную запись: $t = 2^{s_1} + 2^{s_2} + \ldots$, где $s_1 > s_2 > \ldots$. Поскольку x не принадлежит множествам, отобранным стратегиями с номерами $s_1, s_2, \ldots$, кратность x среди первых 2^{s_1} множеств меньше $2^k/i$, кратность x среди следующих 2^{s_2} множеств также меньше $2^k/i$ и т.д. Таким образом, после хода t слово x покрыто менее, чем $i \cdot 2^k/i = 2^k$ множествами первого игрока. $\triangleright$

Как и в случае неограниченного класса гипотез, из этой теоремы следует теорема об упрощении гипотез, содержащих «лишнюю» информацию.

Theorem 231. [improving-descriptions-2-gen] *Пусть семейство $\mathcal{A}$ перечислимо. Если слово x имеет $i * j$-описание $A \in \mathcal{A}$, для которого $\mathit{KS}(A|x) \geqslant k$, то x имеет и $(i-k) * j$-описание из $\mathcal{A}$ (и, следовательно, $i * (j-k)$-описание, если семейство $\mathcal{A}$ удовлетворяет требованию (3)).*

(Как и раньше, мы опускаем логарифмические слагаемые, необходимые в точной формулировке.)

16.5. Дефект оптимальности и дефект случайности

Мы рассматривали две характеристики конечного множества A, содержащего слово x: (а) дефект случайности $d(x|A) = \log|A| - \mathit{KS}(x|A)$ и (б) разность $\log|A| + \mathit{KS}(A) - \mathit{KS}(x)$, которая показывает, насколько длина двухчастного описания слова x с помощью множества A отличается от минимально возможной. Эту разность мы будем называть *дефектом оптимальности* гипотезы A о слове x и обозначать $\delta(x|A)$

Как связаны две эти характеристики? Начнём с очевидного наблюдения:

Theorem 232. [two-defects] *Дефект случайности слова x относительно конечного множества A, содержащего это слово, не превосходит дефекта оптимальности (с логарифмической точностью):*

$$d(x|A) \leqslant \delta(x|A) + O(\log l(x)).$$

◁ Нам нужно доказать, что

$$\log|A| - \mathit{KS}(x|A) \leqslant \log|A| + \mathit{KS}(A) - \mathit{KS}(x) + O(\log l(x)).$$

После сокращения $\log|A|$ мы получаем неравенство

$$\mathit{KS}(x) \leqslant \mathit{KS}(A) + \mathit{KS}(x|A) + O(\log l(x));$$

правая часть его (с логарифмической точностью) есть сложность пары $\langle x, A\rangle$ и потому не меньше $\mathit{KS}(x)$. ▷

Как видно из этого рассуждения, разница между дефектом случайности и дефектом оптимальности равна $\mathit{KS}(x, A) - \mathit{KS}(x)$, то есть $\mathit{KS}(A|x)$ с точностью $O(\log l(x) + \log \mathit{KS}(A))$, что есть $O(\log l(x))$, если $\mathit{KS}(A) = O(\mathit{KS}(x))$. (Более сложных гипотез мы рассматривать не будем: какой смысл в объяснении, которое значительно сложнее объясняемого объекта?)

Легко привести пример гипотезы, для которой дефект оптимальности значительно больше дефекта случайности. Например, для случайного слова x длины n рассмотрим гипотезу B, состоящую из всех слов длины n, и добавим в B случайное слово y длины $n-1$, независимое от x. Тогда $\mathit{KS}(B|x)$ примерно равно n, и дефект оптимальности превосходит дефект случайности (который по-прежнему близок к нулю, несмотря на добавление слова y), примерно на n. Видно, что в этом случае гипотеза «плоха» с интуитивной точки зрения, поскольку содержит явно бессмысленный элемент y, никакого отношения к наблюдаемому x не имеющий; удаление y позволило бы улучшить гипотезу, приблизив дефект оптимальности к дефекту случайности (который практически не меняется).

Доказанная нами теорема 228 говорит, что нечто подобное можно сделать всегда: если для данной гипотезы B, объясняющей слово x, дефект оптимальности $\delta(x|B)$ больше дефекта случайности $d(x|B)$ (на величину $\mathit{KS}(B|x)$, как мы знаем), то можно найти другую гипотезу A (не большей сложности, чем B), у которой дефект оптимальности $\delta(x|A)$ которой не превосходит $d(x|B)$.

Поэтому вопрос о том, можно ли для данного x найти гипотезу A с $\mathit{KS}(A) \leqslant \alpha$ и $d(x|A) \leqslant \beta$, который мы задавали в определении (α, β)-стохастичности, равносилен (с логарифмической точностью) вопросу о том, можно ли найти гипотезу A с $\mathit{KS}(A) \leqslant \alpha$ и $\delta(x|A) \leqslant \beta$.

Тем самым (для данного слова x) устройство множества P_x (см. теорему 225) полностью определяет, при каких α и β слово x является (α, β)-стохастическим: это будет, как легко проверить, если точка $(\alpha, \mathit{KS}(x) - \alpha + \beta)$ принадлежит P_x.

281 Проверьте это утверждение (как всегда, оно понимается с логарифмической точностью).

282 `[improved-nonstochastic]` Выведите отсюда обещанное выше (с. 386) утверждение о том, что если $\alpha + \beta < n - O(\log n)$, то существуют слова длины n, не являющиеся (α, β)-стохастическими.

[Можно ли таким же способом улучшить оценку на долю нестохастических слов?]

283 Для данного слова x рассмотрим множество Q_x, состоящее из пар $\langle \alpha, \beta\rangle$, для которых слово x является (α, β)-стохастическим. Докажите, что граница множества Q_x является (с логарифмической точностью) убывающей кривой, ведущей из точки $\langle 0, l(x)\rangle$ в точку $\langle KS(x), 0\rangle$, и что любая простая убывающая кривая может являться такой границей.

284 Докажите, что если для данного слова x и данного α существует гипотеза, доставляющая минимум дефекта случайности среди гипотез сложности не выше α, но её дефект оптимальности превосходит её дефект случайности на γ, то граница множества P_x на участке от $\alpha-\gamma$ до α (считая по оси абсцисс) идёт под углом $45°$.

[Указание. Воспользуйтесь более сильным утверждением теоремы 228.]

285 Пусть $\mathcal{A}$ — любое семейство конечных множеств слов, удовлетворяющее требованиям (1)–(3). Докажите, что для любого x утверждения “существует множество $A \in \mathcal{A}$ сложности не более α с $d(x|A) \leqslant \beta$”, “существует множество $A \in \mathcal{A}$ сложности не более α с $\delta(x|A) \leqslant \beta$” и “точка $\langle \alpha, KS(x) - \alpha + \beta\rangle$ принадлежит $P_x^{\mathcal{A}}$” эквивалентны (с логарифмической точностью).

286 [p270] Пусть $\mathcal{A}$ произвольное семейство конечных множеств, перечисляемое программой p. Докажите, что для любого x утверждения “существует множество $A \in \mathcal{A}$ с $d(x|A) \leqslant \beta$” и “существует множество $A \in \mathcal{A}$ с $\delta(x|A) \leqslant \beta$” эквивалентны (с точностью $O(l(p) + \log KS(A) + \log n + \log\log|A|))$.

16.6. Минимальные гипотезы

Пусть фиксировано некоторое слово x. Ему, как мы видели, соответствует множество P_x пар (α, β), для которых x имеет $(\alpha * \beta)$-описание. Такие описания, как мы говорили, естественно рассматривать как «статистические гипотезы», объясняющие происхождение слова x. Интересно понять, как именно устроены эти гипотезы. Оказывается, что можно более или менее явно описать некоторый класс гипотез, к которым в некотором смысле сводится любая гипотеза. Этот класс проистекает из доказательства теоремы 226.

Пусть l — произвольное число, большее $KS(x)$. Тогда перечень всех слов сложности не более l содержит x. Фиксируем такой перечень (простой алгоритм, порождающий все эти слова по одному разу) Пусть N_l — число слов в перечне. Запишем N_l в двоичной системе счисления, то есть представим его в виде суммы убывающих степеней двойки:

$$N_l = 2^{s_1} + 2^{s_2} + \ldots + 2^{s_t}, \text{ где } s_1 > s_2 > \ldots > s_t.$$

В соответствии с этим разложением и перечень можно разбить на группы в порядке порождения элементов: сначала идут 2^{s_1} элементов, затем 2^{s_2} элементов и так далее. Слово x попадает в одну из частей. Эту часть (соответствующее конечное множество) мы и будем рассматривать как описание слова x. Таким образом, мы получаем некоторое семейство описаний (при каждом $l > KS(x)$ — своё описание).

Следующие две теоремы устанавливают обещанные свойства этих описаний (во-первых, являются минимальными, то есть лежат на границе множества P_x; во-вторых, всякое описание слова x в некотором смысле сводится к одному из них).

Theorem 233. [boundary-description] *Пусть в описанной ситуации слово x попало в часть размера 2^s. Тогда эта часть является $((l-s)*s)$-описанием и точка $(l-s, s)$ лежит на границе множества P_x (точнее говоря, эта часть является $((l-s+O(\log l))*s)$-описанием и соответствующая точка лежит в $O(\log l)$-окрестности границы).*

$\lhd$ Чтобы задать часть, достаточно знать её размер, а также сколько элементов имеется перед ней, то есть достаточно знать s, l и все биты числа N_l, кроме s последних (то есть $l-s$ битов). Ещё нужно знать сам перечень, но он по предположению имеет логарифмическую сложность. Поэтому сложность рассматриваемой части есть $l-s+O(\log l)$, а число элементов в ней равно 2^s, что и требовалось.

Если бы точка $(l-s,s)$ не лежала бы на границе множества P_x, а входила бы в P_x вместе с некоторой окрестностью более чем логарифмического размера, то слово x имело бы заметно лучшие двухчастные описания, чем построенное нами (с той же или даже немного меньшей суммарной длиной и с бóльшим размером множества) и по теореме 226 (пункт (г)) слово x появлялось бы в перечне более чем за 2^s элементов до конца (что противоречит построению части). $\rhd$

Следующая теорема показывает, что описаний рассмотренного вида в некотором смысле достаточно. Пусть даны произвольное слово x и произвольное конечное множество A, его содержащее. Пусть максимальная сложность слов из A равна l. Как и раньше, разобьём слова сложности не выше l (их число обозначим N_l) на части, размеры которых есть степени двойки, соответствующие единичным битам в двоичной записи N_l. Пусть B — часть, в которую попало слово x, и пусть она имеет размер 2^s.

Theorem 234. [description-universal] *Гипотеза B (как объяснение для слова x) не хуже гипотезы A с точки зрения зрения сложности и дефекта оптимальности: (а) KS$(B) \leqslant$ KS(A); (б) $\delta(x|B) \leqslant \delta(x|A)$ (с точностью $O(\log l)$). Кроме того, (в) KS$(B|A) = O(\log l)$ (гипотеза B проста относительно A.*

$\lhd$ Зная A и значение l, мы можем дождаться, когда в перечне всех слов сложности не больше l появятся все слова из множества A. К этому моменту появится и слово x (входящее в часть размера 2^s), так что останутся необнаруженными не более $O(2^s)$ слов (из этой и следующих частей). Таким образом, мы знаем N_l с ошибкой не более чем $O(2^s)$ и тем самым можем знаем его первые $l-s$ его битов с ошибкой $O(1)$. А по этой информации (а также l и s) мы можем получить B. Таким образом, KS$(B|A) = O(\log l)$. Утверждение (в) (а тем самым и (а)) доказано.

Утверждение (б) очевидно по построению: если KS$(A) = \alpha$ и $\log|A| = \beta$, то все слова множества A имеют $(\alpha*\beta)$-описание и сложность $\alpha+\beta+O(\ logl)$, так что и максимальная сложность l не превосходит $\alpha+\beta+O(\log l)$. С другой стороны, построенное двухчастное описание, как мы видели в предыдущей теореме, является $((l-s)*s)$-описанием, так что его суммарная длина (и потому дефект оптимальности) не больше, чем у A. $\rhd$

Связь между параметрами гипотез A и B в этой теореме показана на рис. 49: точка изображает параметры гипотезы A, а серым цветом показана область, где могут находиться параметры гипотезы B.

Что будет, если начальная гипотеза A сама уже была минимальной (близкой к границе множества P_x)? Можно ли утверждать, что B в этом случае имеет те же параметры, что и A? Вообще говоря, нет: возможно, что гипотеза B находится на пунктирной границе серой области (рис. 49). (Внутри серой области она находиться не может, поскольку в этом случае гипотеза A была бы, как легко проверить, внутренней для P_x.)

Можно сказать ещё и так: если представлять себе границу множества P_x как кривую, идущую поочерёдно вправо-вниз (под углом 45°) и вертикально вниз, то все точки поворота, в которых вертикальное движение сменяется наклонным, имеют соответствующие

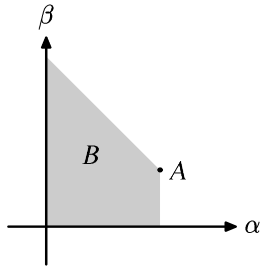

Рис. 49. Параметры гипотезы A и её «упрощения» B

[mdl.3.eps]

им гипотезы нашего семейства (поскольку для таких A серая область, в которой может находиться B, пересекается с P_x в единственной точке).

Заметим, что информация, содержащаяся в гипотезах построенного нами семейства, как это ни странно, почти не зависит от слова x: гипотеза B содержит ту же информацию, что и начало слова N_l длины $l - s$. Как мы видели в задаче 275 (с. 395), это начало можно заменить на слово N_{l-s} или (теорема 93, с. 147) на начальный отрезок числа Чейтина Ω. Тем самым наши гипотезы (по мере движения по границе P_x слева направо) содержат всё больше и больше битов числа Ω, доходя до разности между сложностью x и логарифмом его порядкового номера в списке слов этой (и меньшей) сложности. Так что (как это ни прискорбно с точки зрения философии, которой посвящён следующий раздел) гипотезы семейства содержат всё больше и больше информации не о x (как хотелось бы), а о числе Ω (которое от x не зависит).

Ещё интересно отметить, что это наблюдение (о том, что одна гипотеза содержит часть информации из другой) относится лишь к построенным нами гипотезам, а не к произвольным гипотезам на границе множества P_x. Это показывает следующий пример. Пусть x — случайное слово длины n. Рассмотрим две следующих гипотезы: множество слов длины n, имеющих ту же первую половину, что у x, и множество слов длины n, имеющих ту же вторую половину, что у x. Обе гипотезы имеют совсем небольшой дефект оптимальности, однако содержат разную информацию. (Это не противоречит доказанной теореме, поскольку в качестве гипотезы B, которая лучше обеих, может фигурировать множество всех слов длины n.)

16.7. Немного философии

[mdl-philosophy]

Задачу поиска двухчастного описания для данного слова x можно разукрашивать по-разному. Можно говорить о науке, которая наблюдает внешний мир в лице слова x и строит о нём разные гипотезы, пытаясь отобрать «наиболее правильную». При этом критерием отбора является простота гипотезы (измеряемая колмогоровской сложностью множества A; чем эта сложность меньше, тем гипотеза проще, то есть лучше), а также её «конкретность», или «объясняющая способность» (которая измеряется размером множества A; чем оно меньше, тем гипотеза конкретнее, то есть лучше). Под это дело можно даже подвести философскую базу в лице классика философии Оккама и его бритвы, согласно которой не следует умножать сущности без надобности. Ещё можно стремиться к тому, чтобы дефект случайности наблюдаемого слова x относительно предложенной гипотезы A был мал («не оставалось необъяснённых закономерностей»).

Можно изложить и более прагматичную версию. Колмогоровскую сложность можно рассматривать как предельный вариант науки сжатии файлов: сложность слова x есть нижняя граница для всех способов его сжатия без потери информации. Программа-компрессор тем лучше, чем ближе она к этой границе подходит (для той или иной практически важной категории файлов).

Это относится к сжатию без потерь информации. Однако всё большее применение находят способы сжатия, где некоторая «несущественная» часть информации теряется и за счёт этого достигается большее сжатие.

Пусть, скажем, есть старая граммофонная пластинка, на которой от долгого хранения в пыльном чулане появились царапины (в случайных местах дорожки, содержащей запись). Царапины соответствуют различным паразитным пикам на осциллограмме сигнала (графике зависимости звукового давления от времени). Таким образом, изначальная информация подверглась случайному искажению. Сложность записи при этом сильно возросла, если царапин много. Но если мы хотим передать лишь общее впечатление от проигрывания этой пластинки, то нам неважно, где именно находятся царапины, а важен лишь общий характер дефектов.

Другими словами, фактическая запись на пластинке является одним из элементов большого множества записей с «такого же типа царапинами». Тем самым получается двучастное описание: сначала мы задаём это множество (скажем, указав начальную неискажённую запись и статистические характеристики наложенного шума), а затем указываем конкретный элемент этого множества (где именно царапины). Если при сжатии мы утратим вторую часть описания, сохранив лишь первую, то после восстановления мы получим запись, где на тот же исходный сигнал будет наложен шум с теми же статистическими характеристиками, но уже другой. Можно ожидать, что слушатель не заметит этой подмены. А если удастся шум вообще не накладывать, тем самым «очистив» запись, это ещё лучше (если нас интересует музыка, а не ностальгическое очарование старинного граммофона).

Пользуясь этой аналогией, можно следующим образом интерпретировать утверждение задачи 286. Пусть слово x было получено из неизвестного слова y той же длины наложением шума, то есть, для некоторого известного нам r слово x было выбрано случайно в шаре Хэмминга радиуса r с центром y. Желая удалить шум, мы ищем шар Хэмминга радиуса r, дающий минимально короткое двухчастное описание для x (то есть, шар минимальной сложности). Допустим нам это удалось и мы в самом деле нашли шар минимально возможной сложности. С большой вероятностью дефект случайности x в исходном шаре мал, а значит по задаче 286 (примененной к семейству, состоящему из множества шаров Хэмминга радиуса r) мал будет и дефект оптимальности x в найденном нами шаре. Это означает, что в найденном двухчастном описании x вторая часть не содержит полезной информации. Иными словами, центр найденного шара является очищенной от шума версией x (в частности, и от того шума, который уже был в y).

Другой пример: в фотографии, скажем, песчаной гряды, сделанной с большим разрешением, зафиксированы положения отдельных песчинок, которые случайны, и потому сложность этой фотографии велика. Однако для зрителя она является всего лишь «типичным элементом» множества фотографий, где песчаная гряда находится на том же месте и сложена из такого же песка, хотя конкретные песчинки могут быть в разных местах. Если при сжатии сохранится только информация об этом множестве, а при восстановлении будет построен другой типичный его элемент, то разница будет незаметна.

Надо иметь в виду, что эта аналогия — всего лишь аналогия, причём довольно далёкая, и вряд ли математические результаты о колмогоровской сложности двухчастных описаний могут найти непосредственные «практические применения». (Хотя бы потому, что мы полностью игнорируем вычислительную сложность алгоритмов декодирования, а алгоритмы кодирования вообще не рассматриваем. Может быть, именно этим объясняется парадоксальная независимость оптимальных гипотез от слова x, которую мы отмечали выше.)

Используемые понятия и обозначения

Множество целых чисел обозначается $\mathbb{Z}$, множество натуральных чисел (включая нуль) — $\mathbb{N}$ множество действительных чисел — $\mathbb{R}$. Множество рациональных чисел обозначается $\mathbb{Q}$, среди них выделяются двоично-рациональные числа (конечные двоичные дроби), имеющие вид $m/2^n$ при целых m и n. ч

Число элементов в конечном множестве A обозначается $|A|$.

Если основание логарифмов не указано явно, $\log x$ обозначает логарифм по основанию 2 (как обычно, $\ln x$ — натуральный логарифм).

В некоторых оценках используется обозначение $\lfloor x \rfloor$ для целой части числа x (наибольшего целого числа, не превосходящего x), а также обозначение $\lceil x \rceil$ для наименьшего целого числа, большего или равного x.

Как обычно, запись $f \leqslant g + O(1)$ (где f и g — выражения, которые могут содержать переменные) означает, что существует число c, для которого $f \leqslant g + c$ при всех значениях переменных. Аналогичным образом $f \leqslant g + O(h)$ (при неотрицательном h) означает, что при некотором c и при всех значениях переменных выполняется неравенство $f \leqslant g + ch$. Запись $f = g + O(h)$ (при неотрицательном h) означает, что при некотором c и при всех значениях переменных выполняется неравенство $|f - g| \leqslant ch$.

Через $\mathbb{B}$ мы обозначаем множество $\{0, 1\}$. Конечные последовательности нулей и единиц называются *двоичными словами*, их множество мы обозначаем Ξ. Для любого конечного множества (*алфавита*) A через A^n обозначается множество всех *слов алфавита* A длины n, то есть множество всех последовательностей длины n, составленных из элементов множества A (*букв* алфавита A). Через A^* обозначается множество слов всех длин (включая *пустое* слово Λ длины 0), так что, например, $\Xi = \mathbb{B}^*$. Длина слова x обозначается $l(x)$. Через ab мы обозначаем *конкатенацию* слов a и b, то есть результат приписывания слова b справа к слову a. Говорят, что слово a является *началом*, или *префиксом*, слова b, если $b = ax$ для некоторого слова x. Говорят, что a является *концом*, или *суффиксом*, слова b, если $b = xa$ для некоторого слова x . Говорят, что a является *подсловом* слова b, если $b = xay$ для некоторых слов x и y (другими словами, если a является началом конца b или концом начала b).

Мы рассматриваем также бесконечные последовательности нулей и единиц. Множество таких последовательностей мы называем Ω. Для всякого двоичного слова x можно рассмотреть множество всех бесконечных последовательностей, начинающихся на x. Это множество обозначается Ω_x. Естественным образом определяется конкатенация двоичного слова и бесконечной последовательности нулей и единиц. Рассматриваются также бесконечные последовательности букв произвольного алфавита A; их множество обозначается A^∞.

Иногда полезно рассматривать вместе конечные и бесконечные последовательности. Мы используем обозначение Σ для множества конечных и бесконечных последовательностей нулей и единиц (так что $\Sigma = \Xi \cup \Omega$). Через Σ_x мы обозначаем множество всех конечных и бесконечных продолжений (конечного) слова x.

Мы рассматриваем вычислимые функции, аргументами и значениями которых являются двоичные слова. Функции считаются частичными (не обязательно всюду определёнными), если иное не оговорено специально. Функция f является вычислимой, если существует машина (программа, алгоритм), которая останавливается на тех и только тех входах, где $f(x)$ определено, и выдаёт в качестве результата $f(x)$.

Вместо двоичных слов в качестве аргументов и значений можно использовать и другие конструктивные объекты (натуральные числа, целые числа, конечные множества слов, графы и так далее) — достаточно, чтобы их можно было закодировать двоичными словами и чтобы разные такие кодировки отличались вычислимыми функциями (по слову можно алгоритмически определить, является ли оно кодом какого-либо объекта в данной кодировке, а также найти код того же объекта в другой кодировке).

Можно говорить также и о вычислимости для других объектов (действительных чисел, мер), но это каждый раз требует особого определения.

Множество конструктивных объектов (двоичных слов, натуральных чисел и др.) называется перечислимым, если существует алгоритм, который печатает на выходе элементы этого множества и только их (с произвольными промежутками; алгоритм не обязан завершать работу, даже если множество конечно).

ЕЩЁ:

Полумеры - меры на конечных и бесконечных последовательностях, задаются полуаддитивными (субаддитивными?) функциями

перечислимость снизу - имеется в виду перечислимость снизу соответствующей функции

Колмогоровская сложность и основания теории вероятностей
[приложение к статье в УМН]

Литература

[1] Статья Ahlswede...Yeunga Network Information Flow где рассматривается задача передачи информации из одного источника и доказано, что условия на поток достаточны

[2] Bennett C. H., Gács P., Li M., Vitányi P. M. B., Zurek W., Thermodynamics of computation and information distance, *Proc. 25th ACM Symp. Theory Comput.*, p. 21–30, 1993.

[Статья о $K(A \Leftrightarrow B)$]

[3] Calude C. S., Staiger L. and Terwijn A. S., On partial randomness. *Annals of Pure and Applied Logic*, **138(1-3)**:20-30, 2006.

[Показывается эквивалентность размерности, выраженной в терминах s-мартингалов, и эффективно s-нулевых множеств. При этом определение эффективно s-нулевых множеств немного другое, позволяющее избежать неприятности.]

[4] Chaitin G. J., On the length of programs for computing binary sequences, *J. Assoc. Comput. Mach.*, 13:547–569, 1966.

[5] Chaitin G. J., On the length of programs for computing binary sequences: statictical considerations, *J. Assoc. Comput. Mach.*, 16:145–159, 1969.

[6] Chaitin G. J., A theory of program size formally identical to information theory, *J. Assoc. Comput. Mach.*, 22: 329–340, 1975.

[7] Статья, где Чэйтин приводит определение случайности по Соловею.

[8] Chernov A., Muchnik An. A., Romashchenko A., Shen A., Vereshchagin N. K. Upper semi-lattice of binary strings with the relation "x is simple conditional to y". *Theoretical Computer Science*, v. 271 (2002), issues 1–2, p. 69–95. Preliminary versions: DIMACS Tech. Report, 97-74 (December 1997); Proceedings of 1999 Computational Complexity conference, Atlanta.

[9] Верещагин Н. К., Скворцов Д. П., Скворцова Е. З., Чернов А. В., Варианты понятия реализуемости для пропозициональных формул, приводящие к логике слабого закона исключённого третьего. *Математическая логика и алгебра*, Труды Математического института им. В. А. Стеклова, 2003, т. 242, с. 77–97.

[10] Chernov A., Hutter M., Schmidhuber J., Algorithmic complexity bounds on future prediction errors, *Information and Computation*, 2007, vol. 205, pp. 242–261. DOI 10.1016/j.ic.2006.10.004

[11] Church A., On the concept of a random sequence. *Bull. Amer. Math. Soc*, 1940, v. 46, no. 2, p. 130–135.

[12] Статья R.P.Daley, где рассматриваются частичные правила. [Проверить, какая — в Ли Витаньи много разных]

[13] Downey R., Hirschfeldt D.R., Nies A., Terwijn S., Calibrating randomness., *The Bulletin of Symbolic Logic*, v. 12, no. 3, Sept. 2006, p. 411–491. [есть файл]

[14] Gorbunov K.Yu., On a complexity of the formula $A \vee B \Rightarrow C$, *Theoretical Computer Science*, v. 207, No. 2 (November 1998), p. 383–386. [доказывается, что сложность этой формулы может превышать максимум условных сложностей более чем на константу???]

[15] Халмош П. *Теория меры*, М.: ИЛ, 1953. 292 с.

[16] D. Hammer, A. Romashenko, A. Shen, N. Vereshchagin, Inequalities for Shannon entropies and Kolmogorov complexities. *Proceedings of CCC'97 Conference, Ulm*. Final version: Inequalities for Shannon entropy and Kolmogorov Complexity, *Journal of Computer and System Sciences*, v. 60, p. 442–464 (2000)

[17] Gács P., On the relation between descriptional complexity and algorithmic probability, FOCS 1981. Journal version: *Theoretical Computer Science*, 1983, v. 22, p. 71–93.

[18] Английский перевод: Gacs P., On the symmetry of algorithmic information. Soviet Math. Dokl, **15**:1477-1480, 1974.

[19] Gács P., Every sequence is reducible to a random one, *Inform. Contr.*, 70, no. 2–3, p. 186–192, 1986.

[20] Kleene S. C., On the interpretation of intuitionistic number theory. *Journal of Symbolic Logic*, 1945, v. 10, pp. 109–124.

[21] Kolmogoroff A., Zur Deutung der intuitionistishen Logik. *Mathematische Zeitschrift*, 1932, Bd. 35, H. 1, S. 58–65. (Русский перевод: К толкованию интуиционистской логики. В сборнике: Колмогоров А. Н., *Избранные труды. Математика и механика*, М.: Наука, 1985, с. 142–148.)

[22] Колмогоров А. Н., Фомин С. В. *Элементы теории функций и функционального анализа*, 4-е изд., М.: Наука, 1976. 544 с.

[23] Kolmogorov A. N., On tables of random numbers, *Sankhyā, The Indian Journal of Statistics, Ser. A*, 1963, v. 25, no. 4, p. 369–376. Reprinted in: *Theoretical Computer Science*, v. 207, No. 2 (November 1998), p. 387–395. (Русский перевод: О таблицах случайных чисел. *Семиотика и информатика*, 1982, вып. 18, с. 3–13, М.:ВИНИТИ. Перепечатано в: Колмогоров А. Н., *Теория информации и теория алгоритмов*, М.: Наука, 1987, с. 204–213.)

[24] Колмогоров А. Н., Три подхода к определению понятия «количество информации», *Проблемы передачи информации*, 1965, т. 1, вып. 1, с. 3–11 (Английский перевод: Kolmogorov A. N., Three approaches to the quantitative definition of information. *Problems Inform. Transmission*, 1(1):1–7, 1965.)

[25] Колмогоров А. Н., К логическим основам теории информации и теории вероятностей, *Проблемы передачи информации*, 1965, т. 5, вып. 3, с. 3–7.

[26] Kučera A., Measure, Π^0_1-classes and complete extensions of PA, p. 245–259 in: H.-D. Ebbinghaus, G. H. Müller and G. E. Sacks, eds., *Recursion Theory Week (Oberwolfach, 1984)*. Lect. Notes in Math., Vol. 1141, Springer-Verlag, Heidelberg, 1985.

[27] M. Kummer. On the complexity of random strings, *Symposium on Theoretical Aspects of Computer Science*, 1996. Lecture Notes in Computer Science, v. 1046, p. 25–36.

[28] van Lambalgen M., *Random Sequences*, Ph. D. Thesis, University of Amsterdam, 1987.

[29] Левин Л. А., О понятии случайной последовательности, *Доклады Академии наук СССР*, 1973, т. 212, № 3, с. 548–550. Levin L.A., On the notion of a random sequence, *Soviet Math. Dokl.*, **14**:1413–1416, 1973.

[30] Левин Л. А., Законы сохранения (невозрастания) информации и вопросы обоснования теории вероятностей, *Проблемы передачи информации*, 1974, т. 10, вып. 3, с. 30–35. Английский перевод: Levin L. A., Laws of information conservation (nongrowth) and aspects of the foundation of probability theory, *Problems of Information Transmission*, vol. 10 (1974), p. 206–210.

[31] Левин Л. А., О различных мерах сложности конечных объектов (аксиоматическое описание), *Доклады Академии наук СССР*, 1976, т. 227, № 4, с. 804–807. Levin L. A., Variuos measures of complexity for finite objects (axiomatic description), *Soviet Math. Dokl.*, 17:522–526, 1976.

[Статья, где объясняется, что сложности соответствуют ограничениям на перечислимые снизу функции]

[32] Английский перевод: Levin L. A., On the principle of conservation of information in intuitionistic mathematics, *Soviet Math. Dokl.*, **17** (1976), No. 2, 601–605

[33] Левин Л. А., Равномерные тесты случайности, *Доклады Академии наук СССР*, 1976, т. 227, № 1, с. 33-35. Levin L.A., Uniform tests of randomness, *Soviet Math. Dokl.*, **17**:337, 1976

[34] Левин Л. А., Об одном конкретном способе задания сложностных мер, *Доклады Академии наук СССР*, 1977, том 234, № 3, 536–539

[35] Levin L. A., Vyugin V. A., Invariant properties of informational bulks, *Mathematical Foundations of Computer Science, 1977*, Lecture Notes in Computer Science, v. 153.

[36] Levin L. A., Randomness Conservation Inequalities: Information and Independence in Mathematical Theories, *Information and Control*, **61**, No. 1–2, 15–37 (1984)

[37] Levin L. A., A concept of independence with application in various fields of mathematics, MIT Technical Report, MIT/LCS/TR-235, 21 p.

[38] Li M., Vitányi P., *An Introduction to Kolmogorov Complexity and Its Applications*, Second Edition, Springer, 1997. (638 pp.)

[39] Shoe-Yen Robert Li, Raymond W. Yeung, Linear Network Coding. Preprint, 1999.

статья, где достаточность условий на поток из одного источника доказывается с помощью случайного линейного кодирования где опубликовано???

Дальнейшее развитие: Ralf Koetter, Muriel Médard, An Algebraic Approach to Network Coding (исследуется случай нескольких источников и приёмников и линейных кодов, применяется алгебраическая геометрия). Препринт? Proceedings of INFOCOM, 2002.

Peter Sanders, Sebastian Enger, Ludo Tolhuizen, Polynomial Time Algorithms for Network Information Flow (полиномиальный алгоритм построения кодов для одновременной передачи информации из одного источника в несколько стоков, сравнение с алгоритмами без кодирования).

[40] Loveland D. W., A new interpretation of von Mises' concept of a random sequence, *Z. Math. Logik und Grundlagen Math.*, 12:279–294, 1966. [Статья Лавлэнда, где объясняется пример алгоритм LMS, может быть, тот же самый, что излагается под названием примера Вилля? упоминаются немонотонные правила, вроде бы объясняется, что они более общие. Даётся ссылка на неопубликованную работу Levin, Minsky, Silver 1962]

[41] Loveland D. W., The Kleene hierarchy classification of recursively random sequences, *Trans. Amer. Math. Soc.*, 125:497–510, 1966. [Статья Лавлэнда, где вводятся немонотонные правила (проверить!!!)]

[42] Lutz J., Dimension in Complexity Classes, *SIAM Journal on Computing*, v. 32 (2003), p. 1236–1259. Preliminary version: *Proceedings of the Fifteenth Annual IEEE Conference on Computational Complexity*, 2000, p. 158–169.

[Определяются s-gales (s-мартингалы). Доказывается представимость классической хаусдорфовой размерности в терминах s-мартингалов. Рассматриваются свойства размерностей, получаемых из вычислимых в разных сложностных классах s-мартингалов.]

[43] Lutz J. H., The dimensions of individual strings and sequences, *Information and Computation*, **187(1)**:49-79, 2003. Preliminary version: Gales and the constructive dimension of individual sequences. *Proceedings of the 27th International Colloquium on Automata, Languages, and Programming*, Springer-Verlag, 2000, p. 902–913.

[Эффективная размерность Хаусдорфа определяется через перечислимые снизу s-мартингалы. Строится оптимальный перечислимый снизу s-мартингал. Доказывается, что размерность множества равна точной верхней грани размерностей его элементов. Строится дискретный аналог хаусдорфовой размерности. В первоначальной версии доказано, что размерность последовательности заключена между liminf и limsup удельных сложностей; в окончательной — новый вариант доказательства и ссылка на [44].]

[44] Mayordomo, E., A Kolmogorov complexity characterization of constructive Hausdorff dimension, *Information Processing Letters*, 2002, **84** (1): p. 1–3.

[доказательство того, что эффективная хаусдорфова размерность последовательности есть нижний предел условных сложностей, со ссылкой на предварительную версию [43] про неравенство в одну сторону]

[45] Манин Ю. И. *Вычислимое и невычислимое*, М.: Советское радио, 1980. 128 с.

[46] Martin-Löf P., The Definition of Random Sequences, *Information and Control*, **9**:602–619, 1966.

[47] Martin-Löf P., Complexity Oscillations in Infinite Binary Sequences, *Z. Wahrscheinlichkeitstheorie verw. Geb.*, **19**, 225–230 (1971)

[48] Merkle W., The complexity of stochastic sequences, *Proceedings of the 18th Annual IEEE Conference on Computatil Complexity, 7–10 July 2003, Aarhus, Denmark*, p. 230–235. (Меркле называет Mises–Wald–Church stochastic то, что у нас называется случайными по Мизесу – Чёрчу – Дэли, recursively random — случайные относительно вычислимых мартингалов, partial-recursively random — случайные относительно частичных вычислимых мартингалов.)

Строятся последовательности сколь угодно малой сложности (при известной длине) случайные относительно вычислимых мартингалов, чуть более логарифмической сложности, случайные относительно частичных вычислимых мартингалов и доказано несуществование случайных по Мизесу – Чёрчу – Дэли последовательностей сложности $O(\log n)$

[49] Merkle W. The Kolmogorov–Loveland stochastic sequences are not closed under selecting subsequences, *Journal of Symbolic Logic*, v. 68 (2003), p. 1362–1376. Preliminary version: *International Colloquium on Automata, Languages and Programming, 2002*, Lecture Notes in Computer Science, v. 2380, p. 390–400, Springer, 2002.

[50] Miller J., Yu L., On initial segment complexity and degrees of randomness, *Transactions of the American Mathematical Society*, to appear.

[51] Miller J., Every 2-random real is Kolmogorov random, *Journal of Symbolic Logic*, **69**(3):907–913 (2004).

[52] Miller J., Contrasting plain and prefix-free Kolmogorov complexity, Submitted [23.08.2006].

[53] von Mises, Richard, Grundlagen der Wahrscheinlichkeitsrechnung, *Mathematische Zeitschrift*, Bd. 5, 191, S. 52–99. (Перепечатано в книге: Selected Papers of Richard von Mises. Volume Two. Probability and Statistics, General. American Mathematical Society, 1964. p. 57–106.)

[54] von Mises, Richard, *Wahrscheinlichkeit, Statistik und Wahrheit*, Wien: Springer-Verlag, 1928. 189 p.

[55] Мизес Р., *Вероятность и статистика*, М.–Л.:Госиздат, 1930.

[56] Мучник Ан. А. Об основных структурах дескриптивной теории алгоритмов. *Доклады Академии Наук СССР*, 1985, т. 285, № 2, с. 280–283.

[57] Мучник Ан, А., Нижние пределы частот в вычислимых последовательностях и релятивизованная априорная вероятность. *Теория вероятностей и её применения*, 1987, № 3, с. 563–565.

Статья, где объясняется, что нижние пределы частот в вычислимых последовательностях соответствуют $\mathbf{0}'$-перечислимым снизу полумерам.

[58] Muchnik An. A. Conditional complexity and codes, *Theoretical Computer Science*, v. 271 (2002), issues 1–2, p. 97–109. [Предварительная версия: Andrej Muchnik, Alexej Semenov, Multi-conditional Descriptions and Codes in Kolmogorov Complexity, ECCC Technical Report, 2000, no. 15, January 27, 2000.]

[59] Muchnik An. A., Positselsky S. E., Kolmogorov entropy in the context of computability theory, *Theoretical Computer Science*, v. 271 (2002), issues 1–2, p. 15–35.

[60] Muchnik An.A., On common information, *Theoretical Computer Science*, v. 207, No. 2 (November 1998), p. 319–328.

[61] Muchnik An.A., Semenov A.L., Uspensky V.A., Mathematical metaphysics of randomness, *Theoretical Computer Science*, v. 207, No. 2 (November 1998), p. 263–317.

[62] Nies A., Stephan F., Terwijn S., Randomness, relativization and Turing degrees, *Journal of Symbolic Logic*, **70** (2), 515–535 (2005).

[63] J. Reimann, *Computability and fractal dimension*, PhD thesis, Ruprecht-Karls Universität Heidelberg, 2004. `urn:nbn:de:bsz:16-opus-55430`

[Формулируется определение эффективной размерности через s-нулевые множества, доказывается существование максимального s-нулевого множества для рациональных s.]

[64] Ривест и др. Алгоритмы

[65] Romashchenko A., Shen A., Vereshchagin N., Combinatorial interpretaion of Kolmogorov complexity, *Theoretical Computer Science*, v. 271 (2002), issues 1–2, p. 111–123. Preliminary versions: ECCC Report 7(26):2000; IEEE conference on Computational Complexity.

[66] Schnorr C.P., *Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Begründung der Wahrscheinlichkeitstheorie*, Lecture notes in mathematics, v. 218. IV+212 S. Springer, 1971.

[67] Schnorr C.P, Optimal Gödel numberings. *Information Processing 71*. Proceedings of IFIP congress 71, Lubljana, August 23–28, 1971. V. 1. Foundations and systems. Eds.: C.V. Freiman et al. North-Holland, 1971, p. 56–58

[68] Schnorr C.P., Optimal enumerations and optimal Gödel numberings, *Mathematical Systems Theory*, v. 8, no. 2 (1975), p. 182–191.

[69] Schnorr C.P., Process Complexity and Effective Random Tests, *Journal of Computer and System Sciences*, **7**, p. 376-388 (1973).

[70] Шень А., Алгоритмические варианты понятия энтропии, ??? Shen A., Algoritmic variants of the notion of entropy, *Soviet Math. Dokl.*, 29(3):569–573, 1984.

[71] Шень А., О соотношениях между различными алгоритмическими определениями случайности, *Доклады АН СССР*, 1988, том 302, № 3, с. 548–552. Английский перевод: Shen A., On relations between different algorithmic definitions of randomness, *Soviet Math. Dokl.*, v. 38 (1989), no. 2, p. 316–319.

[72] Theory of computation textbook

[73] Solomonoff R. J., A formal theory of inductive inference, part 1, part 2. *Inform. Contr.*, 7:1–22, 224–254, 1964

[74] Tadaki, K., A generalization of Chaitin's halting probability Ω and halting self-similar sets, *Hokkaido mathematical journal*, **31** (1): 219–253, 2002. See also: arXiv:nlin.CD/0212001

[Определяются s-нулевые множества.]

[75] Успенский В. А., Семёнов А. Л. *Алгоритмы: основные идеи и приложения*, М.: Наука, 1987. 288 с. English translation: Uspensky V., Semenov A., *Algorithms: Main Ideas and Applications*. Translated by A. Shen. Kluwer Academic Publishers, 1993. ISBN: 0-7923-2210-X. 269 pp.

[76] Vereshchagin N. K. Kolmogorov complexity conditional to large integers. *Theoretical Computer Science*, v. 271 (2002), issues 1–2, p. 59–67.

[77] Верещагин Н. К., Шень А. Лекции по математической логике и теории алгоритмов. Часть 3. Вычислимые функции. М.:МЦНМО, 2000(?). (ftp://ftp.mccme.ru/users/shen/logic/comput)

[78] Shafer G., Vovk V. Probability and Finance: It's Only a Game! New York: Wiley, 2001.

[79] Ville J., *Étude Critique de la Notion de Collectif*, Gauthier-Villars, 1939. (Monographies des probabilités. Calcul des probabilites et ses applications. Publiée sous la direction de M. Émile Borel. Fascicule III.)

[80]

[81] V'yugin V.V., Ergodic theorems for individual random sequences, *Theoretical Computer Science*, v. 207, No. 2 (November 1998), p. 343–361.

[82] V'yugin V.V., Non-stochastic infinite and finite sequences, *Theoretical Computer Science*, v. 207, No. 2 (November 1998), p. 363–382. [Вроде бы строится машина, которая с положительной вероятностью выдаёт последовательность, не случайную ни по какой вычислимой мере. Нечто похожее есть также в V'yugin V.V. Bayesianism: An Algorithmic Analysis, *Information and Computation*, 1996, vol. 127, pp. 1–10. В чём разница?!]

[83] Wald A., Die Wiederspruchsfreiheit des Kollektivbegriffs der Wahrscheinlichkeitsrechnung, *Ergebnisse eines matematischen Kolloquiums*, 8:38–72, 1937.

[84] Заславский И. Д., Цейтин Г. С. О сингулярных покрытиях и связанных с ними свойствами конструктивных функций. Труды математического института АН СССР имени В. А. Стеклова, 1962. Том 67, с. 458–502.

[85] Звонкин А. К., Левин Л. А., Сложность конечных объектов и обоснование понятий информации и случайности с помощью теории алгоритмов. *Успехи математических наук*, т. 25, вып. 6 (156), с. 85–127. Английский перевод: A.K. Zvonkin, L.A. Levin, The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Math. Surveys*, **25**:6 (1970), p. 83–124.

(lutz-s-gales в библиографии уже есть)

Это вроде бы покрывает все или почти все, что в книжке написано про эффективную размерность. (В комментариях выше писал, как правило, про то, что в книжке есть. С полным содержанием всего этого я пока не совсем разобрался, надо еще почитать.) Ссылки, которые, видимо, нужно поставить:

- С определения s-нулевых множеств - определены K. Тадаки [74] - С теоремы 94 [effective-hausdorf] - [63] - С теоремы 96 [dimension-sup] - [43] - С теоремы 97 [hdim-formula] - [43,44] - С теоремы 152 [hausdorf-martingale] - [42] - Возможно, с того места, где написано про эквивалентность представлений о

этого жанра. Еще про s-мартингалы туда же.

Про s-мартингалы они еще все ссылаются на C.-P. Schnorr, Zufälligkeit und Wahrscheinlichkeit. Но здесь я только название переписать могу, больше ничего. Она, впрочем, в библиографии и lit-files и без меня есть.