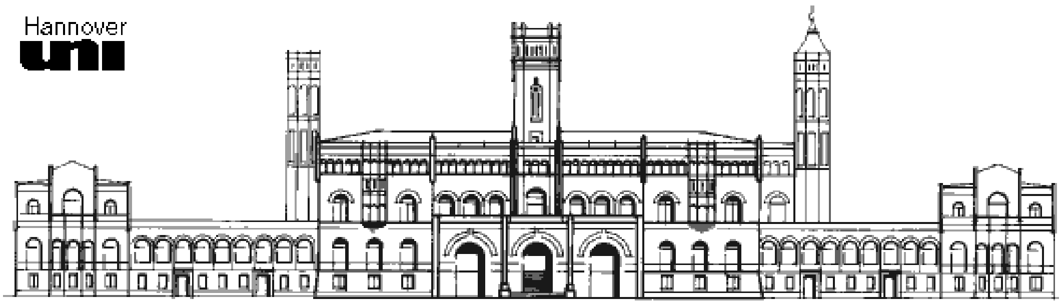


Vorlesungsmitschrift:
“Stochastik I”

Prof. Dr. R. Grübel
Institut für Mathematische Stochastik, Hannover

(SS 96)

24. November 1997

ge-TEX-ed von Sven Meyer
(email: `sm@stochastik.uni-hannover.de`)
(WWW: `www.stochastik.uni-hannover.de/~sm`)

Für diejenigen, die sich die Mühe machen, dieses Skript zu lesen und dabei auf Tippfehler stoßen: Bitte schickt mir, nachdem Ihr auf meiner Homepage nachgesehen und Euch vergewissert habt, daß dieser Fehler noch immer existiert, eine email, damit ich ihn berichtigen kann.

Mit bestem Dank, Sven.

Inhaltsverzeichnis

… und mit einem großen Dankeschön für die Mitarbeit an:

- Britta Kersten
- Luise Regier
- Astrid Vieweg

Kapitel 1

Grundbegriffe, die Axiome von Kolmogorov

1.1 Ergebnisraum und Ergebnisalgebra

Stochastik (=WTheorie + Statistik) beschäftigt sich mit Zufallsexperimenten, bei denen das Ergebnis nicht durch die Randbedingungen des Experiments festgelegt ist.
Der Ergebnisraum Ω ist eine Menge, die die möglichen Ergebnisse des Experiments enthält. Ereignisse werden durch Teilmengen von Ω beschrieben.

Beispiel 1.1

(i) (Würfelwurf) $\Omega = \{1,2,3,4,5,6\}$. Das Ereignis "eine gerade Zahl erscheint" ist $A = \{2,4,6\}$.

(ii) Eine Münze wird n-mal geworfen.
Schreibt man 0 für "Kopf" und 1 für "Zahl", so ist $\Omega = \{(i_1,\dots,i_n) : i_j \in \{0,1\}\ \forall j\} = \{0,1\}^n$ ein geeigneter Ergebnisraum. Hierbei bedeutet $w = (i_1,\dots,i_n)$, daß im j-ten Wurf i_j erscheint.

(iii) Eine Probe radioaktiven Materials emittiert Partikel einer bestimmten Sorte. Zählt man die Emissionen in einem bestimmten Zeitraum, so wird man auf $\Omega = \{0,1,2,\dots\} =: \mathbb{N}_0$ geführt. Die Menge $\{10,11,12,\dots\}$ steht für "mindestens 10 Partikel". Wartet man auf die erste Emission, so ist $\Omega = [0,\infty)$ ein geeigneter Ergebnisraum.

(iv) (Rotierender Zeiger)
$\Theta = 2\pi x, 0 \le x < 1$ sei der Winkel bei Stillstand. $\Omega = [0,1)$. Das Ereignis "Zeigerstop im oberen, rechten Quaranten" wird durch $(0,\frac{1}{4})$ bzw., bei Einschluß der Ränder, durch $[0,\frac{1}{4}]$ beschrieben.

Ein Ereignis A mit exakt einem Element, also $A = \{\omega\}$ mit einem $\omega \in \Omega$, nennt man Elementarereignis.

Kombinationen von Ereignissen können durch mengenalgebraische/-theoretische Operationen beschrieben werden:

$A \cap B$: A und B treten beide ein.
$A \cup B$: A oder B (oder beide) treten ein.
$\overline{\mathrm{A}}, A^c$: A tritt nicht ein.
Beim Würfelwurf: Es erscheint keine gerade Zahl wird beschrieben durch $\{2,4,6\}^c = \{1,3,5\}$

Beispiel 1.2 (Kombination von mehr als zwei Ereignissen)

(i) "Genau eines der Ereignisse A,B,C tritt ein": $A \cap B^c \cap C^c \ + \ A^c \cap B \cap C^c \ + \ A^c \cap B^c \cap C$
($A + B$ steht für $A \cup B$, wenn $A \cap B = \emptyset$)

(ii) Es sei $A_1, A_2, A_3, \dots$ eine Folge von Ereignissen. Dann wird das Ereignis "unendlich viele der A_i treten ein" repräsentiert durch $\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$ (=: $\limsup_{n\to\infty} A_n$, der limes superior der Mengenfolge)

Klar: $\bigcup_{m=n}^{\infty} A_m$ steht für "mindestens eines der Ereignisse mit Index $\ge n$ tritt ein", und es gibt
$\omega \in \limsup_{n\to\infty} A_n \iff \forall n \in \mathbb{N}\ \exists m \ge n : \ \omega \in A_m \iff \#\{n \in \mathbb{N} : \ \omega \in A_n\} = \infty.$

Die Menge der Ereignisse in einem Zufallsexperiment bilden ein Mengensystem $\mathfrak{A}$ ($\subseteq \mathbb{P}(\Omega)$, Potenzmenge von Ω).
Bei endlichen oder abzählbar unendlichen Ergebnisräumen können wir problemlos $\mathfrak{A} = \mathbb{P}(\Omega)$ voraussetzen, bei überabzählbarem Ω geht dies in vielen wichtigen Fällen nicht (wird später präzisiert.)
Die obigen Forderungen führen auf gewisse Mindestvoraussetzungen an $\mathfrak{A}$ und damit zu folgender Definition:

Definition 1.3 : $\mathfrak{A} \subseteq \mathbb{P}(\Omega)$ mit $\Omega \ne \emptyset$ heißt eine σ-Algebra (über Ω), wenn gilt:

(i) $\Omega \in \mathfrak{A}$
(ii) $A \in \mathfrak{A} \implies A^c \in \mathfrak{A}$
(iii) $A_1, A_2, \dots \in \mathfrak{A} \implies \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$

1.2 Wahrscheinlichkeit

Was ist "Wahrscheinlichkeit"? Strenggenommen keine mathematischen Frage.

Als "mathematischer Gegenstand" ist W. eine Funktion, die Ereignissen Zahlen zwischen 0 und 1 zuordnet und dabei gewissen "Axiomen" genügt; diese werden durch den "umgangssprachlichen" WBegriff motiviert.

Zur Erläuterung betrachten wir die Aussage "das Ereignis A hat die W. p" (z.B. "beim Wurf eines fairen Würfels erscheint mit W. $\frac{1}{2}$ eine gerade Zahl.")
Interpretation:

(a) Die Häufigkeitsauffassung (Frequentisten)
Es sei $N_n(A)$ die Häufigkeit des Auftretens von A bei n Wiederholungen des Zufallsexperiments; $\frac{1}{n}N_n(A)$ ist die relative Häufigkeit von A. Bei großem n wird man erwarten, daß die relative Häufigkeit von A in der Nähe von p liegt.

(b) Die Glaubensausfassung (Subjektivisten)
Der Wert p gibt auf einer Skala von 0 bis 1 die "Stärke meines Glaubens" an das Eintreten von A wieder. Dies kann über Wetten formalisiert werden und ist im Gegensatz zu (a) auch bei nicht wiederholbaren Experimenten anwendbar.

Für relative Häufigkeiten gelten die Regeln:
(i) $\frac{1}{n}N_n(\Omega) = 1$
(ii) $\frac{1}{n}N_n(A) \geq 0$
(iii) $\frac{1}{n}N_n(A_1 + \ldots + A_k) = \frac{1}{n}N_n(A_1) + \ldots + \frac{1}{n}N_n(A_k)$ für paarweise disjunkte $A_1, \ldots, A_k \in \mathfrak{A}$

Hier nun unser mathematisches Modell für Zufallsexperimente:

Definition 1.4 (Die Axiome von Kolmogorov)
Ein WRaum ist ein Tripel $(\Omega, \mathfrak{A}, P)$, bestehend aus einer nichtleeren Menge Ω (dem Ergebnisraum), einer σ-Algebra $\mathfrak{A}$ über Ω (dem Ereignissystem) und einer Abbildung $P : \mathfrak{A} \to \mathbb{R}$ mit

(i) $P(\Omega) = 1$
(ii) $P(A) \geq 0 \quad \forall A \in \mathfrak{A}$
(iii) $P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ für alle paarweise disjunkten $A_1, A_2, \ldots \in \mathfrak{A}$

Eine Abbildung mit diesen Eigenschaften heißt WMaß.
Eigenschaft (iii) nennt man $\sigma -$ Additivität.

Beispiel 1.5

(i) Ist Ω eine endliche Menge ($\neq \emptyset$), so wird durch $P(A) := \frac{\#A}{\#\Omega}\ \forall A \subseteq \Omega$ ein WMaß auf $(\Omega, \mathbb{P}(\Omega))$ definiert. Man nennt $(\Omega, \mathfrak{A}, P)$ mit $\mathfrak{A} = \mathbb{P}(\Omega)$ das Laplace $-$ Experiment über Ω. Diese werden häufig durch Symmetrieüberlegungen nahegelegt, wie in Beispiel 1.1 (i),(ii). Beim Wurf eines fairen (symmetrischen) Würfels ergibt sich damit als Wahrscheinlichkeit dafür, daß eine gerade Zahl geworfen wird $P(A) = \frac{\#\{2,4,6\}}{\#\{1,2,3,4,5,6\}} = \frac{1}{2}$ (Anzahl der "günstigen" Fälle dividiert durch die Anzahl der "möglichen" Fälle).

(ii) Ein Experiment, das deterministisch ist in dem Sinne, daß nur ein Ereignis ω_0 möglich ist, kann als degeneriertes Zufallsexperiment $(\Omega, \mathfrak{A}, \delta_{\omega_0})$ betrachtet werden. Hierbei ist Ω irgendeine Menge, die ω_0 enthält, $\mathfrak{A}$ eine σ-Algebra über Ω und δ_{ω_0} das Dirac$-$ oder Einpunktmaß in ω_0: $\delta_{\omega_0}(A) = \begin{cases} 1 \ , \omega_0 \in A \\ 0 \ , \omega_0 \notin A \end{cases}$ (man macht sich leicht klar, daß δ_{ω_0} in der Tat ein WMaß ist.)

Einige erste Folgerungen aus den Axiomen:

Satz 1.6 : Es sei $(\Omega, \mathfrak{A}, P)$ ein WRaum. Dann gilt:

(i) $P(\emptyset) = 0,\ P(A) \leq 1 \quad \forall A \in \mathfrak{A}$
(ii) $P(A^c) = 1 - P(A) \quad \forall A \in \mathfrak{A}$
(iii) (endliche Additivität) $P(A_1 \cup \ldots \cup A_k) = P(A_1) + \ldots + P(A_k)$ für alle paarweise disjunkten $A_1, \ldots, A_k \in \mathfrak{A}$
(iv) (Monotonie) $A, B \in \mathfrak{A}, A \subseteq B \implies P(A) \leq P(B)$.
(v) (Boole'sche Ungleichung) $P(A_1 \cup \ldots \cup A_k) \leq P(A_1) + \ldots + P(A_k)$ für alle (nicht notwendig disjunkten) $A_1, \ldots, A_k \in \mathfrak{A}$
(vi) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(vii) (Formel von (Sylvester-) Poincaré, Siebformel, Einschluß-Ausschluß-Formel)

$$P(A_1 \cup \ldots \cup A_k) = \sum_{\substack{H \subseteq \{1,\ldots,k\} \\ H \neq \emptyset}} (-1)^{\#H-1} P(\bigcap_{i \in H} A_i)$$

Beweis: $\emptyset \in \mathfrak{A}$ wegen (i) $\Omega \in \mathfrak{A}$
(ii) $A \in \mathfrak{A} \implies A^c \in \mathfrak{A}$
(iii) $\emptyset = \Omega^c$

Die Nachweise, daß die übrigen beteiligten Mengen in $\mathfrak{A}$ liegen, ist eine Übungsaufgabe.
(i) Verwendet man die σ-Additivität von P mit $A_1 = A_2 = A_3 = \ldots = \emptyset$, so folgt $P(\emptyset) = P(\emptyset) + P(\emptyset) + \ldots$

$\Longrightarrow P(\emptyset) = 0$
$P(A) \leq 1$ folgt aus $P(\Omega) = 1$ und der Monotonie (iv)
(iii) Setze $A_{k+1} = A_{k+2} = \ldots = \emptyset$, verwende die σ-Additivität und $P(\emptyset) = 0$.
(ii) $A \cup A^c = \Omega,\ A \cap A^c = \emptyset$, verwende nun (iii)
$1 = P(\Omega) = P(A + A^c) = P(A) + P(A^c)$
(iv) Es gilt $B = A + B \cap A^c$, also $P(B) = P(A) + P(B \cap A^c) \geq P(A)$, da $P(B \cap A^c) \geq 0$.
(v) Im Falle $k = 2$ folgt die Aussage aus (vi) und $P(A \cap B) \geq 0$.
Angenommen, die Aussage ist für $k \geq 2$ richtig. Dann folgt:
$P(\underbrace{A_1 \cup \ldots \cup A_k}_{\text{ein Ereignis}} \cup A_{k+1}) \leq P(A_1 \cup \ldots \cup A_k) + P(A_{k+1}) \leq P(A_1) + \ldots + P(A_k) + P(A_{k+1})$
(vi) $A = A \cap B\ +\ A \cap B^c,\ P(A) = P(A \cap B) + P(A \cap B^c)$
$A \cup B = B\ +\ A \cap B^c \Longrightarrow P(A \cup B) = P(B) + P(A \cap B^c) = P(B) + P(A) - P(A \cap B)$.
(vii) Im Fall $k = 2$ erhält man (v).
Induktionsschritt: Übungsaufgabe.

□

Einfache Anwendung: Mit welcher Wahrscheinlichkeit erscheinen beim n-fachen Wurf einer (fairen) Münze beide Seiten mindestens ein Mal?

Wir haben ein Laplace-Experiment über $\Omega = \{0, 1\}^n$
Bezeichnet A das interessante Ereignis, so gilt $A^c = \{(0, 0, \ldots, 0), (1, 1, \ldots, 1)\}$, also
$P(A) = 1 - P(A^c) = 1 - \frac{\# A^c}{\# \Omega} = 1 - \frac{2}{2^n} = 1 - \frac{1}{2^{n-1}}$

Warum σ-Additivität?
Warum nicht nur endliche Additivität?
Warum nicht sogar "Hyper-Additivität" $(P(\sum\limits_{i \in I} A_i) = \sum\limits_{i \in I} P(A_i)$ für bel. $I)$?

zu "endl. Additivität": Man erhält keine schöne (!) Theorie (Mehrdeutigkeiten)
zu "hyp. Additivität": zu stark (im Zeigerexperiment würde man $P(\Omega) = 0$ erhalten)
zu "σ-Additivität": kann als Stetigkeitseigenschaft aufgefaßt werden.

Satz 1.7 ("WMaße sind stetig von unten"):
Es sei $\Omega \neq \emptyset$, $\mathfrak{A}$ eine σ-Algebra auf Ω und $P : \mathfrak{A} \to \mathbb{R}$ eine Abbildung mit den Eigenschaften

(i) $P(\Omega) = 10$
(ii) $P(A) \geq 0 \quad \forall A \in \mathfrak{A}$
(iii) $P(A_1 + \ldots + A_n) = P(A_1) + \ldots + P(A_n)$ für alle paarweise disjunkten $A_1, \ldots A_n \in \mathfrak{A}$ und $n \in \mathbb{N}$.
(Induktion zeigt, daß dies äquivalent zu $P(A \cup B) = P(A) + P(B)\ \forall A, B \in \mathfrak{A}$ mit $A \cap B = \emptyset$ ist).

Dann sind äquivalent:

(a) P ist ein WMaß (also σ-additiv)
(b) P ist stetig von unten, d.h. für jede Folge $A_1, A_2, \ldots \in \mathfrak{A}$, die isoton ist im Sinne von $A_m \subseteq A_{m+1}\ \forall m \in \mathbb{N}$,
gilt: $\lim\limits_{n \to \infty} P(A_n) = P(\bigcup\limits_{n=1}^{\infty} A_n)$.

Beweis: (a)$\Longrightarrow$(b) Sei $(A_n)_{n \in \mathbb{N}}$ eine isotone Folge von Ereignissen.
Es sei $B_1 := A_1;\ B_n := A_n \cap A_{n-1}^c\ (\forall n > 1)$. Klar: $B_n \in \mathfrak{A}\ \forall n$, paarweise disjunkt, $B_1 + \ldots + B_n = A_n$.
Also: $P(\bigcup\limits_{n=1}^{\infty} A_n) = P(\sum\limits_{n=1}^{\infty} B_n) =^1 \sum\limits_{n=1}^{\infty} P(B_n) = \lim\limits_{n \to \infty} \sum\limits_{m=1}^{n} P(B_m) =^2 \lim\limits_{n \to \infty} P(\sum\limits_{m=1}^{n} B_m) = \lim\limits_{n \to \infty} P(A_n)$
^{1}: σ-Additivität.
^{2}: endliche Additivität.
(b)$\Longrightarrow$(a) Wir verwenden die umgekehrte Konstruktion:
Ist $(A_n)_{n \in \mathbb{N}}$ eine disjunkte Mengenfolge in $\mathfrak{A}$ und setzt man $B_n = \sum\limits_{m=1}^{n} A_m$, so gilt $B_n \uparrow \sum\limits_{m=1}^{\infty} A_m$,
$P(\sum\limits_{n=1}^{\infty} A_n) = P(\bigcup\limits_{n=1}^{\infty} B_n) =^3 \lim\limits_{n \to \infty} P(B_n) = \lim\limits_{n \to \infty} P(\sum\limits_{m=1}^{n} A_m) =^4 \lim\limits_{n \to \infty} \sum\limits_{m=1}^{n} P(A_m) = \sum\limits_{n=1}^{\infty} P(A_n)$
^{3}: Stetigkeit von unten.
^{4}: endliche Additivität.

□

Kapitel 2

Bedingte Wahrscheinlichkeiten und Unabhängigkeit

Es seien A, B Ereignisse in einem Zufallsexperiment, das durch einen WRaum $(\Omega, \mathfrak{A}, P)$ beschrieben wird. Was ist die Wahrscheinlichkeit von B unter der Bedingung, daß A eintritt? Bei n Wiederholungen tritt A $N_n(A)$-mal ein, unter diesen ist $N_n(A \cap B)$ die (absolute) Häufigkeit von B. Für die relative Häufigkeit von B unter den Experimenten, die A liefern, gilt: $\frac{N_n(A\cap B)}{N_n(A)} = \frac{\frac{1}{n}N_n(A\cap B)}{\frac{1}{n}N_n(A)}$.

Definition 2.1

Es sei A ein Ereignis mit $P(A) > 0$. Die bedingte Wahrscheinlichkeit eines Ereignisses B unter A wird definiert durch $P(B|A) := \frac{P(A\cap B)}{P(A)}$.
Man sieht leicht, daß $B \to P(B|A)$ ein WMaß ist. Der WRaum $(\Omega, \mathfrak{A}, P(\cdot|A))$ repräsentiert das gegenüber $(\Omega, \mathfrak{A}, P)$ dahingehend veränderte Experiment, daß das Eintreten von A bekannt ist.

Satz 2.2

(i) Die Multiplikationsregel
Es seien $A_1, \ldots, A_n$ Ereignisse mit $P(A_1 \cap \ldots \cap A_n) > 0$. Dann gilt:
$P(A_1 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P(A_n|A_1 \cap \ldots \cap A_{n-1})$.

(ii) Das Gesetz von der totalen Wahrscheinlichkeit
Es sei $A_1, \ldots, A_n$ eine Ereignispartition von Ω, d.h. $A_1, \ldots, A_n \in \mathfrak{A}$, $\bigcup_{i=1}^{n} A_i = \Omega$, $A_i \cap A_j = \emptyset$ für $i \neq j$.
Dann gilt für alle $B \in \mathfrak{A}$: $P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$. (Wir lassen hierbei $P(A_i) = 0$ zu und setzen dann $P(B|A_i)P(A_i) = 0$).

(iii) Die Formel von Bayes
Es seien $A_1, \ldots, A_n, B$ wie in (ii) und $P(B) > 0$. $P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{k=1}^{n} P(B|A_k)P(A_k)}$

Beweis: Verwende bei (ii) $B = \sum_{i=1}^{n} B \cap A_i$ und die Additivität von P.
Alles andere folgt unmittelbar aus den Definitionen. □

Beispiel 2.3

Ein bestimmter medizinischer Test ist zu 95% effektiv beim Erkennen einer bestimmten Krankheit, liefert allerdings bei 1% der gesunden Personen einen "falschen Alarm". Angenommen, 0.5% der Bevölkerung leiden unter dieser Krankheit - mit welcher Wahrscheinlichkeit hat jemand die Krankheit, wenn der Test dies behauptet?

Sei A : die getestete Person hat die Krankheit.
B : der Test zeigt das Vorliegen der Krankheit an.

Übersetzung der Annahmen:

$P(A) = 0.005$, $P(B|A) = 0.95$, $P(B|A^c) = 0.01$

Bayes: $P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B|A)P(A)+P(B|A^c)P(A^c)} = \frac{0.95\cdot 0.005}{0.95\cdot 0.005+0.01\cdot(1-0.005)} \approx 0.323$

Beachte: Die Übersetzung von % -Zahlen in Wahrscheinlichkeiten liegt bestimmten Annahmen über die Auswahl der Testpersonen, etc. zugrunde.
Beispiel 2.3 zeigt auch, daß es nicht immer nötig bzw. sinnvoll ist, $(\Omega, \mathfrak{A}, P)$ explizit anzugeben.

Kernbegriff der Stochastik: Unabhängigkeit
intuitiv: B ist von A unabhängig, wenn die Information, daß A eingetreten ist, die Wahrscheinlichkeit von B nicht verändert. Führt auf die Bedingung: $P(B|A) = P(B)$, besser:

Definition 2.4
Zwei Ereignisse A und B heißen stochastisch unabhängig, wenn gilt: $P(A \cap B) = P(A) \cdot P(B)$.

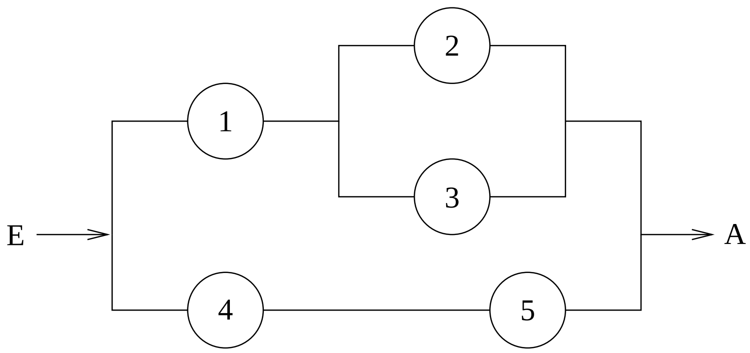
Bei mehr als zwei Ereignissen:
Definition 2.5
Eine Familie $\{A_i : i \in I\}$ von Ereignissen heißt paarweise unabhängig, wenn $P(A_i \cap A_j) = P(A_i)P(A_j)\ \forall i, j \in I,\ i \neq j$ gilt, sie heißen unabhängig, wenn für jede endliche Teilmenge H von I $P(\bigcap_{i \in H} A_i) = \prod_{i \in H} P(A_i)$ gilt.

Beispiel 2.6 (paarweise unabhängigig $\not\Rightarrow$ unabhängig)
Eine faire Münze wird zweimal geworfen ($\rightarrow$ Beispiel 1.1.(i)).
Laplace-Experiment $\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$

Es seien $A_1 := \{(0,0), (0,1)\}$	"Kopf im 1. Wurf"
$A_2 := \{(0,0), (1,0)\}$	"Kopf im 2. Wurf"
$A_3 := \{(0,1), (1,0)\}$	"Resultate verschieden"

Man sieht leicht (Die Durchschnitte sind jeweils einelementig):
$P(A_1 \cap A_2) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1) \cdot P(A_2)$
$P(A_1 \cap A_3) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1) \cdot P(A_3)$
$P(A_2 \cap A_3) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_2) \cdot P(A_3)$
Also $\{A_1, A_2, A_3\}$ ist paarweise unabhängig.
Es gilt $P(A_1 \cap A_2 \cap A_3) = 0 \neq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = P(A_1) \cdot P(A_2) \cdot P(A_3)$, die Familie ist also nicht unabhängig.

Beispiel 2.7
Typische Fragestellung der angewandten WRechnung: Funktionieren von Netzwerken. Wir betrachten einen einfachen Fall: Ein System bestehend aus 5 wie folgt angeordneten Komponenten:



Wir nehmen an, daß die Komponenten unabhängig voneinander mit W. p funktionieren. Das Gesamtsystem funktioniert, wenn es einen Pfad funktionierender Komponenten vom Eingang zum Ausgang gibt. Mit welcher Wahrscheinlichkeit funktioniert das gesamte System?

Sei A_i das Ereignis, daß Komponente i funktioniert.
B das interessierende Ereignis.

Es gilt $B = B_1 \cup B_2$ mit $B_1 = A_4 \cap A_5$ (unterer Pfad passierbar) und $B_2 = A_1 \cap (A_2 \cup A_3)$ (oberer Pfad ok)
Mit Hilfe der Unabhängigkeit und der Formel $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ folgt:
$P(B_1) = P(A_4) \cdot P(A_5) = p^2$
$P(B_2) = P(A_1 \cap A_2 \cup A_1 \cap A_3) = P(A_1 \cap A_2) + P(A_1 \cap A_3) - P(A_1 \cap A_2 \cap A_3) = 2p^2 - p^3$
$P(B_1 \cap B_2) = P(A_4 \cap A_5 \cap A_1 \cap A_2) + P(A_4 \cap A_5 \cap A_1 \cap A_3) - P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) = 2p^4 - p^5$,
also insgesamt: $P(B) = P(B_1) + P(B_2) - P(B_1 \cap B_2) = p^2(3 - p - 2p^2 + p^3)$.

Beispiel 2.8
Bei einer Game-Show befindet sich hinter einer von drei Türen ein Auto. Kandidat wählt eine Tür, Quizmaster öffnet eine der beiden übrigen bzw. die übrige Tür, hinter der kein Auto steht. Sollte der Kandidat wechseln?
Wir numerieren die Türen mit 1,2,3, wobei Tür 1 die ursprünglich vom Kandidaten gewählte bezeichnet.

Sei A_i: Auto hinter Tür i
B_i: Quizmaster wählt Tür i
C: Auto steht weder hinter der ursprünglich gewählten noch der Quizmaster-Tür.

klar: $C = A_2 \cap B_3 + A_3 \cap B_2$
$P(C) = P(A_2 \cap B_3) + P(A_3 \cap B_2) = P(B_3|A_2)P(A_2) + P(B_2|A_3)P(A_3)$.
Aus der Problemstellung geht hervor: $P(B_3|A_2) = P(B_2|A_3) = 1,\ P(A_i) = \frac{1}{3}$ $(i = 1, 2, 3)$, also $P(C) = \frac{2}{3}$.

Kapitel 3

Laplace-Experimente

3.1 Allgemeines

Ist Ω eine endliche, nichtleere Menge, so beschreibt $(\Omega, \mathfrak{A}, P)$ mit $\mathfrak{A} = \mathbb{P}(\Omega)$, $P(A) = \frac{\#A}{\#\Omega}$ $\forall A \subseteq \Omega$ das Laplace-Experiment über Ω. Es gibt nur endlich viele, mögliche Elementarereignisse, und diese haben alle dieselbe W. (Wir schreiben gelegentlich $P(\omega)$ anstelle $P(\{\omega\})$.)

Zufallsexperimente dieser Art tauchen auf:

- beim Werfen eines symmetrischen Gegenstandes (Münze, Würfel, ... : symmetrisch heißt, daß alle Seiten mit derselben W. oben landen)
- beim Mischen von Karten, allgemeiner: bei zufälligen Reihenfolgen (gut gemischt bzw. zufällige Reihenfolge heißt: alle möglichen Anordnungen haben dieselbe Wahrscheinlichkeit.)
- beim Entnehmen einer (zufälligen) Stichprobe aus einer (endlichen) Grundgesamtheit (zufällige Entnahme einer Stichprobe vom Umfang k aus einer Grundgesamtheit M von n Gegenständen/Personen o.ä. heißt: alle Teilmengen von M mit Umfang k haben dieselbe W.)

Wir betrachten nun "Kopplungen" von solchen Experimenten: Sind $(\Omega_i, \mathfrak{A}_i, P_i)$, $1 \leq i \leq n$, Laplace-Experimente, so können wir das Ausführen aller dieser Zufallsexperimente als ein neues Zufallsexperiment betrachten.
Geeigneter Ereignisraum: $\Omega := \mathop{\times}\limits_{i=1}^{n} \Omega_i = \Omega_1 \times \ldots \times \Omega_n = \{\omega = (\omega_1, \ldots, \omega_n) : \omega_i \in \Omega_i \text{ für } i = 1, \ldots, n\}$. Dies ist wieder eine endliche Menge, also sei $\mathfrak{A} := \mathbb{P}(\Omega)$.
Wie sieht ein geeignetes WMaß aus? Wenn die Experimente unabhängig voneinander ausgeführt werden, dann sollten für alle $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$ die Ereignisse $A_1, \ldots, A_n$, A_i : im i-ten Teilexperiment erscheint ω_i, im Sinne von §2 unabhängig sein, und natürlich sollte $P(A_i) = P_i(\{\omega_i\})$ gelten.
Mit $A_i = \Omega_1 \times \ldots \times \Omega_{i-1} \times \{\omega_i\} \times \Omega_{i+1} \times \ldots \times \Omega_n$ folgt nun

$$P(\{\omega\}) = P(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i) = \prod_{i=1}^{n} P_i(\{\omega_i\}) = \prod_{i=1}^{n} \frac{1}{\#\Omega_i} = \frac{1}{\#\Omega}.$$

Da dies nicht von ω abhängt, erhalten wird $P(A) = \frac{\#A}{\#\Omega}$ $\forall A \subseteq \Omega$, d.h. die unabhängige Kopplung (auch das Produkt genannt) der Laplace-Experimente ist wieder ein Laplace-Experiment. Bei dieser Konstruktion sind Ereignisse, die sich auf verschiedene Teilexperimente beziehen stochastisch unabhängig im Sinne von Def. 2.5.

Beispiel 3.1 Zwei Münzen werden (gleichzeitig) geworfen. Als Modell würde ein Laplace-Experiment über $\Omega = \{\omega_0, \omega_1, \omega_2\}$ mit:

ω_0	beide Münzen zeigen "Kopf"
ω_1	beide Münzen zeigen "Zahl"
ω_2	Ergebnisse verschieden

auf $P(\text{Ergebnisse verschieden}) = \frac{1}{3}$ führen.
Das widerspricht der Einführung! Das "korrekte" Modell ist das Laplace-Experiment über $\{0,1\}^2$, dies führt auf $P(\text{Ergebnisse verschieden}) = \frac{1}{2}$.

Wichtig: Bestimmung des korrekten Modells hängt von der "Außenwelt" ab, ist kein rein mathematisches Problem. Münzen sind "unterscheidbar", im Gegensatz zu manchen Objekten der Elementarteilchenphysik.

3.2 Etwas Kombinatorik

$A \times B = \{(a,b) : a \in A, b \in B\}$, $A^k = \underbrace{A \times \ldots \times A}_{k-\text{mal}}$.

Zwei elementare Grundregeln:

Regel 1: $\exists \varphi : A \longrightarrow B$ bijektiv $\Longrightarrow \#A = \#B$

Regel 2: Ist für jedes $x \in A$ B_x eine Menge mit n Elementen (insbesondere: alle B_x haben dieselbe Anzahl an Elementen), so gilt:$\#\{(x,y) : x \in A, y \in B_x\} = n \cdot \#A$.

Spezialfall: $\#(A \times B) = \#A \cdot \#B$.

Wir schreiben abkürzend $M_n = \{1, \dots, n\}$
Regel 2 liefert $\#{M_n}^k = \#\{(i_1, \dots, i_k) : 1 \leq i_j \leq n \ \forall j\} = n^k$
Die Elemente von ${M_n}^k$ werden auch k-Permutationen von M_n mit Wiederholungen genannt.

Interpretationen:
(i) Einer Menge von n Elementen kann man n^k Stichproben vom Umfang k mit Zurücklegen bei Berücksichtigung der Reihenfolge des Ziehens entnehmen. Das Element $(i_1, \dots, i_k)$ von ${M_n}^k$ steht hierbei für die Stichprobe, bei der im l-ten Zug das Element mit der Nummer i_l erscheint.
(ii) Es gibt n^k Möglichkeiten, k verschiedene Objekte auf n mögliche Plätze zu verteilen, wieder bei Berücksichtigung der Reihenfolge. Hierbei steht $(i_1, \dots, i_k) \in {M_n}^k$ für die Verteilung, bei der im l-ten Schritt das Objekt mit der Nummer l auf den Platz mit der Nummer i_l gelegt wird.

Satz 3.2 Für $1 \leq k \leq n$ gilt: $\#\{(i_1, \dots, i_k) \in {M_n}^k : i_l \neq i_j \text{ für } l \neq j\} = \frac{n!}{(n-k)!}$.
Beweis: Wende Regel 2 an: es gibt n Möglichkeiten für i_1.
Bei gegebenem i_1 bleiben $n-1$ Möglichkeiten für i_2.
Bei gegebenen i_1, i_2 bleiben $n-2$ Möglichkeiten für i_3, etc.
Die gesuchte Anzahl ist also $n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)$. □

Wichtiger Spezialfall mit $k = n$: es gibt genau $n!$ Permutationen einer Menge mit n Elementen. Die Elemente der Menge aus Satz 3.2 werden auch k-Permutationen von M_n ohne Wiederholung genannt.

Interpretation:
(i) Einer Menge von n Elementen kann man $\frac{n!}{(n-k)!}$ verschiedene Stichproben vom Umfang k ohne Zurücklegen bei Berücksichtigung der Reihenfolge entnehmen.
(ii) Es gibt $\frac{n!}{(n-k)!}$ verschiedene Möglichkeiten, k Objekte auf n Plätze so zu verteilen, daß keine Mehrfachbesetzungen vorkommen.

Satz 3.3 Für $1 \leq k \leq n$ gilt: $\#\{(i_1, \dots, i_k) \in {M_n}^k : i_1 < i_2 < \ldots < i_k\} = \binom{n}{k}$.
Beweis: Zu jedem Element dieser Menge gehören genau $k!$ Elemente der Menge aus Satz 3.2, nämlich alle die k-Tupel, die durch Permutationen der Koordinaten aus dem geordneten Tupel hervorgehen. □

Die Elemente der Menge aus Satz 3.3 werden auch k-Kombinationen von M_n ohne Wiederholung genannt.
Wichtige Anwendung: Eine Menge mit n Elementen hat $\binom{n}{k}$ Teilmengen mit (genau) k Elementen.

Interpretation:
(i) $\binom{n}{k}$ Möglichkeiten für Stichproben vom Umfang k ohne Zurücklegen ohne Berücksichtigung der Reihenfolge.
(ii) $\binom{n}{k}$ Möglichkeiten, k Objekte ohne Mehrfachbesetzung auf n Plätze ohne Berücksichtigung der Reihenfolge zu verteilen.

Satz 3.4 Es gibt $\#\{(i_1, \dots, i_k) \in {M_n}^k : i_1 \leq \ldots \leq i_k\} = \binom{n+k-1}{k}$ k-Kombinationen mit Wiederholungen.
Beweis: Wir definieren eine bijektive Abbildung

$$\varphi : \begin{cases} \{(i_1, \dots, i_k) \in {M_n}^k : i_1 \leq \ldots \leq i_k\} & \longrightarrow & \{(i_1, \dots, i_k) \in M_{n+k-1}^k : i_1 < \ldots < i_k\} \\ \varphi((i_1, \dots, i_k)) & = & (i_1, i_2+1, i_3+2, \dots, i_k+k-1) \end{cases}$$

Verwende nun Regel 1 in Satz 3.3. □

Interpretation:
(i) Einer Menge von n Elementen kann man $\binom{n+k-1}{k}$ verschiedene Stichproben vom Umfang k entnehmen, wenn zurückgelegt wird und die Ziehungsreihenfolge nicht berücksichtigt wird.
(ii) Es gibt $\binom{n+k-1}{k}$ Möglichkeiten, k Objekte mit möglicher Mehrfachbesetzung auf n Plätze zu verteilen, wenn die Verteilungsreihenfolge nicht berücksichtigt wird.

Anwendung: Es gibt $\binom{n+k-1}{k}$ Möglichkeiten, eine natürliche Zahl k als Summe von n nichtnegativen, ganzen Zahlen zu schreiben.

3.3 Lösung einiger typischer Probleme

3.3.1 Paradox von de Méré

(i) Ein Würfel wird 4mal geworfen, sei A das Ereignis, daß mindestens eine 6 erscheint.
(ii) Zwei Würfel werden 24mal geworfen, sei B das Ereignis, daß mindestens eine Doppel-6 erscheint.

Da in (i) das Einzelereignis ein Sechstel der W. hat, und in (ii) das Sechsfache an Wiederholungen ausgeführt wird, wurde vermutet, daß A und B dieselbe W. haben.

(i) Sei A_i das Ereignis, daß im i-ten Wurf 6 erscheint.
$A_1, \dots, A_4$ sind unabhängig und haben W. $\frac{1}{6}$, also:
$P(A) = 1 - P(A^c) = 1 - P(A_1^c \cap \dots \cap A_4^c) = 1 - P(A_1^c) \cdot \ldots \cdot P(A_4^c)$
$= 1 - (1 - P(A_1))(1 - P(A_2))(1 - P(A_3))(1 - P(A_4)) = 1 - (1 - \frac{1}{6})^4 \approx 0.518\dots$

(ii) Sei B_i das Ereignis, daß im i-ten Wurf das Paar (6,6) erscheint. $B_1, \dots, B_{24}$ sind unabhängig und haben Wahrscheinlichkeit $\frac{1}{36}$:
$P(B) = 1 - P(B^c) = 1 - P(B_1^c \cap \dots \cap B_{24}^c) = 1 - (1 - \frac{1}{36})^{24} \approx 0.491\dots$

3.3.2 Geburtstagsproblem

In einem Raum befinden sich n Personen. Mit welchen W. haben mindestens 2 der Personen am gleichen Tag Geburtstag? ($n \leq 365$)

Vereinfachung der Annahme: der 29. Februar wird vernachlässigt, ebenso Zwillinge, etc., saisonale Schwankungen in der Geburtenrate bleiben unbeachtet. Dann ist ein Laplace-Experiment über $\Omega := \{(i_1, \dots, i_n) : 1 \leq i_1, \dots, i_n \leq 365\} = \{1, \dots, 365\}^n$ plausibel, wobei $i_j = k$ bedeutet, daß Person j am Tag k Geburtstag hat.
$A := \{(i_1, \dots, i_n) \in \Omega : \exists l \neq j \text{ mit } i_l = i_j\}$.
Man hat $A^c = \{(i_1, \dots i_n) \in \Omega : i_l \neq i_j \text{ für } l \neq j\}$. Mit den Formeln aus §3.2 folgt also
$P(A) = 1 - P(A^c) = 1 - \frac{\#A^c}{\#\Omega} = 1 - \frac{\frac{365!}{(365-n)!}}{365^n}$
Dies ist eine in n (streng) steigende Folge: ab $n = 23$ gilt $P(A) \geq 0.5$, bei $n = 50$ gilt $P(A) \approx 0.97\dots$

3.3.3 Bridge

Beim Kartenspiel Bridge werden 52 Karten an 4 Spieler (Norden, Süden, Osten, Westen) verteilt. Wir wollen die W. der Ereignisse

- A einer der Spieler erhält alle 4 Asse
- B jeder der Spieler erhält ein As

bestimmen.

Das Mischen der Karten liefert eine zufällige Reihenfolge.
$\Omega' := \{(\omega_1, \dots, \omega_{52}) \in \{1, \dots, 52\}^{52} : \omega_i \neq \omega_j \text{ für } i \neq j\}$, also Menge der Permutationen von $\{1, \dots, 52\}$. Hierbei werden die Karten mit 1, ... ,52 durchnumeriert und $\omega_k = j$ bedeutet, daß die k-te Karte im Stapel die Nummer j hat. Alle Elementarereignisse haben dieselbe Wahrscheinlichkeit $\frac{1}{52!}$ (Def. von "gut gemischt")
Die Ereignisse hängen nicht von der Reihenfolge ab, mit der die Karten bei den Spielern ankommen; man kann also mit $\Omega := \{(D_1, D_2, D_3, D_4) : D_i \subseteq \{1, \dots, 52\} \; \forall i, \#D_i = 13, \text{ disjunkt}\}$ arbeiten. Hierbei ist D_i die Menge der Karten für Spieler i. Die Austeilreihenfolge definiert eine Abbildung von Ω' in Ω, die jeweils $(13!)^4$ Elemente von Ω' auf genau 1 Element von Ω abbildet (alle 13! Permutationen der an Spieler 1 ausgegebenen Karten liefern dieselbe Menge D_1, etc.) Betrachten wir also als Resultat des Zufallsexperiments das Vierertupel der "Hände", so liegt noch stets ein Laplace-Experiment vor (es werden jeweils gleichviele Elemente von Ω' zu einem Element von Ω zusammengefaßt). Hieraus folgt auch:
$\#\Omega = \frac{\#\Omega'}{(13!)^4} = \frac{52!}{13!13!13!13!}$

Alternatives Argument: D_1 ist eine Teilmenge vom Umfang 13 einer Menge mit 52 Elementen. Es gibt also $\binom{52}{13}$ Möglichkeiten für D_1. D_2 ist eine Teilmenge vom Umfang 13 von $\{1, \dots, 52\} \setminus D_1$, die 52–13=39 Elemente hat. Ist also D_1 festgelegt, so bleiben $\binom{39}{13}$ Möglichkeiten für D_2. Für D_3 bleiben $\binom{26}{13}$ Möglichkeiten, Spieler 4 erhält den Rest ($\binom{13}{13} = 1$ Möglichkeit).

Anwendung von Regel 2 aus §3.1 führt auf:
$\#\Omega = \binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13} = \frac{52!}{13!13!13!13!}$

Sei nun A_i das Ereignis, daß Spieler i alle vier Asse erhält (mit den Zahlen 1, ... ,4).
$A_1 = \{(D_1, D_2, D_3, D_4) \in \Omega : D_1 \supseteq \{1, 2, 3, 4\}\}$
Für $D_1' := D_1 \cap \{1, 2, 3, 4\}^c$ bleiben $\binom{48}{9}$ Möglichkeiten (9 Karten aus der Menge der 48 "Nicht–Asse"). Die Anzahl der Möglichkeiten für D_2, D_3, D_4 bleiben unverändert, also gilt: $P(A_1) = \frac{\#A_1}{\#\Omega} = \frac{\binom{48}{9}\binom{39}{13}\binom{26}{13}}{\binom{52}{13}\binom{39}{13}\binom{26}{13}} = \frac{13 \cdot 12 \cdot 11 \cdot 11}{52 \cdot 51 \cdot 50 \cdot 49}$.
Dieselben Argumente funktionieren bei A_2, A_3, A_4 und liefern dasselbe Ergebnis, also gilt:
$P(A) = P(A_1) + \ldots + P(A_4) = 4P(A_1) \approx 0.01056\dots$

Behandlung von B ganz analog, abgekürzt:
Es gibt 4! Möglichkeiten, die Asse so an die Spieler zu verteilen, daß jeder genau ein As erhält. Sind die Asse verteilt, so bleiben $\binom{48}{12}\binom{36}{12}\binom{24}{12}\binom{12}{12} = \frac{48!}{(12!)^4}$ Möglichkeiten für die übrigen Karten. Damit ist

$P(B) = \frac{\#B}{\#\Omega} = \frac{4!\frac{48!}{(12!)^4}}{\frac{52!}{(13!)^4}} = \frac{4! \cdot 13^4}{52 \cdot 51 \cdot 50 \cdot 49} \approx 0.1055\dots$

3.3.4 Der zerstreute Postbote

Ein Postbote verteilt n Briefe "zufällig" auf n Briefkästen (einer pro Kasten); Wir nehmen an, daß zu jeder der n Adressen genau einer der n Briefe gehört. Wieviele Personen erhalten den für sie bestimmten Brief?

Wir numerieren Briefe und Briefkästen so, daß Brief i in Kasten i gehört, $1 \leq i \leq n$. Die möglichen Austeilungen entsprechen dann den Permutationen von $\{1, \ldots, n\}$, "zufällig" soll heißen, daß ein Laplace-Experiment über $\Omega_n := \{(\omega_1, \ldots, \omega_n) \in \{1, \ldots n\}^n : \omega_i \neq \omega_j$ für $i \neq j\}$ vorliegt. Sei zunächst $A_n = \{\omega \in \Omega_n : \omega_i \neq i \; \forall i = 1, \ldots, n\}$, die Menge der fixpunktefreien Permutationen, $B_{ni} = \{\omega \in \Omega_n : \omega_i = i\}, \; 1 \leq i \leq n$.

Offensichtlich gilt $A_n{}^c = \bigcup_{i=1}^{n} B_{ni}$, also folgt nach Poincaré:

$$P_n(A_n) = 1 - P_n(\bigcup_{i=1}^{n} B_{ni}) = 1 - \sum_{\substack{J \subseteq \{1, \ldots, n\} \\ J \neq \emptyset}} (-1)^{\#J-1} P_n(\bigcap_{i \in J} B_{ni})$$

Es gilt: $\bigcap_{i \in J} B_{ni} = \{\omega \in \Omega_n : \omega_i = i \; \forall i \in J\}$

Für ein ω aus diesem Durchschnitt sind $\#J$ Positionen festgeleht; die übrigen $n - \#J$ Positionen können beliebig permutiert werden, also: $\# \bigcap_{i \in J} B_{ni} = (n - \#J)!$

Es gibt $\binom{n}{k}$ Teilmengen J mit k Elementen, also:

$$P_n(A_n) = 1 - \sum_{\substack{J \subseteq \{1, \ldots n\} \\ J \neq \emptyset}} (-1)^{\#J-1} \tfrac{(n-\#J)!}{n!} = 1 - \sum_{k=1}^{n} \tbinom{n}{k} (-1)^{k-1} \tfrac{(n-k)!}{n!} = 1 + \sum_{k=1}^{n} (-1)^k \tfrac{1}{k!} = \sum_{k=0}^{n} (-1)^k \tfrac{1}{k!}$$

Aus Analysis ist $\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x$ bekannt, also folgt: für großes n ist die Wahrscheinlichkeit dafür, daß kein Brief beim richtigen Empfänger landet, ungefähr $e^{-1} \approx 0.3679 \ldots$.

Wie haben hier ein erstes Limesresultat. Da hier eine alternierende Reihe vorliegt, kann man sogar eine Fehlerabschätzung angeben: $\left| P_n(A_n) - e^{-1} \right| \leq \frac{1}{(n+1)!}$.

Gleichzeitig haben wir eine Aussage bewiesen, die nicht auf W. Bezug nimmt: Die Anzahl der fixpunktefreien Permutationen einer Menge von n Elementen ist $n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}$

Sei nun $X_n : \Omega \longrightarrow \mathbb{N}_0, \; X_n(\omega) := \#\{i : \omega_i = i\}$ die Anzahl der Fixpunkte von Ω (im Anwendungsbeispiel: die Anzahl der Briefe, die beim richtigen Empfänger landen.)

Es gilt: $X_n(\omega) = k \iff \exists \; J \subseteq \{1, \ldots, n\}, \; \#J = k : \omega_i \begin{cases} = i & , \quad i \in J \\ \neq i & , \quad i \notin J \end{cases}$

Die Anzahl aller $\omega \in \Omega_n$ mit $X_n(\omega) = k$ ist also das Produkt aus der Anzahl der Möglichkeiten für J und der Anzahl der fixpunktefreien Permutationen der Menge $\{1, \ldots, n\} \setminus J$ (Regel 2).

$$P_n(X = k) = \frac{\#\{\omega \in \Omega_n : X(\omega) = k\}}{\#\Omega_n} = \frac{1}{n!} \binom{n}{k} (n-k)! \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \quad \longrightarrow \quad \frac{1}{k!} e^{-1} \; \forall k \in \mathbb{N}_0$$

Kapitel 4

Diskrete WRäume und Zufallsgrößen

4.1 Allgemeines

Wir nennen $(\Omega, \mathfrak{A}, P)$ einen diskreten WRaum, wenn Ω eine endliche oder abzählbar unendliche Menge ist (und $\mathfrak{A} := \mathbb{P}(\Omega)$ gilt). Aufgrund der σ-Additivität ist P dann durch die zug. WMassenfunktion p, $p : \Omega \to \mathbb{R}$, $p(\omega) = P(\{\omega\})$ eindeutig festgelegt: $P(A) = \sum\limits_{\omega \in A} p(w)\ \forall A \in \mathfrak{A}$.

Verallgemeinerte Laplace-Experimente $(p(\omega) = \frac{1}{\#\Omega})$
Wie bei Laplace-Experimenten lassen sich Produkte von (endlich vielen) diskreten WRäumen bilden, diese ergeben wieder diskrete WRäume.

Im Beispiel 3.1 erhalten wir Ereignisse mit unterschiedlichen Wahrscheinlichkeiten durch Zusammenfassen von unterschiedlich vielen Ergebnissen bei einem Laplace-Experiment.

Definition 4.1 Es sei $(\Omega, \mathfrak{A}, P)$ ein diskreter WRaum und S eine nichtleere Menge. Dann heißt eine Abbildung $X : \Omega \to S$ eine (S-wertige) diskrete Zufallsgröße, Zufallsvariable (ZV) bei $S = \mathbb{R}$, Zufallsvektor bei $S = \mathbb{R}^k$ $(k > 1)$.

Mit Ω ist auch $X(\Omega)$ zufällig. Es wird bei der Behandlung von Zufallsgrößen nicht darum gehen (können), welchen Wert X annimmt, sondern darum, mit welcher W. X in einer Teilmenge von S liegt.

Satz und Definition 4.2 Es seien $(\Omega, \mathfrak{A}, P)$ ein diskreter WRaum und $X : \Omega \to S$ eine diskrete Zufallsgröße. Dann wird durch $P^X : \mathbb{P}(S) \to \mathbb{R}$, $P^X(A) := P(X^{-1}(A))\ \forall A \in \mathbb{P}(S)$ ein WMaß definiert.
Hierbei: $X^{-1}(A) := \{\omega \in \Omega : X(\omega) \in A\}$. Für $P(X^{-1}(A))$ schreiben wir auch $P(X \in A)$. P^X heißt die Verteilung von X, alternative Schreibweise: $\mathfrak{L}(X)$ ("law")

Beweis: (i) $P^X(S) = P(\{\omega \in \Omega : X(\omega) \in S\}) = P(\Omega) = 1$
$P^X(A) = P(\{\omega \in \Omega : X(\omega) \in A\}) \geq 0$
(ii) Sind $A_1, A_2, \ldots \subseteq S$ paarweise disjunkt, so sind auch $B_1, B_2, \ldots \subseteq \Omega$ mit $B_i := \{\omega \in \Omega : X(\omega) \in A_i\}$ paarweise disjunkt, also liefert die σ-Additivität von P:
$P^X(\sum\limits_{i=1}^{\infty} A_i) = P(X^{-1}(\sum\limits_{i=1}^{\infty} A_i)) = P(\sum\limits_{i=1}^{\infty} X^{-1}(A_i)) = \sum\limits_{i=1}^{\infty} P(X^{-1}(A_i)) = \sum\limits_{i=1}^{\infty} P^X(A_i)$.
Dies zeigt, daß P^X σ-additiv ist. $\square$

Beispiel 4.3 Wie oft erscheint "Zahl" beim 5-maligen Wurf einer (fairen) Münze?
Das Ausgangsexperiment ist ein Laplace-Experiment über $\Omega = \{0,1\}^5$ (0=Kopf, 1=Zahl), die Anzahl der Zahlwürfe ist $X(\omega) = \omega_1 + \omega_2 + \ldots + \omega_5$, $\omega = (\omega_1, \ldots, \omega_5) \in \Omega$
Als Bildbereich kommt beispielsweise $S = \{0, 1, \ldots 5\}$ in Frage. Als WMaß auf einer endlichen Menge wird $\mathfrak{L}(X)$ wieder durch die zugehörige Massenfunktion $p_X : S \to \mathbb{R}$, $p_X(k) = P(X = k)$ beschrieben, $k = 0, \ldots, 5$.
Aus §3 ist bekannt: $P(\{\omega \in \Omega : X(\omega) = k\}) = \frac{\#\{\omega \in \Omega : \sum_{i=1}^{5} \omega_i = k\}}{\#\Omega} = \frac{\binom{5}{k}}{32} =: p_X(k)$ für $k = 0, 1, \ldots, 5$, denn es gibt $\binom{5}{k}$ Möglichkeiten, die k Einswerte auf die 5 zugehörigen Positionen zu verteilen. p_X heißt Wahrscheinlichkeitsmassenfunktion zur Zufallsgröße X.

4.2 Einige wichtige Verteilungen

4.2.1 Binomial- und Bernoulli-Verteilung

Eine diskrete ZV X heißt binomialverteilt mit den Parametern n und p, kurz: $\mathfrak{L}(X) = Bin(n,p)$ oder $X \sim Bin(n,p)$, wobei $n \in \mathbb{N}$, $p \in [0,1]$, wenn $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, \ldots, n$ gilt.

Die ZV X aus Beispiel 4.3 ist $Bin(5, \frac{1}{2})$-verteilt.

Auch die ZV $Y = 5 - X$, also Anzahl der Kopf-Würfe, hat diese Verteilung.
Dies ist ein Beispiel dafür, daß durchaus verschiedene Zufallsvariablen dieselben Verteilungen haben können.

Binomialverteilungen tauchen bei "Erfolgsanzahl bei unabhängigen Versuchswiederholungen" auf, wenn "Erfolg" das Eintreten eines bestimmten Ereignisses A in einem Einzelexperiment bezeichnet. Hierbei ist n die Anzahl der Wiederholungen und p die W. von A.
Begründung: jedes Ereignis $\underbrace{A \times A \times A^c \times \ldots \times A^c \times A}_{k\text{-mal } A,\ (n-k)\text{-mal } A^c}$ im Produktexperiment, das auf $X = k$ führt, hat wegen der Unabhängigkeit die Wahrscheinlichkeit $p^k(1-p)^{n-k}$ und es gibt $\binom{n}{k}$ Möglichkeiten, die k A-Faktoren auf die n Positionen zu verteilen.
Im Falle $n = 1$ spricht man auch von Bernoulli-Verteilung; dann gilt: $P(X \in \{0,1\}) = 1$.

4.2.2 Poisson-Verteilung

X heißt poisson-verteilt mit Parameter $\lambda > 0$, wenn $P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$ für alle $k \in \mathbb{N}$ gilt.

Diese Verteilung spielt eine wichtige Rolle als Grenzverteilung; mit $\lambda = 1$ tritt sie beispielsweise bei der Anzahl der Fixpunkte einer zufälligen Permutation als Grenzverteilung auf ($\rightarrow$ 3.3.4)

Die Poisson-Verteilung kann als Approximation für $Bin(n,p)$ bei großem n und kleinem p verwendet werden:
Satz 4.4 Ist $(p_n)_{n\in\mathbb{N}}$ eine Nullfolge auf $[0,1]$ mit $\lim\limits_{n\to\infty} np_n = \lambda \in (0,\infty)$, so gilt für alle $k \in \mathbb{N}_0$

$$\lim_{n\to\infty} \binom{n}{k} {p_n}^k (1-p_n)^{n-k} = e^{-\lambda} \cdot \frac{\lambda^k}{k!}.$$

Beweis: $\binom{n}{k}{p_n}^k(1-p_n)^{n-k} = \underbrace{\frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{n^k}}_{\longrightarrow 1} \underbrace{\frac{(np_n)^k}{k!}}_{\longrightarrow \frac{\lambda^k}{k!}} \underbrace{(1 - \frac{\overbrace{np_n}^{\longrightarrow\lambda}}{n})^{n-k}}_{\longrightarrow e^{-\lambda}}$ für festes k □

In Worten: Bei einer großen Anzahl Wiederholungen n mit kleiner Erfolgsw. p ist die Anzahl X der Erfolge näherungsweise poissonverteilt mit Parameter $\lambda = np$.

Diese Verteilung taucht also häufig im Zusammenhang mit seltenen Ereignissen auf (Anzahl der Druckfehler pro Seite, Anzahl der emittierten Partikel pro Zeiteinheit bei radioaktivem Material, etc.)

4.2.3 Geometrische Verteilung und negative Binomialverteilung

Angenommen, wir werfen einen (symmetrischen) Würfel solange, bis eine Sechs erscheint. X sei die hierfür notwendige Anzahl der Würfe (einschließlich des Wurfes, der die '6' liefert.) Es gilt $X = n$ ($n \in \mathbb{N}$) genau dann, wenn die ersten $n-1$ Versuche keine '6' ergeben und im n-ten Versuch die '6' erscheint. Aufgrund der Unabhängigkeit der Würfe hat dieses Ereignis die W. $\underbrace{(1-\frac{1}{6}) \cdot (1-\frac{1}{6}) \cdot \ldots \cdot (1-\frac{1}{6})}_{n-1 \text{ mal}} \cdot \frac{1}{6}$.
Allgemeiner, wenn X nur Werte aus $\mathbb{N}$ annimmt und $P(X = n) = (1-p)^{n-1}p$ für alle $n \in \mathbb{N}$ gilt, dann heißt X geometrisch verteilt mit Parameter $p \in (0,1)$.

Diese Verteilung tritt also bei der Anzahl der Versuche auf, wenn man ein Zufallsexperiment so oft wiederholt, bis ein bestimmtes Ereignis, das W. p hat, eingetreten ist.

Verallgemeinerung:
Warten auf das r-te Eintreten des Ereignisses führt auf $P(X = k) = \binom{k-1}{r-1}p^r(1-p)^{k-r}$ für $k \in \mathbb{N},\ k \geq r$.
Man nennt diese Verteilung die negative Binomialverteilung mit Parametern p und r, $r \in \mathbb{N},\ 0 < p < 1$.[1]

4.2.4 Hypergeometrische Verteilung

Eine Urne enthalte N Kugeln, hiervon M weiße und $N-M$ schwarze. Es werden n Kugeln entnommen, ohne Zurücklegen ($n, M \leq N$). Es sei X die Anzahl der weißen Kugeln in der Stichprobe. Dann gilt: $P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$, denn es gibt $\binom{M}{k}$ Möglichkeiten für die weißen und $\binom{N-M}{n-k}$ Möglichkeiten für die schwarzen Kugeln in der Stichprobe, alle $\binom{N}{n}$ Stichproben werden als gleichwahrscheinlich vorausgesetzt.
Man nennt diese Verteilung die hypergeometrische Verteilung mit den Parametern n, N und M.

Beispielsweise ist in der in Absatz 3.3.3 beschriebenen Situation die Anzahl der Asse, die "Nord" erhält, hypergeometrisch verteilt mit den Parametern 13, 52 und 4.

[1] In der Literatur wird häufig stattdessen die Anzahl der Mißerfolge, also $Y = X - r$, betrachtet.

Anderes Beispiel: Die Wahrscheinlichkeit für k Richtige im Zahlenlotto '6 aus 49' ist $\frac{\binom{6}{k}\binom{43}{6-k}}{\binom{49}{6}}$, $k = 0, 1, \ldots, 6$ (wir ignorieren Zusatzzahlen, etc.), also hypergeometrisch verteilt mit Parametern 6, 49 und 6.

4.2.5 Multinomialverteilung

Es sei $(\Omega, \mathfrak{A}, P)$ ein Zufallsexperiment und $A_1, \ldots, A_r$ eine Ereignispartition (siehe Satz 2.2 (ii)) von Ω, $p_i = P(A_i)$. Das Experiment wird n-mal wiederholt, $X = (X_1, \ldots, X_r)$ sei der Zufallsvektor, dessen l-te Komponente zählt, wie oft das Ereignis A_l eingetreten ist. Dann gilt in Verallgemeinerung von 4.2.1:
$P(X = (k_1, \ldots, k_r)) = \frac{n!}{k_1! \cdot \ldots \cdot k_r!} {p_1}^{k_1} \cdot \ldots \cdot {p_r}^{k_r}$ für alle $k_1, \ldots, k_r \in \mathbb{N}_0$ mit $k_1 + \ldots + k_r = n$. Man nennt diese Verteilung die Multinomialverteilung mit Parametern n und $p = (p_1, \ldots, p_r)$.

4.3 Erwartungswert und Varianz von Zufallsgrößen

Es sei $(\Omega, \mathfrak{A}, P)$ ein diskreter WRaum und $X : \Omega \to \mathbb{R}$ eine ZV. Welches Resultat erhält man für X im Mittel, wenn das zugehörige Zufallsexperiment n-mal wiederholt wird (für großes n)?
Das Ergebnis ω tritt in n Versuchen $N_n(\omega)$-mal auf, der Mittelwert der Resultate ist dann $\frac{1}{n}(X(\omega_1)N_n(\omega_1) + X(\omega_2)N_n(\omega_2) + \ldots)$.
Für großes n sollte $\frac{1}{n}N_n(\omega_k)$ in der Nähe von $p(\omega_k) = P(\{\omega_k\})$ liegen, d.h. der obige Ausdruck ist ungefähr $X(\omega_1)P(\{\omega_1\}) + X(\omega_2)P(\{\omega_2\}) + \ldots$.
Dies motiviert folgende Definition:

Definition 4.5 $(\Omega, \mathfrak{A}, P)$, X wie oben. Der Erwartungswert von X, Schreibweise EX, wird definiert durch $EX = \sum_{\omega \in \Omega} X(\omega)P(\{\omega\})$, vorausgesetzt, die Summe konvergiert absolut, d.h. $\sum_{\omega \in \Omega} |X(\omega)|P(\{\omega\}) < \infty$.
Ist dies nicht der Fall, so sagen wir, daß der Erwartungswert von X nicht existiert.

Man kann die Summation auf den Bildraum verlagern und erhält dann:

Satz 4.6 $(\Omega, \mathfrak{A}, P)$, X wie oben, $f : \mathbb{R} \to \mathbb{R}$, $Y := f(X)$. Dann ist Y eine (diskrete) Zufallsgröße und mit p_X, p_Y als Massenfunktionen zu X, Y gilt:

$$EX = \sum_{x \in \mathbb{R}} x p_X(x) \quad \Big(:= \sum_{x \in \mathbb{R},\, p_X(x) > 0} x p_X(x)\ \Big)$$

$EY = \sum_{y \in \mathbb{R}} y p_Y(y) = \sum_{x \in \mathbb{R}} f(x) p_X(x)$, vorausgesetzt die beteiligten Summen konvergieren absolut.

Beweis: Die Menge $A_x := \{\omega \in \Omega : X(\omega) = x\}$, $x \in Bild(X)$ bilden eine Partition von Ω. Da absolut konvergente Reihen beliebig umgeordnet werden können, gilt:

$$\sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) = \sum_{x \in Bild(X)} \sum_{\omega \in A_x} X(\omega)P(\{\omega\}) = \sum_{x \in \mathbb{R}} x \sum_{\omega \in A_x} P(\{\omega\}) = \sum_{x \in \mathbb{R}} x \underbrace{P(X = x)}_{p_X(x)}$$

$EY = \sum_{\omega \in \Omega} Y(\omega)P(\{\omega\}) = \sum_{\omega \in \Omega} f(X(\omega))P(\{\omega\}) = \sum_{x \in Bild(X)} f(x) \sum_{\omega \in A_x} P(\{\omega\}) = \sum_{x \in \mathbb{R}} f(x)p_X(x)$, denn f ist auf A_x konstant. □

Wichtige Konsequenz: EX hängt nur von der Verteilung von X ab.
Analogie zur Mechanik: plaziert man Massen $\Pi_1, \Pi_2, \Pi_3, \ldots$ auf die Punkte $x_1, x_2, x_3 \ldots \in \mathbb{R}$, so ist $\sum x_i p_i$ mit $p_i := \frac{\Pi_i}{\sum_j \Pi_j}$, der Schwerpunkt des Gesamtgebildes (EX ist ein Lageparameter für $\mathfrak{L}(X)$).

Beispiel 4.7 Im Falle $X \sim Bin(n, p)$ ergibt sich:

$$EX = \sum_{k=0}^{n} kP(X = k) = \sum_{k=1}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!}p^{k-1}(1-p)^{(n-1)-(k-1)} =$$

$$np \underbrace{\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k}}_{(p+(1-p))^{n-1}} = np.$$

Definiert man $Y := X(X-1)$, so folgt (für $f(k) = k(k-1)$):

$$EY = \sum_{k=0}^{n} f(k)P(X = k) = \sum_{k=1}^{n} k(k-1)\binom{n}{k}p^k(1-p)^{n-k} = \ldots = n(n-1)p^2.$$

Satz 4.8 Es seien X und Y ZV mit existierendem Erwartungswert, $c \in \mathbb{R}$.
(i) Dann existiert auch zu $X + Y$, cX der Erwartungswert und es gilt: $E(X + Y) = EX + EY$, $E(cX) = cEX$ (der Erwartungswertoperator ist linear)
(ii) Wenn $X \leq Y$ (d.h. $X(\omega) \leq Y(\omega)\ \forall \omega \in \Omega$), dann gilt: $EX \leq EY$ (Monotonie)

Beweis: $\sum_{\omega \in \Omega} \underbrace{|(X+Y)(\omega)|}_{\leq |X(\omega)| + |Y(\omega)|} P(\{\omega\}) \leq \sum_{\omega \in \Omega} |X(\omega)|P(\{\omega\}) + \sum_{\omega \in \Omega} |Y(\omega)|P(\{\omega\}) < \infty$

$$E(X + Y) = \sum_{\omega \in \Omega} (X + Y)(\omega)P(\{\omega\}) = \sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) + \sum_{\omega \in \Omega} Y(\omega)P(\{\omega\}).$$

Die anderen Teile folgen analog.

Erste Folgerung: $X \leq |X|,\ -X \leq |X| \implies EX \leq E|X|,\ -EX = E(-X) \leq E|X|$, also: $|EX| \leq E|X|$. □

Definition 4.9 Das k-te Moment (meist $k \in \mathbb{N}$) einer ZV ist EX^k, vorausgesetzt $\sum_{x\in\mathbb{R}} |x|^k P(X = k) < \infty$ (sonst sagen wir, daß das k-te Moment zu X nicht existiert.)

Existiert das zweite Moment zu X, so nennen wir $var(X) := E(X - E(X))^2$, $\sigma := \sqrt{var(X)}$ die Varianz bzw. Standardabweichung von X. Die Varianz ist also die mittlere, quadratische Abweichung der ZV X von ihrem Erwartungswert; Standardabweichung hat dieselbe Dimension wie X.

Nützliche Formel:

Lemma 4.10 $var(X) = E(X^2) - (EX)^2$

Beweis: $var(X) = E(X^2 - 2(EX)X + (EX)^2) = E(X^2) - 2(EX)(EX) + (EX)^2 = E(X^2) - (EX)^2$, wobei wir Satz 4.8 gebraucht haben, und die Tatsache, daß der Erwartungswert einer konstanten ZV gleich dieser Konstanten ist. □

Beispiel 4.11

(i) Im Falle $X \sim Bin(n,p)$ gilt $EX = np,\ EX(X-1) = n(n-1)p^2$ (Bsp. 4.7), also:
$EX^2 = E(X(X-1) + X) = EX(X-1) + E(X) = n(n-1)p^2 + np$, also:
$var(X) = EX^2 - (EX)^2 = n^2p^2 - np^2 + np - (np)^2 = np(1-p)$

(ii) Ist X poisson-verteilt mit Parameter λ (Absatz 4.2.2), so erhält man:

$$EX = \sum_{k=1}^{\infty} k e^{-\lambda}\frac{\lambda^k}{k!} = \lambda e^{-\lambda} \underbrace{\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}}_{e^{\lambda}} = \lambda$$

$$EX(X-1) = \sum_{k=2}^{\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} = \lambda^2$$

$$\implies var(X) = E(X(X-1) + X) - (EX)^2 = \lambda$$

Bei der Poissonverteilung stimmen also Erwartungswert und Varianz überein.

Bemerkung 4.12

Ist M eine beliebige Menge und $A \subseteq M$, so heißt $1_A : M \longrightarrow \mathbb{R},\ 1_A(x) := \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$ die Indikatorfunktion zu A.

Ist $(\Omega, \mathfrak{A}, P)$ ein diskreter WRaum und $A \subseteq \Omega$ ein Ereignis, so ist $X := 1_A$ eine Zufallsvariable. Diese zeigt an, ob das Ereignis A eingetreten ist (1) oder nicht (0). Sei $p := P(A)$. Dann gilt $X \sim Bin(1,p)$.

Mit dieser Konstruktion sieht man, daß Erwartungswerte Wahrscheinlichkeiten verallgemeinern:

$E1_A = 0 \cdot P(1_A = 0) + 1 \cdot P(1_A = 1) = P(A)$, d.h. die Wahrscheinlichkeit eines Ereignisses ist gleich dem Erwartungswert der zugehörigen Indikatorfunktion.

4.4 Bedingte Verteilungen und Unabhängigkeit

Sind $X : \Omega \to S_1,\ Y : \Omega \to S_2$ Zufallsgrößen auf einem diskreten WRaum $(\Omega, \mathfrak{A}, P)$, so ist $Z : \Omega \to S_1 \times S_2$, $Z(\omega) := (X(\omega), Y(\omega))$ eine (diskrete) Zufallsgröße mit Werten in $S_1 \times S_2$. Die Verteilung von Z nennt man auch die gemeinsame Verteilung von X und Y.

Beispiel 4.13 In der Situation aus §3.3.3 (Bridge) sei X die Anzahl der Asse von "Nord", Y die von "Süd". Dann ist $Z := (X,Y)$ eine Zufallsgröße mit Werten in $\{0,\ldots,4\} \times \{0,\ldots,4\}$ und es gilt:

$P(Z = (k,l)) = \frac{\binom{4}{k}\binom{48}{13-k}\binom{4-k}{l}\binom{35+k}{13-l}\binom{26}{13}\binom{13}{13}}{\frac{52!}{(13!)^4}}$.

Tabelle:

	0	1	2	3	4	Zeilensumme
0	1150	2600	1950	572	55	6327
1	2600	4225	2028	286	0	9137
2	1950	2028	468	0	0	4446
3	572	286	0	0	0	858
4	55	0	0	0	0	55

Die Tabelle enthält die Werte $20825 \cdot P(X = k, Y = l)$

Aus den Werte n der Tabelle ergibt sich mit $P(X = i) = P(X = i, Y = 0) + P(X = i, Y = 1) + \ldots + P(X = i, Y = 4)$ für $i = 0,\ldots,4$, und analog für Y, die Marginalverteilung (oder Randverteilung) der Verteilung von Z, dies sind die Verteilungen der Komponenten X und Y. Die gemeinsame Verteilung enthält i.A. "mehr Informationen" als die einzelnen Verteilungen.

Man kann dann die W. von Ereignissen, die von X und Y abhängen, ausrechnen, z.B.:

$P(X = Y) = P(X = 0, Y = 0) + \ldots + P(X = 4, Y = 4) = \frac{1}{20825}(1150 + 4225 + 468 + 0 + 0) = 0.280576\ldots$

"Verlagerungsformel" nützlich für Rechnungen:

$X : \Omega \to S_1,\ Y : \Omega \to S_2,\ f : S_1 \times S_2 \to \mathbb{R}$

$Ef(X,Y) = Ef(Z) = \sum_{z\in\mathbb{R}} f(z)P(Z = z) = \sum_{x\in Bild(X)} \sum_{y\in Bild(Y)} f(x,y)P(X = x, Y = y)$

Satz und Definition 4.14 $(\Omega, \mathfrak{A}, P)$, S_1, S_2, X, Y wie oben. Dann gilt für alle $x \in S_1$ mit $P(X=x)>0$: durch $A \to P[Y \in A|X=x] = \frac{P(\{\omega:Y(\omega)\in A, X(\omega)=x\})}{P(\{\omega:X(\omega)=x\})}$ wird ein WMaß auf $(S_2, \mathbb{P}(S_2))$ definiert, die <u>bedingte Verteilung von Y unter $X=x$</u>, Schreibweise $P^{Y|X=x}$ oder $\mathfrak{L}(Y|X=x)$.
Ist $S_2 = \mathbb{R}$ und $\sum\limits_{y\in\mathbb{R}} |y| P^{Y|X=x}(\{y\}) < \infty$, so nennen wird

$$E[Y|X=x] := \sum_{y\in\mathbb{R}} y P^{Y|X=x}(\{y\}) \quad (= \tfrac{1}{P(X=x)} \sum_{y\in\mathbb{R}} yP(X=x, Y=y))$$

den <u>bedingten Erwartungswert von Y unter $X=x$</u>.
Beweis: klar □

In Bsp. 4.13 ergibt sich als bedingte Erwartung der Anzahl der Asse des Partners, wenn man selbst 2 Asse hat
$E[Y|X=2] = 0\cdot P[Y=0|X=2] + 1\cdot P[Y=1|X=2] + \ldots + 4\cdot P[Y=4|X=2] = 0\cdot\frac{1950}{4446} + 1\cdot\frac{2028}{4446} + 2\cdot\frac{468}{4446} + 3\cdot 0 + 4\cdot 0 = \frac{2}{3}$
Beachte: $EY = 0\cdot P(Y=0) + \ldots + 4\cdot P(Y=4) = 1$

Beispiel 4.15 Es sei $(\Omega', \mathfrak{A}', P')$ ein diskretes Zufallsexperiment, in dem ein bestimmtes Ereignis A mit Wahrscheinlichkeit $p>0$ eintritt. Unser Modell für das n-malige, unabhängige Wiederholen hiervon ist $(\Omega, \mathfrak{A}, P)$ mit $\Omega := \Omega'^n$, $\mathfrak{A} := \mathbb{P}(\Omega)$, $P(\{(\omega_1, \ldots, \omega_n)\}) = P'(\{\omega_1\}) \cdot \ldots \cdot P'(\{\omega_n\})$. Es sei $X : \Omega \to \mathbb{R}$, $X(\omega) := \#\{1 \le i \le n : \omega_i \in A\}$ die Anzahl der Wiederholungen, bei denen A eintritt, und $Y : \Omega \to \mathbb{P}(\{1, \ldots, n\})$, $Y(\omega) := \{i : \omega_i \in A\}$ die Menge der Versuchsnummern, in denen A eintritt.
Die gemeinsame Verteilung von X und Y ist auf $\{(k, B) : k \in \{0, \ldots, n\}, \#B = k\}$ konzentriert, und für jedes Element dieser Menge gilt: $P(X=k, Y=B) = \prod\limits_{j\in B} p \prod\limits_{j\notin B} (1-p) = p^k(1-p)^{n-k}$.
Aus §4.2.1 ist bekannt: $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$, also: $P^{Y|X=k}(\{B\}) = \frac{p^k(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}} = \frac{1}{\binom{n}{k}}$
Die bedingte Verteilung von Y unter $X=k$ ist also die Gleichverteilung (Laplaceverteilung) auf der Menge der Teilmengen vom Umfang k von $\{1, \ldots, n\}$.
Num. Beispiel: $n=3$, $k=2$, also 2 Erfolge in 3 Wiederholungen: Die möglichen "Sequenzen" sind: (1,1,0), (1,0,1), (0,1,1). Die (bed.) W. für jede Sequenz ist $\frac{1}{3}$.
<u>Beachte:</u> Der Parameter p (Erfolgswahrscheinlichkeit) taucht hier nicht mehr auf!
Alle möglichen Anordnungen für die "Erfolge" sind gleichwahrscheinlich. In der Statistik (später) ist es wichtig, daß hier bei der bedingten Verteilung der Parameter p nicht mehr auftritt. ("Alle Informationen zu p sitzen bereits in X")

Beispiel 4.16 Auf dem Produktraum $(\Omega, \mathfrak{A}, P)$ sind die Projektionen $X_i : \Omega \to \Omega_i$, $X_i((\omega_1, \ldots, \omega_n)) = \omega_i$ die Resultate der Einzelexperimente. Für $i<j$ gilt:

$$\begin{aligned} &P(X_i = \omega_i, X_j = \omega_j) = P(\Omega_1 \times \ldots \times \Omega_{i-1} \times \{\omega_i\} \times \Omega_{i+1} \times \ldots \times \Omega_{j-1} \times \{\omega_j\} \times \ldots \times \Omega_n) \\ &= P_i(\{\omega_i\}) \cdot P_j(\{\omega_j\}) = P(X_i = \omega_i) \cdot P(X_j = \omega_j) \end{aligned}$$

Hieraus folgt sofort: $P^{X_i|X_j} = P^{X_i}$ und umgekehrt.

Definition 4.17 Für jedes $i \in I$ sei $X_i : \Omega \to S_i$ eine diskrete Zufallsgröße. Die Familie $\{X_i : i \in I\}$ heißt <u>stochastisch unabhängig</u>, wenn für jede Wahl von $A_i \subseteq S_i$, $i \in I$ die Ereignisfamilie $\{X_i^{-1}(A_i) : i \in I\}$ stochastisch unabhängig ist im Sinne von Def. 2.5.

Satz 4.18
Eine Familie von Mengen $\{X_i : i \in I\}$ von Zufallsgrößen auf einem diskreten WRaum ist genau dann unabhängig, wenn für alle $\{i_1, \ldots, i_n\} \subseteq I$ gilt:
$P(X_{i_1} = x_{i_1}, \ldots, X_{i_n} = x_{i_n}) = \prod\limits_{j=1}^n P(X_{i_j} = x_{i_j})$ für alle $x_{i_1} \in S_{i_1}, \ldots, x_{i_n} \in S_{i_n}$.

Beweis: Notwendigkeit folgt mit $A_i := \{x_i\}$.
Hinreichend: für bel. $A_i \subseteq S_i$, $\{i_1, \ldots, i_n\} \subseteq I$ gilt:

$$\begin{aligned} &P(\bigcap_{j=1}^n X_{i_j}^{-1}(A_{i_j})) = \sum_{x_{i_1}\in A_{i_1}, \ldots, x_{i_n}\in A_{i_n}} P(X_{i_1} = x_{i_1}, \ldots, X_{i_n} = x_{i_n}) \\ &= \sum_{x_{i_1}\in A_{i_1}} P(X_{i_1} = x_{i_1}) \cdot \ldots \cdot \sum_{x_{i_n}\in A_{i_n}} P(X_{i_n} = x_{i_n}) = \prod_{j=1}^n P(X_{i_j} \in A_{i_j}). \end{aligned}$$

□

In Worten: Bei einer endlichen Familie $X_1, \ldots, X_n$ hat man also Unabhängigkeit genau dann, wenn die gemeinsame Massenfunktion $p : p(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$ das "Produkt" der marginalen Massenfunktionen $p_i(x_i) = P(X_i = x_i)$, $1 \le i \le n$, ist in folgendem Sinn: $p(x_1, \ldots, x_n) = p_1(x_1) \cdot \ldots \cdot p_n(x_n)$.
Bei Unabhängigkeit ergibt sich also die gemeinsame Verteilung aus den Randverteilungen, i.A. gilt dies <u>nicht</u>.

4.5 Reellwertige (diskrete) Zufallsgrößen

Mit $\mathbb{R}$ als Wertebereich hat man zusätzliche Strukturen.

Satz 4.19
Sind X und Y unabhängige ZV mit existierendem Erwartungswert, so existiert auch zu $X\cdot Y$ der Erwartungswert und es gilt: $EX\cdot Y = EX \cdot EY$.
Beweis: Die Mengen $A_{xy} := \{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\}$, $x \in \mathit{Bild}(X)$, $y \in \mathit{Bild}(Y)$, bilden eine Partition

von Ω, also $\sum |X \cdot Y(\omega)| P(\{\omega\}) = \sum_{x \in Bild(X)} \sum_{y \in Bild(Y)} |x \cdot y| P(X = x, Y = y) = \sum_x \sum_y |x||y| P(X = x) P(Y = y) = \underbrace{(\sum_x |x| P(X = x))}_{< \infty, \text{ da } EX \text{ existiert}} \underbrace{(\sum_y |y| P(Y = y))}_{< \infty, \text{ da } EY \text{ existiert}} < \infty$, also existiert der Erwartungswert zu $X \cdot Y$

Dies liefert auch die behauptete Formel, wenn man alle Betragsstriche wegläßt. □

Satz 4.20 (Cauchy-Schwarz-Ungleichung)
Existiert zu den ZV X und Y das zweite Moment, so existiert auch der Erwartungswert des Produktes EXY und es gilt $(EXY)^2 \leq EX^2 \cdot EY^2$.

Beweis: Wegen $|(X \cdot Y)(\omega)| = |X(\omega)||Y(\omega)| \leq X(\omega)^2 + Y(\omega)^2$ gilt:
$\sum_{\omega \in \Omega} |(XY)(\omega)| P(\{\omega\}) \leq \underbrace{\sum X(\omega)^2 P(\{\omega\})}_{EX^2 < \infty} + \underbrace{\sum Y(\omega)^2 P(\{\omega\})}_{EY^2 < \infty}$, also existiert der Erwartungswert.

Für beliebiges $t \in \mathbb{R}$ existiert dann auch das zweite Moment zu $X + tY$, und ist nicht-negativ:
$0 \leq E(X + tY)^2 = EX^2 + t^2 EY^2 + 2tEXY \ \forall t \in \mathbb{R}$.
Im Falle $EY^2 = 0$ hat man rechts eine Gerade (über t), die nur dann überall ≥ 0 sein kann, wenn die Steigung 0 ist, d.h. $EXY = 0$.
Im Falle $EY^2 > 0$ erhält man als kleinsten Wert der Parabel auf der rechten Seite $\frac{1}{EY^2}(EX^2 EY^2 - (EXY)^2)$.
Dies ist genau dann ≥ 0, wenn die Ungleichung gilt. □

Definition 4.21 Es seien X und Y ZV mit endlichem, zweitem Moment und den Standardabweichungen σ_X, σ_Y.
Dann heißt $cov(X, Y) := E(X - EX)(Y - EY)$ die Kovarianz von X und Y.
Im Falle $cov(X, Y) = 0$ nennt man X und Y unkorreliert.
Ist σ_X, $\sigma_Y > 0$, so nennt man $\varrho(X, Y) := \frac{cov(X,Y)}{\sigma_x \sigma_y}$ den Korrelationskoeffizienten von X und Y. □

Satz 4.22 Es seien X und Y ZV mit existierendem zweiten Moment. Dann gilt:
(i) $cov(X, Y) = EXY - (EX)(EY)$
(ii) X, Y unabhängig $\Longrightarrow$ X, Y unkorreliert.
(iii) $-1 \leq \varrho(X, Y) \leq 1$

Beweis: (i) Mit der "Linearität des E-Operators" folgt:
$$cov(X, Y) = E(XY - (EX)Y - X(EY) + (EX)(EY)) = EXY - (EX)(EY) - (EX)(EY) + (EX)(EY)$$
(ii) folgt unmittelbar aus (i) und Satz 4.19.
(iii) $\sigma_X^2 \sigma_Y^2 \varrho(X, Y)^2 = (E(E - EX)(Y - EY))^2 \leq E(X - EY)^2 E(Y - EY)^2 = var(X) var(Y)$
$\Longrightarrow \varrho(X, Y)^2 \leq 1$

Gemäß Teil (ii) sind unabhängige ZV unkorreliert, die Umkehrung gilt nicht. Kovarianz und Korrelation können als Maß für die lineare Abhängigkeit von ZV betrachtet werden (siehe Übungen) □

Satz 4.23 Es seien $X_1, \ldots, X_n$ ZV mit existierendem zweiten Moment. Dann gilt:
$var(X_1 + \ldots + X_n) = \sum_{i=1}^n var(X_i) + \sum_{i \neq j} cov(X_i, X_j)$.
Sind $X_1, \ldots, X_n$ unabhängig, so gilt $var(\sum_{i=1}^n X_i) = \sum_{i=1}^n var(X_i)$ (Gleichheit von Bienaymé)

Beweis: Unter Verwendung von Satz 4.22 und Lemma 4.10 folgt:
$$var(\sum_{i=1}^n X_i) = E(\sum_{i=1}^n X_i)^2 - (E(\sum_{i=1}^n X_i))^2 = \sum_{i,j=1}^n EX_i X_j - \sum_{i,j=1}^n EX_i EX_j$$
$$= \sum_{i=1}^n EX_i^2 - (EX_i)^2 + \sum_{i \neq j} (EX_i X_j - EX_i EX_j) = \sum_{i=1}^n var(X_i) + \sum_{i \neq j} cov(X_i, X_j).$$
Der zweite Teil folgt unmittelbar aus 4.22 (ii). □

Beispiel 4.24

(i) In einem Zufallsexperiment sei A ein Ereignis mit W. p. Das Experiment werde n-mal unabhängig wiederholt; X_i zeige an, ob in der i-ten Wiederholung A eingetreten ist ($X_i = 1$) oder nicht ($X_i = 0$).
Dann sind $X_1, \ldots, X_n$ unabhängig mit
$EX_i = 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1) = p$, $EX_i^2 = EX_i = p$.
$var(X_i) = p - p^2 = p(1 - p)$, also gilt für $S_n = X_1 + \ldots + X_n$:
$ES_n = \sum_{i=1}^n EX_i = np$, $var(S_n) = \sum_{i=1}^n var(X_i) = np(1 - p)$(Bienaymé).
Da $S_n \sim Bin(n, p)$ ist dies ein neuer Beweis für die Formel aus Beispiel 4.11. (i).

(ii) Es sei Y die Anzahl der Fixpunkte einer zufälligen Permutation von $\{1, \ldots, n\}$ (siehe 3.3.4)
Sei $X_i(\omega) = \begin{cases} 1 & \omega_i = i \\ 0 & \omega_i \neq i \end{cases}$ $\forall \omega = (\omega_1, \ldots, \omega_n) \in \Omega_n$ (Permutationen). Dann gilt:

$EX_i = P(X_i = 1) = \frac{(n-1)!}{n!} = \frac{1}{n}$ und für $i \neq j$ $EX_iX_j = P(X_i = 1, X_j = 1) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$ und damit:
$var(X_i) = EX_i{}^2 - (EX_i)^2 = EX_i - (EX_i)^2 = \frac{n-1}{n^2}$ für $i \neq j$ $cov(X_i, X_j) = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}$.
(insbesondere sind $X_1, \dots, X_n$ nicht unabhängig.)

Offensichtlich gilt: $Y = X_1 + \ldots + X_n$, also folgt $EY = E\sum_{i=1}^{n} X_i = n \cdot \frac{1}{n} = 1$.

$$var(Y) = \sum_{i=1}^{n} \frac{n-1}{n^2} + \sum_{i \neq j} \frac{1}{n^2(n-1)} = \frac{n-1}{n} + n(n-1)\frac{1}{n^2(n-1)} = 1.$$

Spezieller Fall: ganzzahlige ZV:

Satz und Definition 4.25

(i) Es seien P, Q WMaße auf $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$ mit Massenfunktionen p und q (d.h. $p_k = P(\{k\})$, $q_k = Q(\{k\})$) und $r : \mathbb{Z} \to \mathbb{R}$. $r_n := \sum_{k \in \mathbb{Z}} p_k q_{n-k}$ eine WMassenfunktion. Das zugehörige WMaß R nennen wir die Faltung von P und Q, in Zeichen: $R = P * Q$.

(ii) Sind X und Y unabhängige ZV mit Werten in $\mathbb{Z}$, so ist auch $X + Y$ eine ZV mit Werten in $\mathbb{Z}$, und es gilt:

$$\mathfrak{L}(X + Y) = \mathfrak{L}(X) * \mathfrak{L}(Y)$$

$$\underbrace{P(X + Y = n)}_{r_n} = \sum_{k \in \mathbb{Z}} \underbrace{P(X = k, Y = n - k)}_{=P(X=k) \cdot P(Y=n-k)} = \sum p_k q_{n-k}$$

Beweis: (i) p, q seien Massenfunktionen aus $\mathbb{Z}$ (also $p = (p_k)_{k \in \mathbb{Z}}$ mit $p_k \geq 0$, $\sum p_k = 1$, etc.)
$r : \mathbb{Z} \to \mathbb{R}$ wird definiert durch $r_n := \sum_{k \in \mathbb{Z}} p_k q_{n-k}$

Beh.: $\underbrace{r_k \geq 0}_{\text{trivial}}$, $\sum r_n = 1$

$$\sum_{n \in \mathbb{Z}} r_n = \sum_{n \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} p_k q_{n-k} = \sum_{k \in \mathbb{Z}} p_k \underbrace{\sum_{n \in \mathbb{Z}} q_{n-k}}_{= \sum_{n \in \mathbb{Z}} q_n = 1} = \sum_{k \in \mathbb{Z}} p_k = 1,$$ also definiert r ein WMaß auf $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$

(ii) X, Y unabhängig mit Massenfunktionen p, q $\Longrightarrow$ $X + Y$ hat Massenfunktion r $(= p * q)$
Beweis: $P(X + Y = n) = \sum_{k \in \mathbb{Z}} P(X = k, Y = n - k) = \sum_{k \in \mathbb{Z}} P(X = k)P(Y = n - k) = \sum_{k \in \mathbb{Z}} p_k q_{n-k}$ $\square$

Beispiel 4.26 Es seien X und Y unabhängig, $X \sim Poisson(\lambda)$, $Y \sim Poisson(\mu)$. Dann gilt für alle $n \in \mathbb{N}_0$:

$$P(X + Y = n) = \sum_{k=0}^{n} (e^{-\lambda} \tfrac{\lambda^k}{k!}) \cdot (e^{-\mu} \tfrac{\mu^{n-k}}{(n-k)!}) = e^{-(\lambda+\mu)} \frac{1}{n!} \underbrace{\sum_{k=0}^{n} \binom{n}{k} \lambda^k \mu^{n-k}}_{(\lambda+\mu)^n} = e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^n}{n!}$$

$X+Y$ ist also wieder Poisson-verteilt, und zwar mit Parameter $\lambda+\mu$ (Die Poissonverteilungen bilden eine Faltungsgruppe)
Was ist die bedingte Verteilung von X unter $X + Y$?
Für alle $n \in \mathbb{N}_0$, $k \in \{0, \dots, n\}$ gilt:

$$P(X = k | X + Y = n) = \frac{P(X=k, X+Y=n)}{P(X+Y=n)} = \frac{P(X=k) \cdot P(Y=n-k)}{P(X+Y=n)} = \frac{e^{-\lambda} \frac{\lambda^k}{k!} \cdot e^{-\mu} \frac{\mu^{n-k}}{(n-k)!}}{e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^n}{n!}} = \binom{n}{k} (\tfrac{\lambda}{\lambda+\mu})^k (1 - \tfrac{\lambda}{\lambda+\mu})^{n-k},$$

also $P(X = k | X + Y = n) = Bin(n, \frac{\lambda}{\lambda+\mu})$

4.6 Das schwache Gesetz der großen Zahlen

$X_1, \dots, X_n$ seien unabhängige Zufallsgrößen mit Erwartungswert μ und Varianz σ^2; $\overline{X_n} := \frac{1}{n} \sum_{i=1}^{n} X_i$

$$E\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} EX_i = \mu$$

$$var(\overline{X_n}) = \frac{1}{n^2} var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2} \sum_{i=1}^{n} var(X_i) = \frac{\sigma^2}{n}$$

Also: "Variabilität" geht mit $n \to \infty$ gegen 0

Satz 4.27

(i) (Die Markovsche Ungleichung) Es sei $p > 0$ und $E|X|^p < \infty$. Dann gilt für alle $\alpha > 0$: $P(|X| \geq \alpha) \leq \frac{1}{\alpha^p} E|X|^p$
(ii) (Die Ungleichung von Chebychev) Es sei $EX^2 < \infty$. Dann gilt für alle $\alpha > 0$: $P(|X - EX| \geq \alpha) \leq \frac{1}{\alpha^2} var(X)$

Beweis:

(i) Wir definieren eine neue (diskrete) ZV Y durch $Y(\omega) = \begin{cases} \alpha & , |X(\omega)| \geq \alpha \\ 0 & , \text{sonst} \end{cases}$

Offensichtlich gilt $|Y(\omega)|^p \leq |X(\omega)|^p \quad \forall \omega \in \Omega$, die Monotonieeigenschaft des Erwartungswertes (Satz 4.8) liefert also $E(|Y|^p) \leq E(|X|^p)$.
Da Y nur die Werte 0 und α annimmt, gilt: $E|Y|^p = 0^p \cdot P(Y = 0) + \alpha^p \cdot P(Y = \alpha) = \alpha^p \cdot P(|X| \geq \alpha) \leq E|X|^p$.

(ii) Sei $Y = X - EX$. Verwende Teil (i) mit $p = 2$:
$P(|X - EX| \geq \alpha) = P(|Y| \geq \alpha) \leq \frac{1}{\alpha^2} EY^2 = \frac{1}{\alpha^2} E(X - EX)^2 = \frac{1}{\alpha^2} var(X)$ $\square$

Satz 4.28 (eine einfache Version des schwachen Gesetzes der großen Zahlen)
Es sei $X_1, X_2, \ldots$ eine Folge von paarweise unkorrelierten ZV mit Erwartungswert μ und Varianz σ^2;
$\overline{X_n} := \frac{1}{n} \sum_{i=1}^{n} X_i$
Dann gilt: $P(|\overline{X_n} - \mu| > \epsilon) \longrightarrow 0$ für alle $\epsilon > 0$ und $n \to \infty$

Beweis: Bienaymé und die Rechenregel $var(\alpha X) = \alpha^2 var(X)$ liefern $var(\overline{X_n}) = \frac{1}{n}\sigma^2$, also folgt mit Chebychev

$$P(|\overline{X_n} - \mu| > \epsilon) \leq \frac{1}{\epsilon^2} var(\overline{X_n}) \longrightarrow 0 \quad \forall \epsilon > 0 \text{ (für } n \to \infty)$$

Nimmt man ein festes $\epsilon > 0$ (wie klein auch immer), so geht die W., daß der Mittelwert der Beobachtungen vom gemeinsamen Mittelwert um mehr als ϵ abweicht, gegen 0. □

Wichtiger Spezialfall: Ein Experiment wird unendlich oft wiederholt, X_i zeigt an, ob in der i-ter Wiederholung ein bestimmtes Ereignis A eingetreten ist ($X_i = 1$) oder nicht ($X_i = 0$).
Was ist $\overline{X_n}$ in dieser Situation? Die relative Häufigkeit von A! Was ist μ? $\mu = EX_i = P(A)$
Der obige Satz besagt, daß die relative Häufigkeit von A "in einem gewissen Sinne" gegen die W. von A konvergiert (und liefert die Rechtfertigung für den axiomatischen Aufbau)

Beispiel 4.29 (Eine Anwendung in der Analysis)
Der Approximationssatz von Weierstraß besagt, daß eine stetige, reellwertige Funktion auf einem kompakten Intervall $[a, b] \subseteq \mathbb{R}$ gleichmäßig approximiert werden kann.
Ziel: Ein konstruktiver Beweis mit den Mitteln der Stochastik.
Wir können $[a, b] = [0, 1]$ annehmen.

Sei $p_n : [0,1] \to \mathbb{R}$, $p_n(x) := \sum_{k=0}^{n} f(\frac{k}{n})\binom{n}{k} x^k (1-x)^{n-k}$ das n-te Bernsteinpolynom zu f.

Wir behaupten: $(*)$ $\forall \epsilon > 0\ \exists n_0 \in \mathbb{N}\ \forall n \geq n_0\ \forall x \in [0,1] :\ |f(x) - p_n(x)| \leq \epsilon$

Sei also $\epsilon > 0$. Da eine stetige Funktion auf einem kompakten Intervall gleichmäßig stetig ist, existiert ein $\delta = \delta(\epsilon) > 0$ mit $\forall x, y \in [0,1] :\ |x - y| < \delta \implies |f(x) - f(y)| \leq \frac{\epsilon}{2}$.
Außerdem: Stetige Funktionen sind auf einem kompakten Intervall beschränkt, d.h. es gibt ein $K < \infty$ mit $|f(x)| \leq K\ \forall x \in [0,1]$.

Verbindung zur Stochastik:
Wähle $x \in [0,1]$. Sei $\Omega = \{0,1\}$, $P(\{0\}) = 1 - x$ ($P = Bin(1,x)$). Sei $\Omega_n := \Omega^n$, P_n die Laplace-Verteilung hierauf, X_i die Projektion auf die i-te Koordinate (Münzwurf mit W. für Kopf $= x$, wird n-mal wiederholt, X_i das Resultat des i-ten Wurfes). Dann gilt:
$n\overline{X_n}$ $(= \sum_{i=1}^{n} X_i$, "Anzahl der Erfolge") $\sim Bin(n, x)$, also $Ef(\overline{X_n}) = \sum_{k=0}^{n} f(\frac{k}{n}) P(n\overline{X_n} = k)$ und damit
$Ef(\overline{X_n}) = p_n(x)$
Wie im Beweis von Satz 4.28 folgt: $P(|\overline{X_n} - x| \geq \delta) \leq \frac{x(1-x)}{n \cdot \delta^2} \leq \frac{1}{4n\delta^2}$ (aus AGM-Gleichheit)
Wähle nun $n_0 \in \mathbb{N}$ so groß, daß für alle $n \geq n_0$ gilt: $\frac{2K}{4n\delta^2} < \frac{\epsilon}{2}$. Für alle solchen n gilt dann: $|f(x) - p_n(x)| =$

$$|Ef(\overline{X_n}) - f(x)| \leq \overbrace{E|f(\overline{X_n}) - f(x)| \cdot 1_{\{|\overline{X_n}-x|<\delta\}}}^{\leq \frac{\epsilon}{2} P(|\overline{X_n}-x|<\delta)} + \underbrace{E\,\overbrace{|f(\overline{X_n}) - f(x)|}^{\leq 2K} \cdot 1_{\{|\overline{X_n}-x|\geq\delta\}}}_{2KP(|\overline{X_n}-x|\geq\delta)} \leq \frac{\epsilon}{2} \cdot 1 + \underbrace{2K\frac{1}{4n\delta^2}}_{\leq \frac{\epsilon}{2}} \leq \epsilon$$

Kapitel 5

Allgemeine Wahrscheinlichkeitsräume

5.1 "You can't always get what you want"

In Beispiel 1.1.(iv) haben wir ein Experiment genannt, das eine "Gleichverteilung auf $[0,1)$" erfordert, also einen WRaum $(\Omega, \mathfrak{A}, P)$ mit $\Omega = [0,1)$ und (5.1) $P(x + A$ "modulo 1"$) = P(A)$ für alle $x \in \Omega,\ A \in \mathfrak{A}$.

Satz 5.1 Es gibt kein WMaß auf $\mathbb{P}([0,1))$ mit (5.1).
Beweis: (unter Verwendung des Auswahlaxioms)
Auf $[0,1)$ wird durch $x \sim y \iff x - y \in \mathbb{Q}$ eine Äquivalenzrelation definiert. Das Auswahlaxiom erlaubt es, aus jeder Äquivalenzklasse ein Element auszuwählen; sei A die so erhaltene Menge. Äquivalenzklassen sind disjunkt, also enthält A von jeder Klasse genau ein Element.

Zwischenbehauptung: (i) $(A + x) \cap (A + y) = \emptyset \qquad x, y \in \mathbb{Q} \cap [0,1),\ x \neq y$
(ii) $\bigcup_{x \in \mathbb{Q} \cap [0,1)} (A + x) = [0,1)$ (Addition wieder modulo 1)

zu (i): Angenommen, man hat $a + x = b + y$ mit $x, y \in \mathbb{Q} \cap [0,1),\ x \neq y,\ a, b \in A$. Wegen $x - y \notin \mathbb{Z}$ bedeutet dies $a \neq b$, wegen $a - b \in \mathbb{Q}$ ist dies im Widerspruch zu: "A enthält von jeder Klasse höchstens ein Element"

zu (ii): "$\subseteq$" ist klar
Ist andererseits $z \in [0,1)$, dann existiert ein $a \in A$ mit $a \sim z$, d.h. $x := a - z \in \mathbb{Q}$ (mit dem üblichen $-$). Ersetzt man ggf. x durch $x + 1$, so erhält man die gewünschte Darstellung von z. Ist nun P ein WMaß auf $\mathbb{P}([0,1))$ mit (5.1), so muß P auch der Menge A einen Wert $P(A)$ zuordnen. Mit (5.1),(i),(ii) und der σ-Additivität würde dann $1 = P([0,1)) = P(\bigcup_{x \in \mathbb{Q} \cap [0,1)} (A + x)) = \sum_{x \in \mathbb{Q} \cap [0,1)} P(A + x) = \sum_{x \in \mathbb{Q} \cap [0,1)} P(A)$ folgen. Dies ist unmöglich! □

Die Potenzmenge ist also zu groß, wir werden uns mit einer kleineren σ-Algebra zufrieden geben müssen.

5.2 Mengensysteme

Man wird zumindest gewissen Mengen, wie beispielsweise den Intervallen im Falle $\Omega = \mathbb{R}$, W. zuordnen wollen.

Definition 5.2 Es sei $\Omega \neq \emptyset$ und $\mathfrak{E} \subseteq \mathbb{P}(\Omega)$.
Dann heißt $\sigma(\mathfrak{E}) := \bigcap_{\substack{\mathfrak{A} \supset \mathfrak{E} \\ \mathfrak{A}\ \sigma\text{-Algebra}}} \mathfrak{A}$ die von $\mathfrak{E}$ erzeugte σ-Algebra; $\mathfrak{E}$ nennt man ein Erzeugendensystem von $\mathfrak{A}$.

In dieser Definition haben wir stillschweigend von der (trivialen) Tatsache Gebrauch gemacht, daß der Durchschnitt von beliebig vielen σ-Algebren (über derselben Grundmenge) wieder eine σ-Algebra ist.
Der Durchschnitt ist wegen "$\mathbb{P}(\Omega)$ ist σ-Algebra" nicht leer.

Wichtiger Spezialfall: $\Omega = \mathbb{R}$

Definition 5.3 Die von den LORA-Intervallen $(a, b],\ -\infty < a < b < \infty$, erzeugte σ-Algebra heißt die σ-Algebra der Borel-Mengen von $\mathbb{R}$; Schreibweise: $\mathfrak{B}, \mathfrak{B}_{\mathbb{R}}$.

Eine σ-Algebra $\mathfrak{A}$ kann durchaus verschiedene Erzeugendensysteme haben. Trivialerweise gilt: $\sigma(\mathfrak{A}) = \mathfrak{A}$.

Satz 5.4 $\mathfrak{B}_{\mathbb{R}}$ wird auch erzeugt von
$\mathfrak{E}_1 := \{[a, b) : -\infty < a < b < \infty\}$ (ROLA-Intervalle),
$\mathfrak{E}_2 := \{(-\infty, a] : -\infty < a < \infty\}$ und
$\mathfrak{E}_3 := \{U \subseteq \mathbb{R} : U \text{ offen}\}$.

Beweis: Es sei $\mathfrak{E} := \{(a, b] : -\infty < a < b < \infty\}$. Es reicht, jeweils $\mathfrak{E}_i \subset \mathfrak{B}$ (dies impliziert: $\sigma(\mathfrak{E}_i) \subset \mathfrak{B}$) und $\mathfrak{E} \subset \sigma(\mathfrak{E}_i)$ (dies impliziert: $\mathfrak{B} = \sigma(\mathfrak{E}) \subset \sigma(\mathfrak{E}_i)$) zu zeigen. Hierbei können wir die mengenalgebraischen Abgeschlossenheitseigenschaften von σ-Algebren (gegenüber endlichen und abzählbaren Vereinigungen und Durchschnitten, sowie

Komplementen) verwenden. In diesem Sinne ergibt sich:

$\sigma(\mathfrak{E}_1) \subset \mathfrak{B}$ aus $[a,b) = \bigcap\limits_{n=1}^{\infty} \underbrace{\bigcup\limits_{m=1}^{\infty} (a-\frac{1}{n}, b-\frac{1}{m}]}_{(a-\frac{1}{n},b)}$, und $\mathfrak{B} \subset \sigma(\mathfrak{E}_1)$ durch $(a,b] = \bigcup\limits_{n=1}^{\infty}\bigcap\limits_{m=1}^{\infty} [a+\frac{1}{n}, b+\frac{1}{m})$

$\sigma(\mathfrak{E}_2) = \mathfrak{B}$ folgt aus $(-\infty,a] = \bigcup\limits_{n=1}^{\infty} (a-n,a]$, $(a,b] = (-\infty,b] \cap (-\infty,a]^c$.

Bei $\mathfrak{E}_3$ verwenden wir, daß es zu jedem x aus einer offenen Menge U ein x enthaltendes Intervall $(a,b] \subseteq U$ gibt, wobei wir $a,b \in \mathbb{Q}$ annehmen können. Damit: $U = \bigcup\limits_{\{(a,b)\in\mathbb{Q}^2 \mid (a,b]\subset U\}} (a,b]$. Jede offene Menge läßt sich als abzählbare Vereinigung von LORA-Intervallen schreiben. Dies liefert $\sigma(\mathfrak{E}_3) \subset \mathfrak{B}$, die Gegenrichtung folgt mit $(a,b] = \bigcap\limits_{n=1}^{\infty} (a, b+\frac{1}{n})$.

Dieser Satz impliziert, daß die Intervalle $[a,b)$, $(-\infty,a]$ Borel-Mengen sind, ebenso alle offenen und abgeschlossenen Mengen. Wegen $\{a\} = \bigcap\limits_{n=1}^{\infty} (a-\frac{1}{n}, a]$ (bzw. ihrer Abgeschlossenheit) sind auch alle Einpunktmengen in $\mathfrak{B}$ und damit auch alle abzählbaren Mengen (wie beispielsweise $\mathbb{Q}$) und deren Komplemente (die Menge der irrationalen Zahlen), kompakte Intervalle, etc. □

Kurz: $\mathfrak{B}$ ist "für alle praktischen Zwecke" reichhaltig genug.

Ist A eine nichtleere Teilmenge von $\mathbb{R}$, so wird durch $\mathfrak{B}_A := \{B \cap A : B \in \mathfrak{B}\}$ eine σ-Algebra von A definiert (Übungsaufgabe, die Spur von $\mathfrak{B}$ auf A); wie nennen $\mathfrak{B}_A$ auch die σ-Algebra der Borel-Mengen von A.

Satz 5.5 Es gibt ein WMaß P auf $([0,1), \mathfrak{B}_{[0,1)})$ mit der Eigenschaft (5.2). $P([a,b)) = b-a$ für alle a,b mit $0 \leq a < b < 1$.

Bemerkung 5.6

(i) Man kann leicht zeigen, daß (5.1) auf (5.2) führt. Wir werden später noch sehen, daß für das P aus Satz 5.5 auch (5.1) gilt mit $\mathfrak{A} = \mathfrak{B}_{[0,1)}$. Satz 5.5 zeigt also, daß durch (eine unwesentliche) Verkleinerung des Definitionsbereiches das in 5.1 besprochene Problem gelöst werden kann.

(ii) Man kann P auf $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$ fortsetzen durch $P_{\mathbb{R}}(B) := P(B \cap [0,1))\ \ \forall B \in \mathfrak{B}_{\mathbb{R}}$.
Hat man umgekehrt ein WMaß P auf $(\mathbb{R}, \mathfrak{B})$ und ein $A \in \mathfrak{B}$ mit $P(A) = 1$, so erhält man ein WMaß P_A auf $(A, \mathfrak{B}_A)$ durch $P_A(B) := \underbrace{P(A \cap B)}_{=B}\ \ \forall B \in \mathfrak{B}_A$.

In diesem Sinne nennt man das P aus Satz 5.5 "die" Gleichverteilung auf dem Einheitsintervall, ohne i.A. zu spezifizieren, ob $[0,1)$, $(0,1]$, $(0,1)$, $[0,1]$ gemeint ist, denn wegen $P(\{x\}) = \lim\limits_{n\to\infty} P([x, x+\frac{1}{n})) = 0$ spielen die Randpunkte keine Rolle. $P = unif(0,1)\quad (\Omega = [0,1),\ \mathfrak{A} = \mathfrak{B}_{[0,1)})$.

(iii) In der Maßtheorie nennt man ein Paar $(\Omega, \mathfrak{A})$, $\Omega \neq \emptyset$ und $\mathfrak{A}$ eine σ-Algebra über Ω, einen meßbaren Raum, und eine Abbildung $\mu : \mathfrak{A} \to [0,\infty]$ ein Maß, wenn $\mu(\emptyset) = 0$ und $\mu(\sum\limits_{i=1}^{\infty} A_i) = \sum\limits_{i=1}^{\infty} \mu(A_i)$ für alle $A_1, A_2, \ldots \in \mathfrak{A}$ mit $A_i \cap A_j = \emptyset$ für $i \neq j$ gilt. In diesem Sinne sind Wahrscheinlichkeiten einfach normierte Maße.

Die "geometrische" Version des Problems aus 5.1 lautet: Läßt sich allen Teilmengen von $\mathbb{R}$ ($\mathbb{R}^d$) sinnvoll eine verschiebungsinvariante Länge (bzw. ein Volumen) zuordnen? Es ist wieder eine Einschränkung des Definitionsbereiches notwendig, und man erhält: es gibt ein Maß l (das Lebesgue-Maß) auf $(\mathbb{R}, \mathfrak{B})$ mit $l((a,b]) = b-a$ für alle $-\infty < a < b < \infty$. Man kann $unif(0,1)$ als "Spur" von l auf dem Einheitsintervall auffassen.

Nächstes Problem: Eindeutigkeit (Ist $unif(0,1)$ durch (5.2) eindeutig bestimmt?)
Wichtiges (abstraktes) Hilfsmittel:

Definition 5.7 $\Omega \neq \emptyset$, $\mathfrak{D} \subset \mathbb{P}(\Omega)$ heißt Dynkin-System, wenn gilt:

(i) $\Omega \in \mathfrak{D}$

(ii) $A \in \mathfrak{D} \implies A^c \in \mathfrak{D}$

(iii) $A_1, A_2, \ldots \in \mathfrak{D}, A_i \cap A_j = \emptyset$ für $i \neq j \implies \bigcup\limits_{i=1}^{\infty} A_i \in \mathfrak{D}$

Gegenüber σ-Algebren wird die Forderung "Abgeschlossenheit gegenüber abzählbaren Vereinigungen" auf disjunkte Vereinigungen abgeschwächt.

Der Durchschnitt von (beliebig vielen) Dynkin-Systemen ist wieder ein Dynkin-System, wir können also von

$\delta(\mathfrak{E}) = \bigcap\limits_{\substack{\mathfrak{D} \supset \mathfrak{E} \\ \mathfrak{D}\ \text{Dynkin-System}}} \mathfrak{D}$ als dem von $\mathfrak{E}$ erzeugten Dynkin-System sprechne.

Wir nennen ein Mengensystem $\cap$-stabil, wenn gilt: $A, B \in \mathfrak{E} \implies A \cap B \in \mathfrak{E}$

Satz 5.8 (i) Ein $\cap$-stabiles Dynkin-System ist eine σ-Algebra.
(ii) Ist $\mathfrak{E}$ $\cap$-stabil, so gilt $\delta(\mathfrak{E}) = \sigma(\mathfrak{E})$.

Beweis: (i) Es seien $A_1, A_2, \ldots \in \mathfrak{D}$ (nicht notwendigerweise disjunkt), z.z.: $\bigcup\limits_{i=1}^{\infty} A_i \in \mathfrak{D}$.

Setze $B_1 := A_1$, $B_n := A_n \cap {A_1}^c \cap \ldots \cap A_{n-1}^c$ $(n \geq 1)$. Durchschnittsstabilität und Eigenschaft (ii) liefern $B_n \in \mathfrak{D}\ \forall n \in \mathbb{N}$. Die B_n's sind disjunkt, also gilt nach Eigenschaft (iii): $\bigcup_{i=1}^{\infty} B_n \in \mathfrak{D}$.

Beachte nun: $\bigcup_{i=1}^{\infty} B_n = \bigcup_{i=1}^{\infty} A_n$ (eine ähnliche Konstruktion wurde in Zusammenhang von Satz 1.7 verwendet).

(ii) Da jede σ-Algebra ein Dynkin-System ist, folgt $\delta(\mathfrak{E}) \subset \sigma(\mathfrak{E})$ unmittelbar aus der Definition.
Es sei nun, für $A \in \delta(\mathfrak{E})$, $\mathfrak{D}_A := \{B \subseteq \Omega : B \cap A \in \delta(\mathfrak{E})\}$. Dann ist $\mathfrak{D}_A$ ein Dynkin-System: (i) und (iii) sind trivial, (ii) folgt mit $B^c \cap A = (A^c + B \cap A + \Omega^c + \Omega^c \ldots)^c$.
Da $\mathfrak{E}$ $\cap$-stabil ist, gilt $E' \in \mathfrak{D}_E$ für alle $E, E' \in \mathfrak{E}$, also $\mathfrak{E} \subset \mathfrak{D}_E$ und damit $\delta(\mathfrak{E}) \subseteq \mathfrak{D}_E$, denn $\mathfrak{D}_E$ ist ja Dynkin-System. Dies heißt $D \in \delta(\mathfrak{E})$, $E \in \mathfrak{E} \implies D \cap E \in \delta(\mathfrak{E})$.
Dies wiederum liefert: $\mathfrak{E} \subseteq \mathfrak{D}_D$ für alle $E \in \mathfrak{E}$, $D \in \delta(\mathfrak{E})$, hieraus folgt wieder $\delta(\mathfrak{E}) \subseteq \mathfrak{D}_D$ und damit:

$$A \in \delta(\mathfrak{E}),\ D \in \delta(\mathfrak{E}) \implies A \cap D \in \delta(\mathfrak{E}).$$

Also ist $\delta(\mathfrak{E})$ $\cap$-stabil, damit nach Teil (i) eine σ-Algebra, d.h. $\delta(\mathfrak{E}) \subset \sigma(\mathfrak{E})$. □

Satz 5.9 Es sei $\mathfrak{A}$ eine σ-Algebra mit $\cap$-stabilem Erzeuger $\mathfrak{E}$. Sind dann P und Q WMaße auf $\mathfrak{A}$ mit $P(E) = Q(E)$ für alle $E \in \mathfrak{E}$, so folgt $P(A) = Q(A)$ für alle $A \in \mathfrak{A}$. (Stimmen zwei WMaße auf einem $\cap$-stabilen Erzeuger überein, so sind sie gleich.)
Beweis: Es sei $\mathfrak{D} := \{A \in \mathfrak{A} : P(A) = Q(A)\}$. Dann ist $\mathfrak{D}$ ein Dynkin-System (leicht nachzuweisen), und es gilt: $\mathfrak{E} \subseteq \mathfrak{D}$. Satz 5.8 (ii) liefert nun $\mathfrak{A} \supset \mathfrak{D} \supset \delta(\mathfrak{E}) = \sigma(\mathfrak{E}) = \mathfrak{A}$. □

Die Mengen $[a, b)$, $0 \leq a \leq b < 1$ bilden ein Erzeugendensystem von $\mathfrak{B}_{[0,1)}$ (siehe Aufgabe 26 (ii)), dieses System ist offensichtlich $\cap$-stabil. Also: Es gibt nur <u>ein</u> WMaß auf $\mathfrak{B}_{[0,1)}$ mit der Eigenschaft (5.2); wir können also von <u>der</u> Gleichverteilung auf dem Einheitsintervall sprechen.

5.3 Zufallsgrößen und Verteilungen

Oft interessiert nicht $\omega \in \Omega$ selbst, sondern der Wert $X(\omega)$ einer Funktion X hiervon. Für hinreichend viele (idealerweise: alle) Teilmengen A' des Bildraumes Ω' der Abbildung X wollen wir von der Wahrscheinlichkeit sprechen können, daß X in A' liegt. Hierzu muß $X^{-1}(A') = \{\omega \in \Omega : X(\omega) \in A'\}$ im Definitionsbereich von P, also $\mathfrak{A}$, liegen.

Definition 5.10 Es seien $(\Omega, \mathfrak{A}, P)$ ein WRaum und $(\Omega', \mathfrak{A}')$ ein meßbarer Raum. Eine Abbildung $X : \Omega \to \Omega'$ heißt <u>Zufallsgröße</u> (auf $(\Omega, \mathfrak{A}, P)$ und mit Werten in $(\Omega', \mathfrak{A}')$), wenn X <u>$(\mathfrak{A}, \mathfrak{A}')$-meßbar</u> ist, d.h $X^{-1}(A') \in \mathfrak{A}\ \ \forall A' \in \mathfrak{A}$.

Der Begriff Meßbarkeit stammt aus der Maßtheorie.

Hilfreiche Analogie: Eine Topologie auf einer Menge M wird durch das System $\mathfrak{U}$ der offenen Mengen beschrieben. Eine Abbildung $f : M \to M'$ vom topologischen Raum $(M, \mathfrak{U})$ in einen weiteren topologischen Raum $(M', \mathfrak{U}')$ heißt <u>stetig</u>, wenn $f^{-1}(U') \in \mathfrak{U}$ gilt für alle $U' \in \mathfrak{U}'$. Also:
Stetigkeit: Urbilder offener Mengen sind offen.
Meßbarkeit: Urbilder meßbarer Mengen sind meßbar.

Klar: Im Falle $\mathfrak{A} := \mathbb{P}(\Omega)$ ist $X^{-1}(A') \in \mathfrak{A}$ stets erfüllt. Dies ist der Grund dafür, daß man bei diskreten WRäumen ohne den Meßbarkeitsbegriff auskommt.

Satz und Definition 5.11 Ist X eine $(\Omega', \mathfrak{A}')$-wertige Zufallsgröße auf einem WRaum $(\Omega, \mathfrak{A}, P)$, so wird durch $\mathfrak{A}' \ni A' \to P(X \in A')$ $(= P(\{\omega \in \Omega : X(\omega) \in A'\})$ ein WMaß auf $(\Omega', \mathfrak{A}')$ definiert. Dieses WMaß heißt die <u>Verteilung von X</u>, Schreibweise: P^X, $\mathfrak{L}(X)$.
Bei Beachtung der Meßbarkeit ist der Beweis identisch zum Beweis von Satz 4.2. □

Stochastik	Maßtheorie
Zufallsgröße	meßbare Abbildung
Verteilung der ZG.	Bildmaß

Ein Stab der Länge 1 wird zerbrochen. $\Omega = (0, 1)$, $\mathfrak{A} = \mathfrak{B}_{(0,1)}$, $P = unif(0, 1)$
X sei die Länge des kürzeren Stücks; $X : \Omega \to \mathbb{R}$, $X(\omega) = \min\{\omega, 1 - \omega\}$.

$$P(X < \tfrac{1}{4}) = P(X^{-1}(\underbrace{(0, \tfrac{1}{4})}_{\in \mathfrak{B}})) = P((0, \tfrac{1}{4}) \cup (\tfrac{3}{4}, 1)) = \tfrac{1}{2}.$$

$$X^{-1}((0, \tfrac{1}{4})) = (0, \tfrac{1}{4}) \cup (\tfrac{3}{4}, 1) \in \mathfrak{A}$$

Beim Nachweis der Meßbarkeit kann man sich auf Erzeugendensysteme beschränken:

Satz 5.12 Es seien $(\Omega, \mathfrak{A})$ und $(\Omega', \mathfrak{A}')$ meßbare Räume und $X : \Omega \to \Omega'$ eine Abbildung. Ist $\mathfrak{E}' \subseteq \mathbb{P}(\Omega')$ ein Erzeugendensystem von $\mathfrak{A}'$ und gilt: $X^{-1}(E') \in \mathfrak{A}\ \forall E' \in \mathfrak{E}'$, so ist X $(\mathfrak{A}, \mathfrak{A}')$-meßbar (d.h $X^{-1}(A') \in \mathfrak{A}\ \forall A' \in \mathfrak{A}'$).
Beweis: Es sei $\mathfrak{A}_0 := \{A' \subseteq \Omega' : X^{-1}(A') \in \mathfrak{A}\}$. Dann ist $\mathfrak{A}_0$ eine σ-Algebra (über Ω'), denn:

$$X^{-1}(\Omega') = \Omega \in \mathfrak{A}, \text{ also: } \Omega' \in \mathfrak{A}_0.$$

$X^{-1}(A^c) = \{\omega \in \Omega : X(\omega) \notin A\} = (\{\omega \in \Omega : X(\omega) \in A\})^c = (X^{-1}(A))^c$, also gilt:
$A \in \mathfrak{A}_0 \Rightarrow X^{-1}(A) \in \mathfrak{A} \Rightarrow (X^{-1}(A))^c \in \mathfrak{A} \Rightarrow X^{-1}(A^c) \in \mathfrak{A} \Rightarrow A^c \in \mathfrak{A}_0$.

Analog erhält man mit $X^{-1}(\bigcup_{n=1}^{\infty} A_n) = \bigcup_{n=1}^{\infty} X^{-1}(A_n)$ die dritte definierende Eigenschaft einer σ-Algebra.

Nach Vorlesung gilt: $\mathfrak{E}' \subseteq \mathfrak{A}_0$, also $\mathfrak{A}' = \sigma(\mathfrak{E}') \subseteq \mathfrak{A}_0$ und damit $X^{-1}(A') \in \mathfrak{A}\ \forall A' \in \mathfrak{A}$. □

Beispiel 5.13

Es sei $(\Omega, \mathfrak{A}, P) = ([0,1), \mathfrak{B}_{[0,1)}, unif(0,1))$. Für jedes $x \in \Omega$ werde $T_x : \Omega \to \Omega$ definiert durch

$$T_x(y) := \begin{cases} y - x & , \text{ wenn } y \geq x \\ y - x + 1 & , \text{ sonst} \end{cases}.$$

Für alle $A \in \mathfrak{A}$ gilt dann $T_x^{-1}(A) = \{y \in A : y - x \in A \text{ oder } y - x + 1 \in A\}$, insbesondere:

$$T_x^{-1}([0,a)) = \begin{cases} [x, x+a) & , \text{ wenn } x + a < 1 \\ [0, x+a-1) \cup [x, 1) & , \text{ sonst} \end{cases}.$$

Mit $\sigma(\{[0,a) : 0 < a \leq 1\}) = \mathfrak{A}$ und Satz 5.12 folgt hieraus die $(\mathfrak{A},\mathfrak{A})$-Meßbarkeit von T_x.
Man sieht auch, daß $P(T_x^{-1}([0,a))) = a = P([0,a))$ gilt. Mit Satz 5.9 folgt: $P^{T_x} = P$.
Das liefert $P(x + A) = P(A)$ für alle $A \in \mathfrak{A}$, d.h. das WMaß $unif(0,1)$ hat die Eigenschaft (5.1)
("Translationsinvarianz modulo 1")

"Verknüpfungen meßbarer Abbildungen sind meßbar":

Satz 5.14 Es seien $(\Omega, \mathfrak{A})$, $(\Omega', \mathfrak{A}')$, $(\Omega'', \mathfrak{A}'')$ meßbare Räume sowie $X : \Omega \to \Omega'$, $Y : \Omega' \to \Omega''$ seien $(\mathfrak{A}, \mathfrak{A}')$- bzw. $(\mathfrak{A}', \mathfrak{A}'')$-meßbare Abbildungen. Dann ist $Z := Y \circ X$ $(\mathfrak{A}, \mathfrak{A}'')$-meßbar.

Beweis: Für alle $A'' \in \mathfrak{A}''$ gilt:
$Z^{-1}(A'') = \{\omega \in \Omega : Y(X(\omega)) \in A''\} = X^{-1}(\{\omega' \in \Omega' : Y(\omega') \in A''\}) = X^{-1}(Y^{-1}(A'')) \in \mathfrak{A}$,
denn $A' := Y^{-1}(A'') \in \mathfrak{A}'$ und $X^{-1}(A') \in \mathfrak{A}$ aufgrund der vorausgesetzten Meßbarkeiten. □

5.4 Reellwertige Zufallsgrößen

Wie in der diskreten Situation (4.5) ist $\mathbb{R}$ als Wertebereich besonders wichtig. Es sei $(\Omega, \mathfrak{A}, P)$ ein WRaum; als σ-Algebra auf $\mathbb{R}$ werden wir grundsätzlich die σ-Algebra $\mathfrak{B}$ der Borel-Mengen nehmen.
Aus Satz 5.4 und Satz 5.12 folgt unmittelbar, daß $X : \Omega \to \mathbb{R}$ genau dann eine Zufallsgröße (Zufallsvariable, ZV) ist, wenn gilt: $X^{-1}((-\infty, a]) = \{\omega \in \Omega : X(\omega) \leq a\} \in \mathfrak{A}\ \forall a \in \mathbb{R}$.
Den einfachsten Fall solcher Abbildungen liefern die Indikatorfunktionen. Wegen

$$1_A^{-1}((-\infty, a]) = \begin{cases} \emptyset & , \quad a < 0 \\ A^c & , \quad 0 \leq a < 1 \\ \Omega & , \quad a \geq 1 \end{cases}$$

ist 1_A genau dann eine ZV, wenn $A \in \mathfrak{A}$ gilt ("ZV verallgemeinern Ereignisse").

Häufig werden mit einer ZV X Operationen ausgeführt: Im Zusammenhang mit Streuungen ist beispielweise X^2 interessant. Ist X^2 wieder eine ZV?

Satz 5.15

Ist $g : \mathbb{R} \to \mathbb{R}$ stetig oder (schwach) monoton steigend oder (schwach) monoton fallend, so ist g $(\mathfrak{B}, \mathfrak{B})$-meßbar.
Beweis: Ist g stetig, so ist $g^{-1}(U)$ für jede offene Menge U offen, also eine Borel-Menge.
Hieraus folgt die Behauptung mit den Sätzen 5.4 und 5.12.
Rest: Übung. □

Ist X eine ZV, so kann X^2 als Verknüpfung der $(\mathfrak{A}, \mathfrak{B})$-meßbaren Abbildung X und der $(\mathfrak{B}, \mathfrak{B})$-meßbaren (weil stetigen) Abbildung $g : \mathbb{R} \to \mathbb{R}$, $g(x) := x^2$, angesehen werden, ist also nach Satz 5.14 $(\mathfrak{A}, \mathfrak{B})$-meßbar und damit wieder eine ZV.

Satz 5.16

(i) Sind X und Y ZV auf $(\Omega, \mathfrak{A}, P)$, so liegen die Mengen $\{X < Y\} = \{\omega \in \Omega : X(\omega) < Y(\omega)\}$, $\{X \leq Y\}$, $\{X = Y\}$, $\{X \neq Y\}$ in $\mathfrak{A}$.

(ii) Sind X, Y ZV auf $(\Omega, \mathfrak{A}, P)$ und $\alpha, \beta \in \mathbb{R}$, so sind auch $\alpha X + \beta$, $X + Y$, $X \cdot Y$, $X \wedge Y$ $(= \min\{X, Y\})$ und $X \vee Y$ $(= \max\{X, Y\})$ wieder ZV.

(iii) Ist $(X_n)_{n \in \mathbb{N}}$ eine Folge von ZV auf $(\Omega, \mathfrak{A}, P)$, so sind auch $\sup_{n\in\mathbb{N}} X_n$, $\inf_{n\in\mathbb{N}} X_n$, $\limsup_{n\to\infty} X_n$, $\liminf_{n\to\infty} X_n$ ZV (vorausgesetzt, sie sind $\mathbb{R}$-wertig.) Gilt $X_n(\omega) \to X(\omega)$ für alle $\omega \in \Omega$, so ist auch X eine ZV.

Beweis:

(i) $\{X < Y\} = \bigcup_{q \in \mathbb{Q}} \{X < q\} \cap \{Y > q\}$,$\{X < q\} = X^{-1}((-\infty, q)) \in \mathfrak{A}$, da $(-\infty, q) \in \mathfrak{B}$ und X-meßbar.
Abzählbare Vereinigungen von Elementen aus $\mathfrak{A}$ sind wieder in $\mathfrak{A}$.
$\{X \leq Y\} = \{Y < X\}^c \in \mathfrak{A}$, $\{X = Y\} = \{X \leq Y\} \cap \{X < Y\}^c \in \mathfrak{A}$, $\{X \neq Y\} = \{X = Y\}^c \in \mathfrak{A}$.

(ii) Die Abbildung $x \to \alpha x + \beta$ ist stetig, also ist $\alpha X + \beta$ eine ZV nach dem "Hintereinanderschaltungsargument".
Weiter gilt $\{X + Y \leq \alpha\} = \{X \leq Y - \alpha\} \in \mathfrak{A}$, denn $Y - \alpha$ ist wieder ZV, nach Teil (i) ist also $X + Y$ meßbar.
Mit $X \cdot Y = \frac{1}{4}((X+Y)^2 - (X-Y)^2)$ folgt dann die Meßbarkeit von $X \cdot Y$. Mit $\{X \vee Y \leq \alpha\} = \{X \leq \alpha\} \cap \{Y \leq \alpha\}$ und $\{X \wedge Y \leq \alpha\} = \{X \leq \alpha\} \cup \{Y \leq \alpha\}$ folgt die Meßbarkeit von $X \vee Y$ und $X \wedge Y$.

(iii) $\{\omega \in \Omega : \sup\limits_{n\in\mathbb{N}} X_n(\omega) \le a\} = \bigcap\limits_{n=1}^{\infty} \underbrace{\{\omega \in \Omega : X_n(\omega) \le a\}}_{\in\mathfrak{A}}$

Die Meßbarkeit der anderen Abbildungen folgt nun mit $\inf\limits_{n\in\mathbb{N}} X_n = -\sup\limits_{n\in\mathbb{N}}(-X_n)$, $\limsup\limits_{n\to\infty} X_n = \inf\limits_{n\in\mathbb{N}}\sup\limits_{m\ge n} X_m$, $\liminf\limits_{n\to\infty} X_n = -\limsup\limits_{n\to\infty}(-X_n)$.

Konvergiert X_n mit $n\to\infty$ punktweise gegen X, so gilt $X = \limsup\limits_{n\to\infty} X_n$, also ist auch X eine ZV. □

In Teil (iii) läßt sich die Einschränkung auf reellwertige Abbildungen beseitigen, wenn man $\mathbb{R}$ zu wieder $\overline{\mathbb{R}} := \mathbb{R}\cup\{+\infty,-\infty\} = [-\infty,+\infty]$ erweitert und $\mathfrak{B}$ passend ergänzt.

5.5 Verteilungsfunktionen

Die Verteilung einer reellwertigen Zufallsgröße (ZV) ist ein WMaß auf $(\mathbb{R},\mathfrak{B})$, also eine Abbildung von $\mathfrak{B}$ in $\mathbb{R}$ (schwierig).
Ziel: Beschreibung durch eine Abbildung von $\mathbb{R}$ in $\mathbb{R}$ (einfacher).

Definition 5.17 Die Verteilungsfunktion F zu einem Wahrscheinlichkeitsmaß P auf $(\mathbb{R},\mathfrak{B})$ wird definiert durch $F:\mathbb{R}\to\mathbb{R},\ F(x) := P((-\infty,x])$ für alle $x\in\mathbb{R}$.
Ist P eine Verteilung einer ZV X, so nennen wir F auch Verteilungsfunktion zu X.

Da die Mengen $(-\infty,x]$, $x\in\mathbb{R}$, ein $\cap$-stabiles Erzeugendensystem von $\mathfrak{B}$ bilden (Satz 5.14), wird P durch das zugehörige F eindeutig festgelegt (Satz 5.9).

Satz 5.18 Ist F eine Verteilungsfunktion zu einem WMaß P auf $(\mathbb{R},\mathfrak{B})$, so hat F die folgenden Eigenschaften:

(i) $\lim\limits_{x\to-\infty} F(x) = 0,\ \lim\limits_{x\to\infty} F(x) = 1$

(ii) F ist schwach monoton steigend.

(iii) F ist stetig von rechts.

Beweis: (ii) folgt unmittelbar aus der Monotonie von P (siehe Satz 1.6 (iv))

(i) Sei $(x_n)_{n\in\mathbb{N}}\subseteq\mathbb{R}$ mit $\lim\limits_{n\to\infty} x_n = -\infty$. Setze $y_n := \sup\limits_{m\ge n} x_m$. Dann gilt $x_n \le y_n,\ y_n\downarrow-\infty$, also $(-\infty,x]\downarrow\emptyset$, und es folgt mit der "Stetigkeit von P in der leeren Menge" (Aufgabe 5 (a) mit $\bigcap\limits_{n=1}^{\infty} A_n = \emptyset$):

$$0 \le F(x_n) \le F(y_n) = P((-\infty,y_n]) \longrightarrow 0 \text{ für } n\to\infty, \text{ also } F(x_n)\to 0.$$

Die andere Aussage erhält man analog wegen der Stetigkeit von P von unten (in der Gesamtmenge $\mathbb{R}$, Satz 1.7 (b)).

(iii) Ist $(x_n)_{n\in\mathbb{N}}\subseteq\mathbb{R}$ mit $x_n\ge x,\ x_n\to x$, so gilt für $y_n := \sup\limits_{m\ge n} x_m$, $y_m\downarrow x$, also

$$F(x) = P((-\infty,x]) \le P((-\infty,x_n]) \le P((-\infty,y_n]) \longrightarrow P((-\infty,x]) = F(x),$$

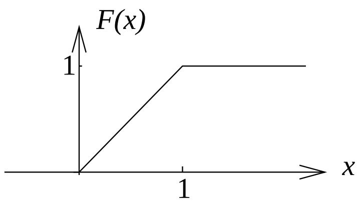
da die $(-\infty,y_n]_{n\in\mathbb{N}}$ eine antitone Folge bilden. □

Wir wollen nun zeigen, daß die Eigenschaftsliste vollständig ist, d.h. daß zu jedem F mit den Eigenschaften (i)-(iii) ein WMaß P existiert, dessen Verteilungsfunktion F ist.

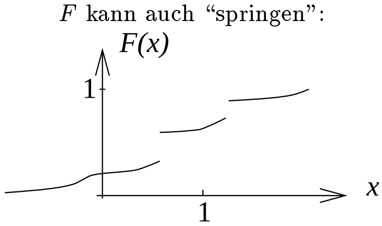
Definition 5.19 Es sei F eine Funktion mit den Eigenschaften (i)-(iii) aus Satz 5.18. Dann definieren wir die Quantilfunktion Q zu F durch $Q:(0,1)\to\mathbb{R},\ Q(y) := \inf\{x\in\mathbb{R} : F(x)\ge y\}$ (wir schreiben auch F^{-1} für Q).

Ist X eine ZV mit Verteilungsfunktion F, so nennt man $F^{-1}(\alpha)$ $(0<\alpha<1)$ auch das α-Quantil zu X (bzw. $\mathfrak{L}(X)$, F). Es ist dies der kleinste Wert q_α mit der Eigenschaft, daß der Wert von X mit W. $\ge\alpha$ nicht größer ist.

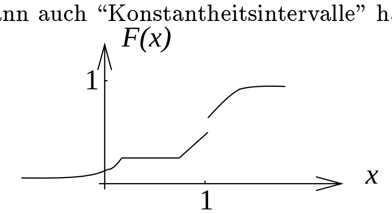
Im Falle $X\sim unif(0,1)$ erhält man: $F(X) = P(X\le x) = \begin{cases} 0 & , \quad x<0 \\ x & , \quad 0\le x\le 1 \\ 1 & , \quad x>1 \end{cases}$



F kann auch "springen":



F kann auch "Konstantheitsintervalle" haben:



Nur wenn F stetig und streng monoton wächst ist, ist F^{-1} die Umkehrfunktion zu F "im üblichen Sinne".

Lemma 5.20 $y\le F(x) \iff F^{-1}(y)\le x$

Beweis: "$\Longrightarrow$" folgt unmittelbar aus der Definition von F^{-1}.

Da außerdem $F(x) < y \implies F(x + \frac{1}{n}) < y$ für ein $n \in \mathbb{N}$, denn f ist stetig von rechts $\implies F^{-1}(y) \geq x + \frac{1}{n}$, denn F ist schwach monoton steigend $\implies F^{-1}(y) > x$ gilt, hat man auch die Gegenrichtung. $\square$

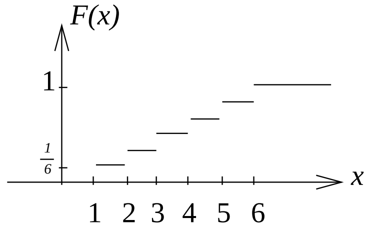
Satz 5.21 Es sei $F : \mathbb{R} \to \mathbb{R}$ eine Funktion mit den Eigenschaften (i)-(iii) aus Satz 5.18. Dann existiert ein WMaß P auf $(\mathbb{R}, \mathfrak{B})$ mit Verteilungsfunktion F.

Beweis: Es sei $\Omega = (0,1)$, $\mathfrak{A} := \mathfrak{B}_{(0,1)}$ und $P_0 = unif(0,1)$. Wir definieren nun $X : \Omega \to \mathbb{R}$ durch $X(\omega) := F^{-1}(\omega)$.

Dann ist X eine ZV (Meßbarkeit folgt aus der Monotonie und Aufgabe 29 (c)), und Lemma 5.20 liefert für $P := \mathfrak{L}(X)$:

$P((-\infty, x]) = P_0(X \leq x) = P_0(\{\omega \in \Omega : F^{-1}(\omega) \leq x\}) = P_0(\{\omega \in \Omega : \omega \leq F(x)\}) = P_0((0, F(x)]) = F(x)$. $\square$

Beispiel: Würfelwurf



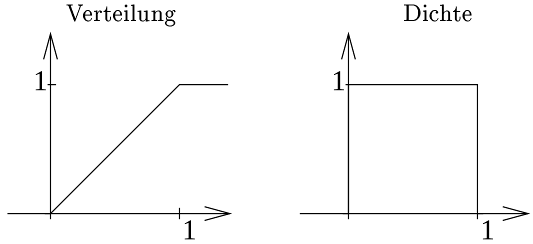
Der Übergang von $P : \mathfrak{B} \to \mathbb{R}$ zu $F : \mathbb{R} \to \mathbb{R}$ wurde durch Eigenschaften von $\mathbb{R}$ ermöglicht und bedeutet eine erhebliche Vereinfachung. Satz 5.21 zeigt auch, daß es zu jedem WMaß auf $(\mathbb{R}, \mathfrak{B})$ eine ZV mit diesem WMaß als Verteilung gibt.
In den Übungen wird gezeigt, daß Verteilungsfunktionen linksseitige Limiten $F(x-)$ haben, und $P(X = x) = F(x) - F(x-)$ (Sprung in x) gilt.
Die Verteilungsfunktionen zu einer diskreten ZV besteht nur aus Sprüngen, deren Höhe durch die Massenfunktion angegeben werden.

Ist $f : \mathbb{R} \to \mathbb{R}_+$ eine Funktion mit $\int_{-\infty}^{\infty} f(x)dx = 1$, so wird nach dem obigen Resultat durch $P((-\infty, x]) := \int_{-\infty}^{x} f(y)dy$ für alle $x \in \mathbb{R}$ ein WMaß auf $(\mathbb{R}, \mathfrak{B})$ definiert, das WMaß mit <u>(Riemann-) Dichte</u> f.

Beispiel 5.22 Bei $P = unif(0,1)$ hat man

$$P((-\infty, x]) = \begin{cases} 0 & , \quad x < 0 \\ x & , \quad 0 \leq x \leq 1 \\ 1 & , \quad x > 1 \end{cases} = \int_{-\infty}^{x} f(y)dy \text{ mit}$$

$$f(y) = \begin{cases} 0 & , \quad x > 1 \text{ oder } x < 0 \\ 1 & , \quad 0 < x < 1 \end{cases} = 1_{(0,1)}(x).$$



zu $unif(0,1)$

WDichten sind in mancher Hinsicht ein infintesimales Analogon zur WMassenfunktion, es können aber Werte > 1 angenommen werden. Ganz allgemein: $P(X \in A) = \int_A f(x)dx$ (Die Wahrscheinlichkeit ergibt sich als Fläche unter f).

5.6 Einige wichtige Verteilungen mit Riemann-Dichten

5.6.1 Gleich- bzw. Rechteck-Verteilung

Die Funktion $f_{a,b} : \mathbb{R} \to \mathbb{R}$, $f_{a,b} = \begin{cases} \frac{1}{b-a} & , \quad a < x < b \\ 0 & , \quad \text{sonst} \end{cases}$ hat die Eigenschaften

$f_{a,b} \geq 0$, $\int_{-\infty}^{\infty} f_{a,b}(x)dx = 1$ (wobei $-\infty < a < b < \infty$), ist also eine WDichte.

Das zugehörige WMaß nennen wir die <u>Gleich-</u> oder <u>Rechteckverteilung</u> auf (a,b) und schreiben $unif(a,b)$.

Alle diese Verteilungen gehen durch affine Transformationen aus $unif(0,1)$ hervor: hat X die Verteilung $unif(0,1)$, so gilt für $Y := a + (b-a)X$ $\quad P(Y \leq y) = P(X \leq \frac{y-a}{b-a}) = \frac{y-a}{b-a}$ für $a < y < b$. Außerdem ist $P(Y \leq y) = 0$ für $y \leq a$, und $P(Y \leq y) = 1$ für $y \geq b$, also insgesamt $P(Y \leq y) = \int_{-\infty}^{y} f_{a,b}(x)dx$, d.h. $Y \sim unif(a,b)$.

Beispiel 5.23 Ein Stab der Länge 1 zerbricht an einer zufälligen Stelle. Wir machen die (etwas unrealistische) Annahme, daß alle Bruchpositionen gleich wahrscheinlich sind und erhalten als Modell für dieses Zufallsexperiment den WRaum $(\Omega, \mathfrak{A}, P)$ mit $\Omega = (0,1)$, $\mathfrak{A} = \mathfrak{B}_{(0,1)}$, $P = unif(0,1)$. Die Länge des kürzeren Bruchstückes ist $X(\omega) = \min\{\omega, 1-\omega\}$. (Nach Satz 5.16 ist dies eine ZV). Welche Verteilung hat X?

Klar: $P(X \leq x) = \begin{cases} 0 & , \quad x \leq 0 \\ 1 & , \quad x \geq \frac{1}{2} \end{cases}$

Für $x \in (0, \frac{1}{2})$ erhält man:

$$P(X \leq x) = P(\{\omega \in \Omega : \omega \leq x \text{ oder } 1 - \omega \leq x\}) = P((0,x] \cup [1-x, 1)) = (x - 0) + (1 - (1-x)) = 2x.$$

Dies ist die Verteilungsfunktion zu $unif(0, \frac{1}{2})$, d.h. es gilt: $X \sim unif(0, \frac{1}{2})$.

5.6.2 Gamma- und Exponential-Verteilung

Die Gamma-Verteilung mit Parametern α und λ ($\alpha > 0,\ \lambda > 0$) ist die Verteilung mit Dichte

$$f_{\alpha,\lambda}(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1}\lambda^{\alpha} e^{-\lambda x} \text{ für } x > 0,\ 0 \text{ sonst.}$$

(Hierbei ist $\Gamma(z) := \int\limits_0^\infty x^{z-1}e^{-x}dx$ die Gammafunktion.)

Schreibweise: $\Gamma(\alpha, \lambda)$. Taucht in verschiedenen Zusammenhängen auf (Wartezeiten, Statistik, ...)
Besonders wichtig ist der Fall $\alpha = 1$, der auf die Exponentialverteilung führt.

5.6.3 Normalverteilung

Die Normalverteilung mit den Parametern μ und σ^2, kurz $N(\mu, \sigma^2)$, wobei $\mu \in \mathbb{R}$ beliebig und $\sigma^2 > 0$ ist die Verteilung mit der Dichte $\varphi_{\mu,\sigma^2} := \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2),\ x \in \mathbb{R}$.

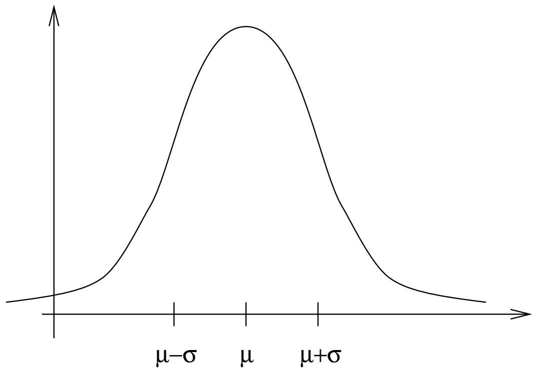
Die Parameter μ und σ^2 beschreiben die Lage und Breite von φ. Im Falle $\mu = 0,\ \sigma^2 = 1$ spricht man von den Standardparametern, $N(0,1)$ ist die Standardverteilung.
Offensichtlich gilt: $\varphi_{\mu,\sigma^2} = \frac{1}{\sigma}\varphi_{0,1}(\frac{x-\mu}{\sigma}) \quad \forall x \in \mathbb{R}$.

Die Verteilungsfunktion zu $N(0,1)$ ist Φ mit

$$\Phi : \begin{cases} \mathbb{R} & \to \quad [0,1] \\ x & \mapsto \quad \Phi(x) := \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{y^2}{2})dy \end{cases}.$$

Eine Variante hiervon ist als "Fehlerfunktion" bekannt.

Das Diagramm zeigt die Gaußsche Glockenkurve:



Φ ist vertafelt und in gängigen Softwarepaketen enthalten. Für statistische Anwendungen sind die zugehörigen α-Quantile von Bedeutung:

α	0.9	0.95	0.975	0.99	0.995
u_α	1.2816	1.6449	1.9600	2.3263	2.5758

$\Phi(u_\alpha) = \alpha$

Lemma 5.24

(i) $\int\limits_{-\infty}^{\infty} \varphi_{\mu,\sigma^2}(x)dx = 1 \quad \forall \mu \in \mathbb{R},\ \sigma^2 > 0.$

(ii) $\Phi(x) = 1 - \Phi(-x) \quad \forall x \in \mathbb{R}.$

(iii) $X \sim N(\mu, \sigma^2),\ a \neq 0,\ b \in \mathbb{R} \implies Y := aX + b \sim N(a\mu + b, a^2\sigma^2)$

Beweis:

(i) Substitution $y := \frac{1}{\sigma}(x - \mu)$ zeigt, daß es reicht, den Fall $\mu = 0,\ \sigma^2 = 1$ zu behandeln:

$$\left(\int\limits_{-\infty}^{\infty} e^{-\frac{x^2}{2}}dx\right)^2 = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \exp(-\tfrac{1}{2}(x^2+y^2))dxdy = \int\limits_0^{2\pi}\int\limits_0^{\infty} \underbrace{r\exp(-\frac{1}{2}r^2)}_{-\frac{d}{dr}\exp(-\frac{1}{2}r^2)}\, dr d\varphi = 2\pi.$$

(ii) folgt sofort mit $\varphi(x) = \varphi(-x)$ (φ steht für $\varphi_{0,1}$).

(iii) Im Falle $a > 0$ hat man $\quad P(Y \leq y) = P(X \leq \frac{y-b}{a}) = \int\limits_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2)dx$

$$= \int\limits_{-\infty}^{y} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2 a^2}} \exp(-\frac{1}{2\sigma^2 a^2}(x' - (a\mu + b))^2)}_{\text{Dichte zu } N(a\mu+b, \sigma^2a^2)}\, dx' \text{ mit } x' = ax + b, \text{ also: } Y \sim N(a\mu + b, \sigma^2 a^2).$$

Wegen (ii) und (iii) reicht es, die Verteilungsfunktion zu $N(\mu, \sigma^2)$ für Standardparameter und Argumente ≥ 0 zu vertafeln; beispielsweise gilt: $u_\alpha = -u_{1-\alpha}$. □

5.7 Erwartungswerte

Die "amtliche" Verallgemeinerung des diskreten Falls erfordert das (allgemeine) Lebesgue-Integral (Stochastik 2), wir begnügen uns mit Andeutungen.
Ist X eine ZV mit Dichte f, und setzt man für alle $x \in \mathbb{R} \quad \lceil x \rceil := \min\{k \in \mathbb{Z} : k \geq x\},\ \lfloor x \rfloor := \max\{k \in \mathbb{Z} : k \leq x\}$, so wird durch $\underline{X_n} := 2^{-n}\lfloor 2^n X \rfloor,\ \overline{X_n} := 2^{-n}\lceil 2^n X \rceil$ eine Familie von diskreten ZV definiert mit

$$\underline{X_n} \leq X \leq \overline{X_n},\ \overline{X_n} - \underline{X_n} \leq 2^{-n}.$$

Bei diesen können wir den Erwartungswert ausrechnen: $E\underline{X_n} = \sum\limits_{k\in\mathbb{Z}} 2^{-n}kP(\underline{X_n} = k \cdot 2^{-n}) = \sum\limits_{k\in\mathbb{Z}} 2^{-n}k \int\limits_{k\cdot 2^{-n}}^{(k+1)\cdot 2^{-n}} f(x)dx$

$$= \int\limits_{-\infty}^{\infty} \frac{\lfloor 2^n x \rfloor}{2^n} f(x)dx \leq \int\limits_{-\infty}^{\infty} xf(x)dx \leq \int\limits_{-\infty}^{\infty} \frac{\lceil 2^n x \rceil}{2^n} f(x)dx = \ldots = E\overline{X_n}.$$

Es liegt also nahe, den Erwartungswert von X (im Falle $\int |x|f(x)dx < \infty$) zu definieren durch $EX := \int xf(x)dx$.

Kritikpunkte: 1. Beweise von Linearität, $Eg(X) = \int g(x)f(x)dx$ wird unsäglich.
2. Es gibt auch ZV, die weder diskret sind, noch eine Dichte haben.

Beispiel 5.25 Im Falle $X \sim N(\mu, \sigma^2)$ erhält man $EX = \int\limits_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2)dx$

$$= \underbrace{\int\limits_{-\infty}^{\infty} (x-\mu)\frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2)dx}_{=0,\text{ wegen Symmetrie}} + \mu \underbrace{\int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2)dx}_{=1,\text{ da Integal über Dichte}} = \mu.$$

5.8 Unabhängigkeit

Bisher waren σ-Algebren nur "notwendiges Übel". Sie spielen aber auch beim Unabhängigkeitsbegriff und als Informationsrepräsentanten eine wichtige Rolle.

Satz und Definition 5.26 Es sei X eine Zufallsgröße auf dem WRaum $(\Omega, \mathfrak{A}, P)$ mit Werten in $(\Omega', \mathfrak{A}')$. Dann ist $\{X^{-1}(A) : A \in \mathfrak{A}'\}$ eine σ-Algebra, die von X erzeugte σ-Algebra $\sigma(X)$.
Ist $\mathfrak{E}'$ ein $\cap$-stabiles Erzeugendensystem von $\mathfrak{A}'$, so ist $\{X^{-1}(E') : E' \in \mathfrak{E}'\}$ ein $\cap$-stabiles Erzeugendensystem von $\sigma(X)$.

Beweis: Der erste Teil folgt leicht mit $\Omega = X^{-1}(\Omega')$, $X^{-1}(A^c) = (X^{-1}(A))^c$, $X^{-1}(\bigcup\limits_{i=1}^{\infty} A_n) = \bigcup\limits_{i=1}^{\infty} X^{-1}(A_n)$.

Zweiter Teil: Übungsaufgabe □

Kennen wir das Resultat ω eines Zufallsexperiments, so können wir von jedem Ereignis $A \in \mathfrak{A}$ sagen, ob es eingetreten ist oder nicht; $\sigma(X)$ ist die Menge der Ereignisse, für die wir diese Entscheidung treffen können, wenn uns $X(\omega)$ bekannt ist.

Definition 5.27 Es sei $(\Omega, \mathfrak{A}, P)$ ein WRaum, $I \neq \emptyset$.

(a) Eine Familie $\{\mathfrak{A}_i : i \in I\}$ von Unter-σ-Algebren von $\mathfrak{A}$ heißt stochastisch unabhängig, wenn für jede endliche Teilmenge $J = \{j_1, \ldots, j_n\}$ von I und alle $A_{j_1} \in \mathfrak{A}_{j_1}, \ldots, A_{j_n} \in \mathfrak{A}_{j_n}$ gilt:
$$(*) \qquad P(\bigcap_{j\in J} A_j) = \prod_{j\in J} P(A_j)$$

(b) Ist für jedes $i \in I$ X_i eine Zufallsvariable auf $(\Omega, \mathfrak{A}, P)$ mit Werten in $(\Omega_i, \mathfrak{A}_i)$, so heißt die Familie $\{X_i : i \in I\}$ stochastisch unabhängig (kurz: die X_i's sind unabhängig), wenn die Familie $\{\sigma(X_i) : i \in I\}$ der erzeugten σ-Algebren im Sinne von (a) unabhängig ist.

Der folgende Satz zeigt, daß man sich beim Nachweis von (*) auf $\cap$-stabile Erzeugendensysteme beschränken kann.

Satz 5.28 Es sei $(\Omega, \mathfrak{A}, P)$ ein WRaum.
Für jedes $i \in I \neq \emptyset$ sei $\mathfrak{A}_i$ eine Unter-σ-Algebra von $\mathfrak{A}$ mit $\cap$-stabilem Erzeugendensystem $\mathfrak{E}_i$.
Gilt dann $P(\bigcap\limits_{j\in J} E_j) = \prod\limits_{j\in J} P(E_j)$ für alle endlichen $J = \{j_1, \ldots, j_n\} \subseteq I$, $E_{j_k} \in \mathfrak{E}_{j_k}$, so sind die $\mathfrak{A}_i$, $i \in I$, unabhängig.

Beweis: Sei $J = \{j_1, \ldots j_n\} \subseteq I$, D_{j_1} die Menge alle $A \in \mathfrak{A}_{j_1}$ mit $P(A \cap E_{j_2} \cap \ldots \cap E_{j_n}) = P(A) \cdot P(E_{j_2}) \cdot \ldots \cdot P(E_{j_n})$ für alle $E_{j_2} \in \mathfrak{E}_{j_2}, \ldots, E_{j_n} \in \mathfrak{E}_{j_n}$.
Man sieht leicht, daß D_{j_1} ein Dynkin-System ist.
Da D_{j_1} den $\cap$-stabilen Erzeuger $\mathfrak{E}_{j_1}$ von $\mathfrak{A}_{j_1}$ enthält, folgt $D_{j_1} = \mathfrak{A}_{j_1}$ mit Satz 5.18 (ii).
Im zweiten Schritt sei D_{j_2} die Menge aller $A \in \mathfrak{A}_{j_2}$ mit
$P(A_{i_1} \cap A \cap E_{j_3} \cap \ldots \cap E_{i_n}) = P(A_{i_1}) \cdot P(A) \cdot P(E_{j_3}) \cdot \ldots \cdot P(E_{j_n})$.
Man sieht wieder, daß D_{j_2} ein Dynkin-System ist, das nach dem bereits bewiesenen Teil $\mathfrak{E}_{j_2}$ enthält, also gilt: $D_{j_2} = \mathfrak{A}_{j_2}$.
Nach insgesamt n Schritten dieser Art hat man die gesamte Beziehung
$P(A_{j_1} \cap A_{j_2} \cap \ldots \cap A_{j_n}) = P(A_{j_1}) \cdot P(A_{j_2}) \cdot \ldots \cdot P(A_{j_n})$ für alle $A_{j_1} \in \mathfrak{A}_{j_1}, \ldots, A_{j_n} \in \mathfrak{A}_{j_n}$. □

Bei einer diskreten Zufallsgröße X bilden die Mengen $X^{-1}(\{x\})$, $x \in Bild(X)$, ein $\cap$-stabiles Erzeugendensystem von $\sigma(X)$ (Satz 5.26); Satz 4.18 zeigt also, daß Definition 5.27 (b) zur Definition 4.17 "abwärtskompatibel" ist. Der Zugang über σ-Algebren hat auch Vorteile, beispielsweise im Beweis von Satz 5.29.

Satz 5.29 Für jedes $i \in I$ sei X_i eine Zufallsgröße mit Werten in $(\Omega_i, \mathfrak{A}_i)$, $(\Omega_i', \mathfrak{A}_i')$ seien weiterere, meßbare Räume und $g_i : \Omega_i \to \Omega_i'$ $(\mathfrak{A}_i, \mathfrak{A}_i')$-meßbare Funktionen. Ist dann $\{X_i : i \in I\}$ eine unabhängige Familie, so ist auch $\{Y_i : i \in I\}$ mit $Y_i := g_i(X_i)$ unabhängig. (kurz: Funktionen von unabhängigen Zufallsgrößen sind wieder unabhängig.)

Beweis: $\sigma(Y_i) \subseteq \sigma(X_i)$ □

Beispiel 5.30 Es sei $(\Omega, \mathfrak{A}, P) = ([0,1), \mathfrak{B}_{[0,1)}, unif(0,1))$.
Für jedes $n \in \mathbb{N}$ werde $X_n : \Omega \to \{0,1\}$ definiert durch $X_n(\omega) := \lfloor 2^n\omega \rfloor - 2\lfloor 2^{n-1}\omega \rfloor$.
Dann gilt $\omega = \sum\limits_{n=1}^{\infty} 2^{-n} X_n(\omega) = 0.X_1(\omega)X_2(\omega)\ldots$ ist "die" Binärdarstellung von ω.
Für alle $k_1, \ldots k_n \in \{0,1\}$ gilt: $P(X_1 = k_1, \ldots, X_n = k_n) = P(\{\omega \in \Omega : \sum\limits_{l=1}^{n} 2^{-l}k_l \leq \omega < \sum\limits_{l=1}^{n} 2^{-l}k_l + 2^{-n}\}) = 2^{-n}$ (das

Intervall besteht aus allen $\omega \in [0,1)$, deren Binärdarstellung mit den Ziffern (Bits) $k_1, \dots, k_n$ beginnt.)
Für beliebige $i_1 < i_2 < \dots < i_n$ erhält man somit

$$P(X_{i_1} = 1, \dots, X_{i_n} = 1) = \sum_{\substack{(k_1,\dots,k_{i_n}) \in \{0,1\}^{i_n} \\ k_{i_j} = 1 \text{ für } j=1,\dots,n}} P(X_1 = k_1, \dots, X_{i_n} = k_{i_n})$$

$= 2^{-i_n} \#\{(k_1, \dots, k_{i_n}) \in \{0,1\}^{i_n} : k_{i_j} = 1 \text{ für } j = 1, \dots, n\} = 2^{-i_n} \cdot 2^{i_n - n} = 2^{-n}$, denn n Positionen sind festgelegt.
Insbesondere folgt $P(X_{i_j} = 1) = \frac{1}{2}$ und damit $P(X_{i_1} = 1, \dots, X_{i_n} = 1) = P(X_{i_1} = 1) \cdot \ldots \cdot P(X_{i_n} = 1)$.
Da $\{X_i^{-1}(\{1\})\}$ ein $\cap$-stabiles Erzeugendensystem von $\sigma(X_i)$ ist, ist damit gezeigt, daß die ZV $X_1, X_2, \dots$ unabhängig sind, außerdem gilt: $\mathfrak{L}(X_i) = Bin(1, \frac{1}{2})$.

Die gesamte Konstruktion kann also Modell für den unendlich oft wiederholten Wurf einer fairen Münze dienen.
Umgekehrt ließe sich aus einer unenlichen Folge $k_1, k_2 \dots$ von Münzwürfen eine auf $[0,1)$ gleichverteilte Zahl x durch $x := \sum_{i=1}^{\infty} k_i 2^{-i}$ konstruieren.

Kapitel 6

Verteilungskonvergenz und Normalapproximation

6.1 Verteilungskonvergenz

Aus Satz 4.4 bekannt: $Bin(n,p)$ kann bei großem n und kleinem p durch die Poissonverteilung mit Parameter $\lambda = np$ approximiert werden.
Jetzt allgemeiner: Approximation von Verteilungen und zugehörige Grenzwertsätze.

Beispiel 6.1 Die Zufallsvariablen $X_1, \dots, X_n$ seien unabhängig und $unif(0,1)$-verteilt. Wir interessieren uns für die Verteilung des Minimums $Y_n := \min\{X_1, \dots, X_n\}$.

Die zugehörige Verteilungsfunktion ist $F_{Y_n}(y) = P(Y_n \le y) = 1 - P(X_1 > y, \dots, X_n > y) =$

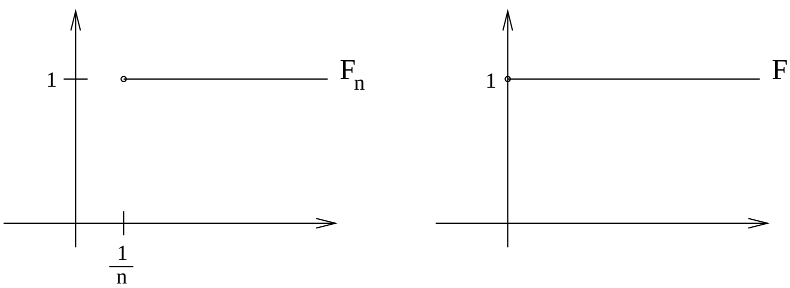
$$1 - P(X_1 > y) \cdot \ldots \cdot P(X_n > y) = 1 - \prod_{i=1}^{n} (1 - F_{X_i}(y)) = \begin{cases} 1-(1-y)^n & , \quad 0 \le y \le 1 \\ 0 & , \quad y < 0 \\ 1 & , \quad y > 1 \end{cases} .$$

Mit $n \to \infty$ erhalten wir den (punktweisen) Limes $\begin{cases} 1 & , \quad y > 0 \\ 0 & , \quad y \le 0 \end{cases}$ (beachte: keine Verteilungsfunktion, da nicht rechtsstetig in 0). Dies gibt uns die intuitive und nicht sonderlich interessante Aussage, daß das Minimum "in gewisser Weise" gegen 0 geht.
Umskalieren: Betrachte $n \cdot Y_n$ (statt Y_n).
Für alle $y > 0$ gilt mit n groß genug:
$F_{nY_n}(y) = P(nY_n \le y) = P(Y_n \le \frac{y}{n}) = 1 - (1 - \frac{y}{n})^n \underset{n\to\infty}{\longrightarrow} 1 - e^{-y}$, d.h. bei großem n ist nY_n näherungsweise exponentialverteilt mit Parameter 1.

Für WMaß gibt es eine ganze Reihe von Konvergenzbegriffen. Der wichtigste:

Definition 6.2 Es seien $P,\ P_n\ (n \in \mathbb{N})$ WMaße auf $(\mathbb{R}, \mathfrak{B})$ mit den Verteilungsfunktionen $F,\ F_n\ (n \in \mathbb{N})$.
$C(F) := \{x \in \mathbb{R} : F \text{ stetig in } x\}$ (Menge der Stetigkeitspunkte von F). Gilt dann: $\lim_{n\to\infty} F_n(x) = F(x)$ für alle $x \in C(F)$, so sagt man, daß P_n schwach gegen P konvergiert ($P_n \xrightarrow{W} P$, w: "weak"). Sind $X,\ X_n\ (n \in \mathbb{N})$ ZV mit Verteilungen $P,\ P_n\ (n \in \mathbb{N})$, so sagt man in dieser Situation, daß X_n in Verteilung gegen X konvergiert ($X_n \xrightarrow{D} X$, "distribution").

Warum Einschränkung auf $C(F)$? Ohne diese Einschränkung würde man mit $X_n \equiv \frac{1}{n}$ nicht den Limes $X \equiv 0$ erhalten.



6.2 Normalapproximation bei Poisson-Verteilungen

Es sei X Poisson-verteilt mit Parameter λ.
Beispiel 4.11.(ii) liefert $EX = \lambda$, $var(X) = \lambda$, mit Chebychev (Satz 4.27.(ii)) $P_\lambda(|X-\lambda| \geq C \cdot \sqrt{\lambda}) \leq \frac{1}{C^2}$
Die Hauptmenge der Verteilung liegt also im Intervall $I(C,\lambda) := [\lambda - C\sqrt{\lambda}, \lambda + C\sqrt{\lambda}] \cap \mathbb{N}_0$
Es sei $\varphi(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2)$ die Dichte zu $N(\mu,\sigma^2)$. Der folgende Satz zeigt, daß $P_\lambda(X=k)$ bei großem λ auf $I(C,\lambda)$ gleichmäßig durch $\varphi(k|\lambda,\lambda)$ approximiert werden kann:

Satz 6.3 (Normalapproximation für Poisson-Verteilungen, lokale Form)
Für alle $C > 0$ gilt: $\sup\limits_{k \in I(C,\lambda)} \left| \frac{e^{-\lambda}\frac{\lambda^k}{k!}}{\frac{1}{\sqrt{2\pi\lambda}}\exp(-\frac{1}{2\lambda}(k-\lambda)^2)} - 1 \right| \underset{\lambda\to\infty}{\longrightarrow} 0.$

Beweis: Sei $C > 0$. Wir setzen, für $k \in \mathbb{N}_0$, $\lambda > 0$

$$g(k|\lambda) := \log P_\lambda(X=k) = -\lambda + k\log\lambda - \log k!$$
$$z(k,\lambda) := \tfrac{1}{\sqrt{\lambda}}(k-\lambda).$$

Dann gilt: $g(k|\lambda) - g(k-1|\lambda) = \log\lambda - \log k$ sowie $|z(k,\lambda)| \leq C$ für alle $k \in I(C,\lambda)$.
Mit $\log(1+x) = x + O(x^2)$ $(x \to 0)$ erhält man $\log k = \log(\lambda + \sqrt{\lambda}z(k,\lambda)) = \log\lambda + \log(1 + \frac{1}{\sqrt{\lambda}}z(k,\lambda)) = \log\lambda + \frac{1}{\sqrt{\lambda}}z(k,\lambda) + O(\frac{1}{\lambda})$ mit $\lambda \to \infty$, gleichmäßig auf $I(C,\lambda)$.

Damit $g(k|\lambda) - g(\lfloor\lambda\rfloor|\lambda) = \sum\limits_{j=\lfloor\lambda\rfloor+1}^{k}(g(j|\lambda) - g(j+1|\lambda)) = -\frac{1}{\sqrt{\lambda}} \sum\limits_{j=\lfloor\lambda\rfloor+1}^{k} z(j,\lambda) + O(\frac{1}{\sqrt{\lambda}}) = \ldots$

Die Anzahl der Summanden ist wegen $k \in I(C,\lambda)$ maximal $C \cdot \sqrt{\lambda}$, $C\sqrt{\lambda} \cdot O(\frac{1}{\lambda}) = O(\frac{1}{\sqrt{\lambda}})$
(die "Ungenauigkeit" landet im $O(\frac{1}{\sqrt{\lambda}})$-Term)

$$\ldots = -\frac{1}{\lambda} \sum_{j=\lfloor\lambda\rfloor+1}^{k}(j-\lambda) + O(\tfrac{1}{\sqrt{\lambda}}) = -\tfrac{1}{2\lambda}(k-\lambda)^2 + O(\tfrac{1}{\sqrt{\lambda}})$$

Setzt man $h(\lambda) := \exp(g(\lfloor\lambda\rfloor|\lambda))$, so haben wir gezeigt: $P_\lambda(X=k) = h(\lambda)\exp(-\frac{1}{2\lambda}(k-\lambda)^2)(1+O(\frac{1}{\sqrt{\lambda}}))$.
Es bleibt zu zeigen:
(*) $\lim\limits_{\lambda\to\infty} \sqrt{2\pi\lambda}h(\lambda) = 1$
Hierfür summieren wir die bereits erhaltene Approximation über $I(C,\lambda)$:
$1 \geq P(X \in I(C,\lambda)) = h(\lambda)(\sum\limits_{k\in I(C,\lambda)} \exp(-\frac{(k-\lambda)^2}{2\lambda})(1+O(\frac{1}{\sqrt{\lambda}})) = \sqrt{2\pi\lambda}h(\lambda)(1+O(\frac{1}{\sqrt{\lambda}})) \sum\limits_{|k-\lambda|\leq C\cdot\sqrt{\lambda}} \frac{1}{\sqrt{\lambda}}\varphi(\frac{k-\lambda}{\sqrt{\lambda}})$

Ganz rechts steht eine Riemann-Summe. Diese konvergiert (mit $\lambda \to \infty$) gegen $\int\limits_{-C}^{C} \varphi(x)dx$, d.h.

$$1 \geq (\limsup_{\lambda\to\infty} \sqrt{2\pi\lambda}h(\lambda)) \int_{-C}^{C} \varphi(x)dx.$$

Mit $C \to \infty$ folgt hieraus wegen $\int\limits_{-\infty}^{\infty} \varphi(x)dx = 1$: $\limsup\limits_{\lambda\to\infty} \sqrt{2\pi\lambda}h(\lambda) \leq 1.$

Andererseits hat man nach Chebychev (siehe Bemerkung vor diesem Satz) $1 - \frac{1}{C^2} \leq P_\lambda(X \in I(C,\lambda)) = \ldots$, also

$$\underbrace{(\liminf_{\lambda\to\infty} \sqrt{2\pi\lambda}h(\lambda))}_{\text{hängt nicht von } C \text{ ab}} \quad \underbrace{\int_{-C}^{C} \varphi(x)dx}_{\to 1 \text{ mit } C\to\infty} \quad \geq \quad \underbrace{1 - \frac{1}{C^2}}_{\to 1 \text{ mit } C\to\infty}$$

und damit $\liminf\limits_{\lambda\to\infty} \sqrt{2\pi\lambda}h(\lambda) \geq 1$.

Damit ist (*) bewiesen. □

Als Korollar ergibt sich eine (be)merkenswerte Formel:

Korollar 6.4 (Stirling-Formel)
$n! \sim \sqrt{2\pi n}e^{-n}n^n$ (in dem Sinne, daß das Verhältnis gegen 1 geht.)
Beweis: Setze $\lambda = \lambda_n = n$, $k = n$, verwende Satz 6.3.

$$\frac{e^{-\lambda}\frac{\lambda^k}{k!}}{\frac{1}{\sqrt{2\pi\lambda}}\exp(-\frac{1}{2\lambda}(k-\lambda)^2)} \overset{\lambda\to\infty}{\longrightarrow} 1 \text{ glm. auf } I(C,\lambda) := [\lambda - C\sqrt{\lambda}, \lambda + C\sqrt{\lambda}]$$ □

Satz 6.5 (Normalapproximation für Poisson-Verteilungen, kummulative Form)
Für alle $\lambda > 0$ sei X_λ eine zum Parameter λ poissonverteilte ZV. Dann gilt $\mathfrak{L}(\frac{X_\lambda - \lambda}{\sqrt{\lambda}}) \xrightarrow{W} N(0,1)$.
Beweis: Für beliebige $a, b \in \mathbb{R}$ mit $a < b$ gibt es ein $C > 0$ mit $[a,b] \subseteq [-C,C]$, also folgt wie im Beweis von Satz 6.3:
$\lim\limits_{\lambda\to\infty} P(a \leq \frac{X_\lambda-\lambda}{\sqrt{\lambda}} \leq b) = \lim\limits_{\lambda\to\infty} \sum\limits_{\lambda+a\sqrt{\lambda}\leq k\leq\lambda+b\sqrt{\lambda}} P(X_\lambda = k) = \lim\limits_{\lambda\to\infty} \sum\limits_{a\leq\frac{k-\lambda}{\sqrt{\lambda}}\leq b} \frac{1}{\sqrt{\lambda}}\varphi(\frac{k-\lambda}{\sqrt{\lambda}}) = \int\limits_a^b \varphi(x)dx = \Phi(b) - \Phi(a)$,
wobei Φ die Verteilungsfunktion zu $N(0,1)$ bezeichnet.
Es bleibt zu zeigen, daß a durch $-\infty$ ersetzt werden kann.
Sei hierzu $\epsilon > 0$. Dann existiert ein $a > -\infty$ mit $\Phi(a) < \frac{\epsilon}{2}$, und $P(\frac{X_\lambda-\lambda}{\sqrt{\lambda}} < a) < \frac{\epsilon}{2}$ für alle $\lambda > 0$ (mit Chebychev erhält man für die W. die von λ unabhängige Oberschranke $\frac{1}{a^2}$). Die Dreiecksungleichung liefert:
$|P(\frac{X_\lambda-\lambda}{\sqrt{\lambda}} \leq b) - \Phi(b)| \leq |P(a \leq \frac{X_\lambda-\lambda}{\sqrt{\lambda}} \leq b) - (\Phi(b) - \Phi(a))| + P(\frac{X_\lambda-\lambda}{\sqrt{\lambda}} \leq a) + \Phi(a)$.

Hieraus und aus den Bedingungen an a folgt, daß jeder Häufungspunkt von $|P(\frac{X_\lambda-\lambda}{\sqrt{\lambda}}) - \Phi(b)|$ bei $\lambda \to \infty$ $\leq \epsilon$ ist. Da ϵ beliebig war, folgt $\lim\limits_{\lambda\to\infty} P(\frac{X_\lambda-\lambda}{\sqrt{\lambda}} \leq b) = \Phi(b)$ für alle $b \in \mathbb{R}$. □

6.3 Normalapproximation bei der Binomialverteilung

Wie betrachten nun das Verhalten von $Bin(n,p)$ bei $n \to \infty$ und festem $p \in (0,1)$.

Satz 6.6 (Normalapproximation für Binomialverteilungen, lokale Form)
Es sei $p \in (0,1)$ und, für alle $C > 0$, $n \in \mathbb{N}$, $I(C,n) := [np - C\sqrt{n}, np + C\sqrt{n}] \cap \{0,\dots,n\}$. Dann gilt für alle $C > 0$:

$$\lim_{n\to\infty} \sup_{k\in I(C,n)} \left| \frac{\binom{n}{k}p^k(1-p)^{n-k}}{\varphi(k|np,np(1-p))} - 1 \right| = 0.$$

Beweis: Setzt man $\lambda = np$, $\mu = n(1-p)$, so gilt $\binom{n}{k}p^k(1-p)^{n-k} = \frac{e^{-\lambda}\frac{\lambda^k}{k!}e^{-\mu}\frac{\mu^{n-k}}{(n-k)!}}{e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^n}{n!}}$.
(Dieser Trick wird durch den aus Beispiel 4.26 bekannten Sachverhalt $P^{X|X+Y} = Bin(X+Y, \frac{\lambda}{\lambda+\mu})$ bei X, Y unabhängig, $X \sim Poisson(\lambda)$, $Y \sim Poisson(\mu)$ nahegelegt.) Wendet man Satz 6.3 dreimal an, so folgt:

$$\binom{n}{k}p^k(1-p)^{n-k} \sim \frac{\varphi(k|\lambda,\lambda)\varphi(n-k|\mu,\mu)}{\varphi(n|\lambda+\mu,\lambda+\mu)} = \frac{\frac{1}{\sqrt{2\pi\lambda}}\exp(-\frac{1}{2\lambda}(k-\lambda)^2)\frac{1}{\sqrt{2\pi\mu}}\exp(-\frac{1}{2\mu}\exp(-\frac{1}{2\mu}(n-k-\mu)^2)}{\frac{1}{\sqrt{2\pi(\lambda+\mu)}}\exp(-\frac{1}{2(\lambda+\mu)}(n-\lambda-\mu)^2)}$$

$$= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{np\cdot n(1-p)}}\exp(-\frac{1}{2np}(k-np)^2 - \frac{1}{2n(1-p)}(n-k-n(1-p))^2) = \frac{1}{\sqrt{2\pi np(1-p)}}\exp(-\frac{1}{2n}(\frac{1}{p}+\frac{1}{1-p})(k-np)^2)$$

$$= \varphi(k|np,np(1-p)).$$ □

Satz 6.7 (deMoivre-Laplace)
Es sei $X_1, X_2, \dots$ eine Folge von unabhängigen ZV mit $P(X_i = 1) = 1 - P(X_i = 0) = p$ für alle $i \in \mathbb{N}$, wobei $p \in (0,1)$ fest. Sei $S_n := \sum_{i=1}^n X_i$ für alle $n \in \mathbb{N}$. Dann gilt $\mathfrak{L}\left(\frac{S_n - np}{\sqrt{np(1-p)}}\right) \xrightarrow{W} N(0,1)$ mit $n \to \infty$.
Beweis: Wegen $S_n \sim Bin(n,p)$ ist dies eine kummulative Form von Satz 6.6. Der Beweis verläuft ähnlich wie bei Satz 6.5.

□

Bemerkung 6.8 Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von ZV mit endlichem zweiten Moment. Dann nennt man $(S_n)_{n\in\mathbb{N}}$ mit $S_n := \sum_{i=1}^n X_i$ die Folge der Partialsummen und $S_n^* := \frac{S_n - ES_n}{\sqrt{var(S_n)}}$ die n-te standardisierte Partialsumme.
(Warum "standardisiert"? $ES_n^* = 0$, $var(S_n^*) = 1$)
Man sagt, daß $(X_n)_{n\in\mathbb{N}}$ dem zentralen Grenzwertsatz (ZGWS) genügt, wenn S_n^* mit $n \to \infty$ in Verteilung gegen $N(0,1)$ konvergiert, wenn also gilt: $\lim\limits_{n\to\infty} P(S_n^* < y) = \Phi(y)$ für alle $y \in \mathbb{R}$.
Satz 6.7 zeigt, daß $(X_n)_{n\in\mathbb{N}}$ dem ZGWS genügt, wenn die X_n unabhängig und $Bin(1,p)$-verteilt sind. Sind die X_n unabhängig und Poissonverteilt mit Parameter λ, so ist nach Beispiel 4.26 S_n Poissonverteilt mit Parameter $\lambda_n := n\cdot\lambda$, mit der "Folgenvariante" von Satz 6.5 folgt, daß auch in diesem Fall $(X_n)_{n\in\mathbb{N}}$ dem ZGWS genügt.

In Stochastik II wird gezeigt, daß dies bei unabhängigen, identisch verteilten X_n immer gilt, wenn nur $E{X_i}^2 < \infty$.

Bemerkung 6.9 Wir haben die jeweilige Aussage zur Verteilungskonvergenz aus einer entsprechenden lokalen Aussage hergeleitet, die stärker ist: Nach Satz 6.6 kann die WMassenfunktion zu $Bin(n,p)$ durch die WDichte der Normalverteilung approximiert werden, deren Lage- und Breiteparameter mit dem Erwartungswert und der Varianz von $Bin(n,p)$ übereinstimmen.

Numerische Beispiele

$n = 100$, $p = \frac{1}{2}$ (Münzwurf)

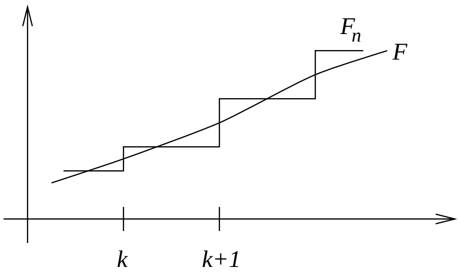
k	40	45	50	99
$P(X=k)$	0.01084	0.04847	0.07959	$0.7889 \cdot 10^{-28}$
Normalapprox.	0.01080	0.04839	0.07979	$0.1115 \cdot 10^{-21}$

und bei $n = 100$, $p = \frac{1}{6}$ (Würfelwurf)

k	5	10	16	17
$P(X=k)$	0.00129	0.02140	0.1065	0.1052
Normalapprox.	0.00080	0.02161	0.1053	0.1066

Grob gilt: Die Approximation für k-Werte in der Nähe von $n\cdot p$ ist besser als "am Rand", und bei gegebenem n ist die Approximation für p-Werte in der Nähe von $\frac{1}{2}$ besser als für p-Werte nahe bei 0 oder 1.

Bemerkung 6.10 Ist F_n die Verteilungsfunktion einer auf $\mathbb{Z}$ konzentrierten Verteilung, so ist F_n auf dem Intervall $[k, k+1)$, $k \in \mathbb{Z}$, konstant. Approximiert man F_n durch eine stetige Verteilungsfunktion F (beispielsweise die einer Normalverteilung), so erhält man häufig eine größere Genauigkeit, wenn man $F_n(k)$ durch $F(k+0.5)$ approximiert (Stetigkeitskorrektur).



Beispiel 6.11 Mit welcher Wahrscheinlichkeit erscheint beim 600-maligen Würfelwurf eines fairen Würfels mindestens 90 und höchstens 105-mal eine 6?

Tatsächlicher Wert: $\sum\limits_{k=90}^{105} \binom{600}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{600-k} = 0.60501\ldots$

Der Satz von deMoivre-Laplace liefert: $P(90 \le S_{600} \le 105) = P(S_{600} \le 105) - P(S_{600} \le 89)$

$= P(S_{600}^* \le \frac{105-100}{\sqrt{\frac{500}{6}}}) - P(S_{600}^* \le \frac{89-100}{\sqrt{\frac{500}{6}}}) \approx \Phi(\frac{5}{\sqrt{\frac{500}{6}}}) - \Phi(\frac{-11}{\sqrt{\frac{500}{6}}}) = 0.5939\ldots$

Mit Stetigkeitskorrektur: $P(90 \le S_{600} \le 105) = P(S_{600} \le 105.5) - P(S_{600} \le 89.5) \approx 0.6015\ldots$

Kapitel 7

Grundbegriffe der Statistik

7.1 Allgemeines

In der WTheorie geht man (stark vereinfacht) von einem Modell $(\Omega, \mathfrak{A}, P)$ für ein Zufallsexperiment aus und berechnet beispielsweise die W. eines Ereignisses A.
In der Statistik soll man, ausgehend von den bei der Ausführung des Experiments gewonnenen Daten, eine Aussage über das zugehörige (unbekannte) P machen.

Beim 10fachen Münzwurf ist beispielsweise eine typische wtheoretische Fragestellung "mit welcher Wahrscheinlichkeit kommt 8-mal Kopf, wenn die Münze fair ist?"
Eine typische statistische Frage ist "Es kam 8 Mal Kopf. Welchen Wert hat p, die W. für Kopf? Ist die Münze fair (d.h. gilt $p = \frac{1}{2}$)"

Allgemeine Struktur: Der Stichprobenraum $\mathfrak{X}$ ist eine Menge, die die möglichen Daten x enthält (in diesem Absatz ist $\mathfrak{X}$ stets "diskret", d.h. endlich oder abzählbar unendlich). Auf $\mathfrak{X}$ hat man eine Familie $\mathfrak{P}$ von WMaßen, die "in Frage kommenden Verteilungen für die Daten". $\mathfrak{P}$ kann die Menge aller WMaße auf $\mathfrak{X}$ sein, hat aber i.A. eine bestimmte Struktur. Im Falle $\mathfrak{P} = \{P_\theta : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^d$ nennt man $\mathfrak{P}$ eine parametrische Klasse, Θ ist die Parametermenge (der Parameterraum). Die Daten $x \in \mathfrak{X}$ können als Realisierung einer Zufallsgröße $X : \Omega \to \mathfrak{X}$ mit unbekannter Verteilung $\mathfrak{L}(\mathfrak{X}) \in \mathfrak{P}$ betrachtet werden. Wird beispielsweise beim 10fachen Münzwurf nur die Anzahl der "Kopf"-Würfe beobachtet, so könnte man $\mathfrak{X} = \{0, 1, \dots, 10\}$, $P_\theta = Bin(10, \theta)$, $\Theta = [0, 1]$ wählen.
Klar: Die Beobachtung $x = 8$ läßt die exakte Bestimmung des unbekannten Parameters θ nicht zu: auf der Basis von zufälligen Beobachtungen lassen sich i.A. keine absolut sicheren (nicht-trivialen) Schlüsse ziehen.[1]

Die hauptsächlichen Verfahren: Schätzen, Tests und Konfidenzbereiche

7.2 Schätztheorie

Ein Schätzer (auch Schätzfunktion) ist eine Abbildung $\hat{\theta} : \mathfrak{X} \to \Theta$, die jeder Beobachtung x einen Schätzwert $\hat{\theta} = \hat{\theta}(x)$ für den unbekannten Parameter θ zuordnet. Naheliegend im Münzwurfbeispiel: $\hat{\theta} = \frac{x}{10}$.
Wie erhält man gute Schätzfunktionen?
Ein plausibles und sehr wichtiges Prinzip besteht darin, daß man den Wert $\hat{\theta}$ wählt, unter dem die Beobachtung x die größte W. hat. Dies ist die Maximum-Likelihood-Methode. ("Maximale Plausibilität")

Definition 7.1 Es sei $\mathfrak{P} = \{P_\theta : \theta \in \Theta\}$ eine Familie von WMaßen auf der abzählbaren Menge $\mathfrak{X}$.
Dann heißt $l(\cdot|x) : \Theta \to \mathbb{R}$, $l(\theta|x) = P_\theta(\{x\})$ die Likelihood-Funktion (bei Vorliegen von $x \in \mathfrak{X}$); $\log P(\cdot|x)$ ist die Log-Likelihood-Funktion. Hat $\hat{\theta} : \mathfrak{X} \to \Theta$ die Eigenschaft $l(\hat{\theta}(x)|x) = \sup\{l(\theta|x) : \theta \in \Theta\}$ $\forall x$, so nennen wir $\hat{\theta}$ einen Maximum-Likelihood-Schätzer für θ (ML).

Es können allerlei Schwierigkeiten auftreten: Das Supremum wird u.U. nicht angenommen, oder ist nicht eindeutig, etc.

Beispiel 7.2 Ein See enthalte eine unbekannte Anzahl N von Fischen. Es werden M Fische gefangen, markiert und wieder freigelassen. Nach einer gewissen Zeit werden n Fische gefangen, unter diesen befinden sich x markierte. Ein Schätzer für N ist gesucht.
Unter gewissen Voraussetzungen (Fische "vermischen" sich, etc.) erscheint das folgende Modell vernünftig: M und n sind bekannt, N ist der unbekannte Parameter und $x \in \mathfrak{X} = \{0, \dots, n\}$ ist die Beobachtung.
Die Beobachtung ist hypergeometrisch verteilt mit den Parametern N, M und n, also $P_N(\{x\}) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$.
Dann gilt: $\frac{P_N(\{x\})}{P_{N-1}(\{x\})} = \frac{\binom{M}{x}\binom{N-M}{n-x}\binom{N-1}{n}}{\binom{N}{n}\binom{M}{x}\binom{N-1-M}{n-x}} = \frac{(N-M)(N-n)}{N(N-M-n+x)}$.

[1] You can't make a silk purse out of a sou's ear ...

Hieraus folgt: $P_N(\{x\}) \underset{>}{\overset{<}{=}} P_{N-1}(\{x\}) \iff (N-M)(N-n) \underset{>}{\overset{<}{=}} N(N-M-n+x) \iff nM \underset{>}{\overset{<}{=}} Nx$ und damit $N \to P_N(\{x\})$ wird maximal für $\hat{N} := \lfloor \frac{nM}{x} \rfloor$ (im Falle $\frac{nM}{x} \in \mathbb{N}$ wird das Maximum in $\hat{N}$ und $\hat{N}-1$ angenommen).

Diesen Schätzer erhält man auch, wenn man davon ausgeht, daß der Anteil der markierten Fische im Fang, $\frac{x}{n}$, ungefähr übereinstimmen sollte mit $\frac{M}{N}$, dem Anteil der markierten Fische im See.

Beispiel 7.3 Ein Zufallsexperiment, in dem ein bestimmtes Ereignis A die unbekannte W. θ hat, wird n-mal unabhängig wiederholt; θ ist zu schätzen.
Schreiben wir 1 für das Eintreten von A und 0 sonst, so sind die resultierenden Daten Elemente von $\mathfrak{X} = \{0,1\}^n$, und die möglichen Verteilungen von P_θ, $0 \le \theta \le 1$ mit $P_\theta(\{(x_1,\dots,x_n)\}) = (1-\theta)^{(n-\sum_{i=1}^n x_i)}\theta^{\sum_{i=1}^n x_i}$. Dies führt auf $l(\theta|x) = (1-\theta)^{(n-\sum_{i=1}^n x_i)}\theta^{\sum_{i=1}^n x_i}$
Wir betrachten zunächst die Randfälle:

- Bei $\sum_{i=1}^n x_i = 0$ erhält man das (eindeutige, globale) Maximum in $\hat{\theta} = 0$,
- bei $\sum_{i=1}^n x_i = n$ erhält man $\hat{\theta} = 1$.
- In den Fällen $\sum_{i=1}^n x_i \in \{1,\dots,n-1\}$ hat man $l(0|x) = l(1|x) = 0$, $l(\theta|x) > 0$ auf $0 < \theta < 1$, und das Maximum kann über Nullsetzen der Ableitung der Log-Likelihood-Funktion[2] gefunden werden:
  $\frac{\partial}{\partial\theta} \log l(\theta|x) = -(n - \sum_{i=1}^n x_i)\frac{1}{1-\theta} + (\sum_{i=1}^n x_i)\frac{1}{\theta} = 0 \iff \theta = \hat{\theta} := \frac{1}{n}\sum_{i=1}^n x_i$.

Als ML ergibt sich also die relative Häufigkeit des Eintretens von A (auch in den Randfällen).

Beispiel 7.4 Eine Urne enthalte N mit den Zahlen 1 bis N numerierte Kugeln. Es werden nacheinander n Kugeln entnommen und zurückgelegt; anhand der beobachteten Zahlen soll N geschätzt werden.
Wir verwenden $\mathfrak{X} = \mathbb{N}^n$ und $\mathfrak{P} = \{P_N : N \in \mathbb{N}\}$, wobei
$P_N(\{(x_1,\dots,x_n)\}) = \begin{cases} N^{-n} & , \text{ falls } \max\{x_1,\dots,x_n\} \le N \\ 0 & , \text{ sonst} \end{cases}$.
Bei festem $n \in \mathbb{N}$ ist N^{-n} als Funktion von N monoton fallend. Dies führt auf den ML-Schätzer $\hat{N}_{ML} = \max\{x_1,\dots,x_n\}$.

Bemerkung 7.5

(i) Unser formales Modell geht von einem "Hintergrund WRaum" $(\Omega, \mathfrak{A}, P)$ aus; die beobachteten Daten $x \in \mathfrak{X}$ werden als Werte (Realisierung) einer Zufallsgröße $X : \Omega \to \mathfrak{X}$ betrachtet (also: Großbuchstaben stehen für die Abbildung selbst, kleine Buchstaben für ihre Werte). Die Verteilung $\mathfrak{L}(X)$ von X ist ein (unbekanntes) Element P von $\mathfrak{P} = \{P_\theta : \theta \in \Theta\}$. Schätzfunktionen sind Abbildungen vom Datenraum $\mathfrak{X}$ in den Parameterraum Θ. Im Falle $\Theta \subseteq \mathbb{R}$ ist also $\hat{\theta} = \hat{\theta}(X)$ eine ZV, deren Erwartungswert die Lage der Verteilung des Schätzers beschreibt. Diese Verteilung (und damit auch der Erwartungswert) hängt von $\theta \in \Theta$ ab: wir schreiben $E_\theta\hat{\theta}$ für den Erwartungswert von $\hat{\theta}(X)$ unter der Voraussetzung, daß $\mathfrak{L}(X) = P_\theta$ gilt (d.h. θ ist der "wahre Parameter").

(ii) Gelegentlich ist man nicht an θ selbst, sondern an einem Wert $\eta = g(\theta)$ interessiert (g heißt Parameterfunktion). Ist $\hat{\theta}_{ML}$ ein Maximum-Likelihood-Schätzer für θ, so nennen wir $\hat{\eta}_{ML} := g(\hat{\theta}_{ML})$ einen ML-Schätzer für η.

(iii) Man kann alle Klassen $\mathfrak{P}$ von WMaßen parametrisieren (beispielsweise durch sich selbst). Wir sprechen beispielsweise vom Schätzer eines Erwartungswertes ohne Bezug auf eine bestimmte Familie $\mathfrak{P}$.
(setze $\Theta = \mathfrak{P}$, $g(P) = E_P X$)

(iv) Man hat häufig $x = (x_1,\dots,x_n)$, wobei die einzelnen Werte Realisierungen von unabhängigen Zufallsgrößen $X_1,\dots,X_n$ sind, die alle dieselbe Verteilung Q haben. Man nennt $X_1,\dots,X_n$ dann eine Stichprobe vom Umfang n aus Q.
Die Beispiele 7.3 und 7.4 haben diese Struktur mit $Q = Bin(1,\theta)$ bzw. $Q = unif\{1,\dots,N\}$.

Im Spezialfall $\Theta \subseteq \mathbb{R}$ kann man die Differenz $\hat{\theta} - \theta$ bilden. Wünschenswert wäre, daß diese klein ist.

Definition 7.6 Es sei $\hat{\eta}$ ein Schätzer für eine reellwertige Parameterfunktion $g(\theta)$.

(i) Der Schätzer $\hat{\theta}$ heißt erwartungstreu (für $\eta = g(\theta)$), wenn gilt: $E_\theta\hat{\eta} = g(\theta)$ für alle $\theta \in \Theta$
(englisch: "unbiased", die Differenz $E_\theta\hat{\eta} - g(\theta)$ wird auch Bias von $\hat{\eta}$ genannt.)

(ii) Die mittere quadratische Abweichung $MSE(\cdot,\hat{\eta})$ von $\hat{\eta}$ wird definiert durch
$MSE(\cdot,\hat{\eta}) : \Theta \to \mathbb{R}^+$, $MSE(\theta,\hat{\eta}) := E_\theta(\hat{\eta} - g(\theta))^2$. ($MSE$ steht für **m**ean **s**quare **e**rror)

Beispiel 7.7

(i) Es sei $X_1,\dots,X_n$ eine Stichprobe aus einer (unbekannten) Verteilung Q mit Erwartungswert μ.
Dann ist $\overline{X_n} := \frac{1}{n}\sum_{i=1}^n X_i$ ein erwartungstreuer Schätzer für μ.
$E_Q\overline{X_n} = \frac{1}{n}\sum_{i=1}^n E_Q X_i = \frac{1}{n}n\mu = \mu$.
Man nennt $\overline{X_n}$ den Stichprobenmittelwert.
Als Stichprobenvarianz bezeichnet man $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X_n})^2$.

[2] da es sich hier um ein Produkt handelt, liegt das nahe ...

Warum $\frac{1}{n-1}$ statt $\frac{1}{n}$? Es gilt:

$$E(\sum_{i=1}^n X_i)^2 = \sum_{i=1}^n EX_i^2 + \sum_{i\neq j} \overbrace{EX_iX_j}^{=EX_i\cdot EX_j} = nEX_1^2 + n(n-1)(EX_1)^2$$

$$\sum_{i=1}^n (X_i - \overline{X_n})^2 = \sum_{i=1}^n X_i^2 - 2\overline{X_n} \underbrace{\sum_{i=1}^n X_i}_{n\overline{X_n}} + n\overline{X_n}^2 = \sum_{i=1}^n X_i^2 - \frac{1}{n}(\sum_{i=1}^n X_i)^2,$$

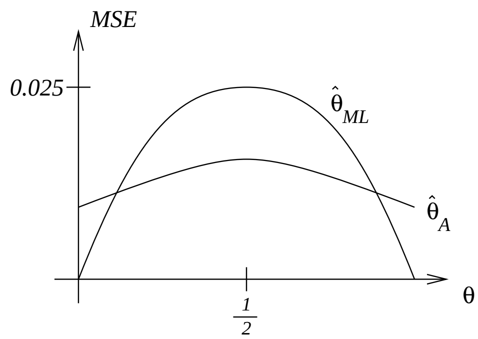
also $ES_n^2 = \frac{1}{n-1}(E(\sum_{i=1}^n X_i^2) - \frac{1}{n}(nEX_1^2 + n(n-1)(EX_1)^2)) = \frac{1}{n-1}((n-1)EX_1^2 - (n-1)(EX_1)^2) = var(X_1)$.

Der Faktor $\frac{1}{n-1}$ führt also auf einen erwartungstreuen Schätzer für die Varianz der Verteilung; mit $\frac{1}{n}$ also Faktor wäre der Schätzer also "im Mittel systematisch zu klein".

(ii) Ist $\hat\theta$ ein erwartungstreuer Schätzer für θ, so ist $\hat\eta := g(\hat\theta)$ nicht unbedingt ein erwartungstreuer Schätzer für $\eta := g(\theta)$. Weiterer Nachteil: In vielen Situationen gibt es keinen erwartungstreuen Schätzer.
Nur in trivialen Fällen existieren Schätzer, die den MSE gleichmäßig minimieren: der "entartete" Schätzer $\hat\theta \equiv \theta_0$, $\theta_0 \in \Theta$ fest, hat MSE 0 in θ_0.
Typisch ist die Situation, die man erhält, wenn man in der Situation von Beispiel 7.3 (Schätzen einer W.) die Schätzer $\hat\theta_{ML} := \frac{1}{n}\sum_{i=1}^n x_i$ und $\hat\theta_A := \frac{1}{n+2}(\sum_{i=1}^n x_i + 1)$ vergleicht.
($\hat\theta_A$ vermeidet, daß die W. von A durch 0 bzw. 1 geschätzt wird, wenn A niemals bzw. immer eingetreten ist.)
$E_\theta\hat\theta_{ML} = \frac{1}{n}\sum_{i=1}^n E_\theta X_i = \theta$, $E_\theta\hat\theta_A = \frac{1}{n+2}(\sum_{i=1}^n E_\theta X_i + 1) = \frac{1}{n+2}(n\theta+1)$ ($\hat\theta_{ML}$ ist also erwartungstreu, $\hat\theta_A$ nicht.)

Bei erwartungstreuen Schätzern ist der MSE gleich der Varianz des Schätzers. Da $\sum_{i=1}^n X_i$ unter P_θ $Bin(n,\theta)$-verteilt ist, erhalten wir: $MSE(\theta,\hat\theta_{ML}) = \frac{1}{n^2}var_\theta(\sum_{i=1}^n X_i) = \frac{\theta(1-\theta)}{n}$,
sowie nach einigen Rechnungen (oder dem Gebrauch von "Mapel") $MSE(\theta,\hat\theta_A) = \frac{1}{(n+2)^2}(1 + n\theta - 4\theta - n\theta^2 + 4\theta^2)$
In $\theta = 0$, $\theta = 1$ hat also $\hat\theta_{ML}$ den kleineren MSE.
Für $\theta = \frac{1}{2}$ erhält man
$MSE(\frac{1}{2};\hat\theta_{ML}) = \frac{1}{4n} > \frac{n}{4(n+2)^2} = MSE(\frac{1}{2};\hat\theta_A)$
Keiner der Schätzer ist also gleichmäßig besser als der andere.

Bild für $n = 10$:



Läßt man nur erwartungstreue Schätzer zu, so kann man (in dieser kleinen Klasse) gelegentlich gleichmäßig beste (im MSE-Sinn also optimale) Schätzer finden.
Der folgende Satz liefert (unter bestimmten Bedingungen) eine Untergrenze für die Varianz (und damit den MSE) eines erwartungstreuen Schätzers.

Satz 7.8 (Die Cramér-Rao- oder Informationsungleichung)
Es sei $\Theta = (a,b)$ mit $-\infty \le a < b \le \infty$, $\hat\theta$ ein erwartungstreuer Schätzer für θ mit existierendem zweiten Moment. Es gelte weiter $p(x|\theta) := P_\theta(\{x\}) > 0 \ \ \forall x \in \mathfrak{X},\ \theta \in \Theta$, sowie

(1) $\frac{d}{d\theta}\sum_{x\in\mathfrak{X}} p(x|\theta) = \sum_{x\in\mathfrak{X}} \frac{\partial}{\partial\theta}p(x|\theta)$.

(2) $\frac{d}{d\theta}\sum_{x\in\mathfrak{X}} \hat\theta(x)p(x|\theta) = \sum_{x\in\mathfrak{X}} \hat\theta(x)\frac{\partial}{\partial\theta}p(x|\theta)$.

Insbesondere sei also $\theta \to p(x|\theta)$ differenzierbar für alle $x \in \mathfrak{X}$.

Schließlich sei $I(\theta) := E_\theta(\frac{\partial}{\partial\theta}\log l(\theta|x))^2 \ \ (= \sum_{x\in\mathfrak{X}} \frac{(\frac{\partial}{\partial\theta}p(x|\theta))^2}{p(x|\theta)}) \ < \infty$

Dann folgt: $var_\theta\hat\theta \ge \frac{1}{I(\theta)}$.

Beweis: Wegen $\sum_{x\in\mathfrak{X}} p(x|\theta) = 1$ ergibt (1) $E_\theta\frac{\partial}{\partial\theta}\log l(\theta|x) = \sum_{x\in\mathfrak{X}} \frac{\frac{\partial}{\partial\theta}p(x|\theta)}{p(x|\theta)}p(x|\theta) = 0$.

Die Erwartungstreue von $\hat\theta$ zusammen mit (2) liefert

$$1 = \frac{d}{d\theta}\theta = \frac{d}{d\theta}E_\theta\hat\theta(x) = \frac{d}{d\theta}\sum_{x\in\mathfrak{X}}\hat\theta(x)p(x|\theta) = \sum_{x\in\mathfrak{X}}\hat\theta(x)(\frac{\partial}{\partial\theta}\log l(\theta|x))p(x|\theta) = E_\theta\hat\theta(x)\frac{\partial}{\partial\theta}\log l(\theta|x) = 1.$$

Verbindet man diese beiden Aussagen, so folgt $E_\theta(\hat\theta(x) - \theta)\frac{\partial}{\partial\theta}\log l(\theta|x) = 1$.

Die Cauchy-Schwarz-Ungleichung (Satz 4.20) liefert nun $var_\theta(\hat\theta(x)) \cdot E_\theta(\frac{\partial}{\partial\theta}\log l(\theta|x))^2 \ge 1$. □

Haben wir eine Stichprobe $X_1,\dots,X_n$ mit einer Verteilung aus $\mathfrak{Q} = \{Q_\theta : \theta \in \Theta\}$, so ist $X = (X_1,\dots,X_n)$ eine Zufallsgröße mit Verteilung aus $\mathfrak{P} = \{P_\theta : \theta\in\Theta\}$, wobei $P_\theta(\{(x_1,\dots,x_n)\}) = \prod_{i=1}^n Q_\theta(\{x_i\})$.

Welcher Zusammenhang besteht zwischen $I_\mathfrak{Q}$ und $I_\mathfrak{P}$? In der naheliegenden Notation gilt $\log l(\theta|X_1,\dots,X_n) = \sum_{i=1}^n \log l(\theta|X_i)$, also $I_\mathfrak{P}(\theta) = E_\theta(\sum_{i=1}^n \frac{\partial}{\partial\theta}\log l(\theta|X_i))^2 = \sum_{i=1}^n E_\theta(\frac{\partial}{\partial\theta}\log l(\theta|X_i))^2 + \sum_{i\neq j} E_\theta(\frac{\partial}{\partial\theta}\log l(\theta|X_i))(\frac{\partial}{\partial\theta}\log l(\theta|X_j)) =$

$n \cdot I_{\mathfrak{Q}}(\theta)$, denn für $i \neq j$ gilt: $E_\theta(\frac{\partial}{\partial\theta}\log l(\theta|X_i))(\frac{\partial}{\partial\theta}\log l(\theta|X_j)) = E_\theta(\frac{\partial}{\partial\theta}\log l(\theta|X_i)) \cdot E_\theta(\frac{\partial}{\partial\theta}\log l(\theta|X_j))$, da X_i, X_j unabhängig sind mit Satz 5.29 und Satz 4.19.
Im Spezialfall $Q = \{(1-\theta)\delta_0 + \theta\delta_1 : 0 < \theta < 1\}$ ergibt sich

$$\log l(\theta|x) = \log((1-\theta)\delta_0 + \theta\delta_1))(\{x\}) = \begin{cases} \log(1-\theta) & , \quad x = 0 \\ \log\theta & , \quad x = 1 \end{cases}, \text{ also } \frac{\partial}{\partial\theta}\log l(\theta|x) = \begin{cases} -\frac{1}{1-\theta} & , \quad x = 0 \\ \frac{1}{\theta} & , \quad x = 1 \end{cases},$$

und damit $I_{\mathfrak{Q}}(\theta) = (\frac{1}{1-\theta})^2(1-\theta) + (\frac{1}{\theta})^2\theta = \frac{1}{\theta(1-\theta)}$.
Dieses läßt sich in der Situation von Beispiel 7.3 (Schätzen einer W. beim n-maligen Ausführen eines Experiments) anwenden.

Die Informationsungleichung liefert als Unterschranke für die Varianz eines erwartungstreuen Schätzers den Wert $(nI_{\mathfrak{Q}}(\theta))^{-1} = \frac{1}{n}\theta(1-\theta)$. Aus den Bemerkungen nach Def. 7.6 ist bekannt, daß diese Unterschranke durch den Schätzer "relative Häufigkeit" angenommen wird.
In der Familie der erwartungstreuen Schätzer ist dieser Schätzer also optimal in dem Sinn, daß er den MSE gleichmäßig auf $(0,1)$ minimiert (in den Randpunkten $\{0,1\}$ hat der Schätzer den MSE 0, ist also auf ganz $[0,1]$ optimal.)
In einigen der obigen Beispiele (7.3, 7.7(i)) haben wir für jeden Stichprobenumfang n einen Schätzer $\hat{\theta}_n$.

Definition 7.9 Eine Folge $(\hat{\eta}_n)_{n\in\mathbb{N}}$ von Schätzern für einen reellwertigen Parameter $\eta = g(\theta)$ heißt konsistent, wenn für alle $\theta \in \Theta$ gilt: $\lim\limits_{n\to\infty} P(|\hat{\eta}_n - \eta| > \epsilon) = 0$ für alle $\epsilon > 0$.

Aus dem schwachen Gesetz der großen Zahlen (Satz 4.28) folgt, daß in der Situation von Beispiel 7.7(i) der Stichprobenmittelwert ein konsistenter Schätzer für den Erwartungswert ist; als Spezialfall hiervon wiederum ergibt sich, daß in der Situation von Beispiel 7.3 die relative Häufigkeit ein konsistenter Schätzer für die Wahrscheinlichkeit ist (genauer: die Folge der relativen Häufigkeiten ist eine konsistente Schätzerfolge.)

Der folgende Satz zeigt, daß Konsistenz unter stetigen Parameterfunktionen erhalten bleibt.

Satz 7.10 Ist $(\hat{\eta}_n)_{n\in\mathbb{N}}$ eine konsistente Schätzerfolge für η und $h : \mathbb{R} \to \mathbb{R}$ stetig auf $\{g(\theta) : \theta \in \Theta\}$, so ist auch $(h(\hat{\eta}_n))_{n\in\mathbb{N}}$ eine konsitente Schätzerfolge für den Parameter $h(\eta)$.
Beweis: Sei $\theta \in \Theta$, $\epsilon > 0$, $\eta := g(\theta)$. Dann existiert ein $\delta > 0$ mit $|\hat{\eta}_n - \eta| < \delta \implies |h(\hat{\eta}_n) - h(\eta)| < \epsilon \quad (*)$, also $P(|h(\hat{\eta}_n) - h(\eta)| > \epsilon) = P(|h(\hat{\eta}_n) - h(\eta)| > \epsilon, |\hat{\eta}_n - \eta| > \delta) + \underbrace{P(|h(\hat{\eta}_n - h(\eta)| > \epsilon, |\hat{\eta}_n - \eta| \leq \delta)}_{\to P(\emptyset)} \leq P(|\hat{\eta}_n - \eta| > \delta) \overset{n\to\infty}{\longrightarrow} 0$, da $(\hat{\eta}_n)_{n\in\mathbb{N}}$ konsistent ist. □

Man kann (mit etwas Aufwand) zeigen, daß die Stichprobenvarianz (siehe Beispiel 7.7(i)) ein konsistenter Schätzer für die Varianz ist, die Wurzel hieraus – die Stichprobenstandardabweichung – ist gemäß Satz 7.10 also ein konsistenter Schätzer für die Standardabweichung.

7.3 Tests

Es sei $\mathfrak{P}$ eine Familie von WMaßen auf dem Stichprobenraum $\mathfrak{X}$, $\mathfrak{X}$ sei abzählbar. Oft soll anhand von Daten entschieden werden, ob die tatsächliche Verteilung P in einer vorgegebenen Teilfamilie $\mathfrak{P}_0$ von $\mathfrak{P}$ liegt, d.h. man will die Hypothese $H : P \in \mathfrak{P}_0$ testen. Bei einer parametrisierten Familie $\mathfrak{P} = \{P_\theta : \theta \in \Theta\}$ läßt sich als Teilfamilie Θ_0 von Θ charakterisieren. Die Hypothese lautet dann $H : \theta \in \Theta_0$, wobei "θ" für den "wahren" Parameter steht. $K : \theta \in \Theta \setminus \Theta_0$ (bzw. $K : P \in \mathfrak{P} \setminus \mathfrak{P}_0$) bezeichnet man Alternative. Man kann H und K auch als Zerlegung von Θ auffassen. H heißt einfach im Fall $\#\mathfrak{P}_0 = 1$ bzw. $\#\Theta_0 = 1$, und zusammengesetzt sonst; analog für K.

Definition 7.11 Eine Funktion $\varphi : \mathfrak{X} \to [0,1]$ heißt (randomisierte) Testfunktion zum Signifikanzniveau α, kurz: Test zum Niveau α, wenn gilt: $E_P\varphi(\mathfrak{X}) \leq \alpha$ für alle $P \in \mathfrak{P}_0$.
Die Abbildung $P \mapsto E_P\varphi(\mathfrak{X})$ ist die Gütefunktion des Tests; im parametrisierten Fall ist dies
$\beta : \Theta \to [0,1]$, $\beta(\theta) = E_\theta\varphi(\mathfrak{X})$.

Interpretation: Bei Vorliegen der Beobachtung $\mathfrak{X}$ wird die Hypothese H mit der Wahrscheinlichkeit $\varphi(\mathfrak{X})$ abgelehnt, bei einem Test zum Niveau α wird die W. für eine irrtümliche Verwerfung der Hypothese nich größer als α. Für α sind die Werte 0.1, 0.05, 0.01 und 0.001 gebräuchlich.
Bei Tests geht es also darum, eine vorgegebene Hypothese anhand der Daten entweder zu verwerfen oder nicht zu verwerfen (beachte: "nicht verwerfen" ist nicht dasselbe wie "als richtig bewiesen".)

Beispiel 7.12 Eine Münze liefert bei einem Wurf mit unbekannter Wahrscheinlichkeit $\theta \in [0,1]$ das Resultat "Kopf". Die Münze wird n-mal geworfen; es soll die Hypothese $H : \theta \leq \frac{1}{2}$ getestet werden. Es liegt nahe, die Hypothese H zu verwerfen, wenn die Anzahl t der "Kopfwürfe" zu groß wird. Verwenden wir wieder $\mathfrak{X} = \{0,1\}^n$ (mit "1" für Kopf), so ist $T(x) = \sum\limits_{i=1}^n x_i$, und man erhält Testfunktionen von der Form $\varphi(x) = \begin{cases} 1 & , \quad \sum\limits_{i=1}^n x_i \geq k \\ 0 & , \quad \text{sonst} \end{cases}$. Da T unter P_θ $Bin(n,\theta)$-verteilt ist, ist $E_\theta\varphi(x) = \sum\limits_{j=k}^n \binom{n}{j}\theta^j(1-\theta)^{n-j}$. Wird die Wahrscheinlichkeit für Kopf größer, so steigt auch die W. dafür, daß bei n Würfen mindestens k mal "Kopf" erscheint; diese Gütefunktion ist also (bei festem

$n \in \mathbb{N},\ k \leq n$) monoton wachsend (kann auch mit Hilfe der Ableitung gesehen werden). Die Testfunktion liefert also genau dann einen Test zum Niveau α, wenn $E_{\frac{1}{2}}\varphi(\mathfrak{X}) \leq \alpha$ gilt. Bei $n = 10$ erhält man beispielsweise die Werte 0.0108 bei $k = 9$ und 0.0546 bei $k = 8$. Gibt man also das Signifikanzniveau $\alpha = 0.05$ vor, so erhält man bei $k = 9$ einen Test zum Niveau α, der dieses Niveau nicht voll "ausschöpft". Bei $k = 8$ wird das vorgegebene Niveau überschritten.

θ: W. für "Kopf", $H : \theta \leq \frac{1}{2}$, T: Anzahl der "Kopfwürfe", $T(x) = \sum\limits_{i=1}^{n} x_i$, Testfunktion $\varphi(x) = \begin{cases} 1 & , \quad \sum\limits_{i=1}^{n} x_i \geq k \\ 0 & , \quad \sum\limits_{i=1}^{n} x_i < k \end{cases}$.

Gütefunktion $E_\theta\varphi(x) = \sum\limits_{i=k}^{n} \binom{n}{i}\theta^i(1-\theta)^{n-i}$ (W. für Ablehnung als Funktion des unbekannten Parameters). Bei einem Test zum Niveau α muß $E_\theta\varphi(x) \leq \alpha \quad \forall \theta \in H$ gelten.

"Erweiterung": $\varphi(x) = \begin{cases} 1 & , \quad > \\ \gamma & , \quad \sum\limits_{i=1}^{n} x_i = k \\ 0 & , \quad < \end{cases}$ (Randomisierung: bei $T(x) = k$ wird mit W. γ abgelehnt.)

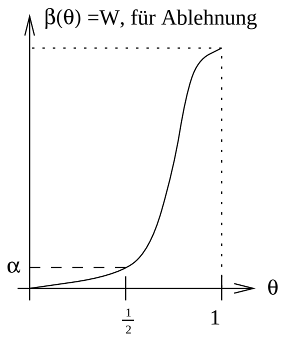
Bei $n = 10,\ k = 8$ erhält man mit $\gamma = \frac{\alpha - P_{0.5}(T>8)}{P_{0.5}(T=8)} \approx 0.8933$.

$E_{0.5}\varphi(x) = 0 \cdot P_{0.5}(\varphi(x) = 0) + \gamma \cdot P_{0.5}(\varphi(x) = \gamma) + 1 \cdot P_{0.5}(\varphi(x) = 1) = \alpha - P_{0.5}(T > 8) + P_{0.5}(T > 8) = \alpha$

Das führt auf folgende Vorschrift: erhält man beim 10fachen Münzwurf 9- oder 10-mal "Kopf", so ist $H : \theta \leq \frac{1}{2}$ zu verwerfen. Erhält man genau 8-mal "Kopf", so führt man ein weiteres Zufallsexperiment aus, in dem ein bestimmtes Ereignis A die W. γ hat. Tritt A ein, so ist H zu verwerfen, sonst nicht. Bei 7 oder weniger "Kopfwürfen" ist H nicht zu verwerfen. Insgesamt erhält man damit einen Test zum Niveau $\alpha = 0.05$, der dieses Niveau auch voll auschöpft.

Im obigen Beispiel liefert die Testgröße (auch Teststatistik) T, die die Eigenschaft hat, daß große Werte von T gegen die Hypothese H sprechen, eine ganze Familie von Tests φ_c, $\varphi_c = \begin{cases} 1 & , \quad T(x) > c \\ 0 & , \quad T(x) < c \end{cases}$. Man nennt c den kritischen Wert, die Menge $\{x \in \mathfrak{X} : T(x) > c\}$ heißt Ablehnung- oder Verwerfungsbereich.

Bei Tests geht es um zwei Entscheidungen: H wird verworfen oder H wird nicht verworfen. Als Konsequenz hiervon gibt es zwei Fehlerarten:
Fehler 1. Art: Die Hypothese wird verworfen, obwohl sie richtig ist.
Fehler 2. Art: Die Hypothese wird nicht verworfen, obwohl sie falsch ist.
Die Wahrscheinlichkeit für eine falsche Entscheidung hängt von dem unbekannten Parameter θ ab. Bei einem Test zum Niveau α ist die W. für einen Fehler 1. Art höchstens α. FehlerW. lassen sich an der Gütefunktion ablesen.
Man wird versuchen, bei vorgegebener Schranke für den Fehler 1. Art einen Test zu finden, bei dem die W. für einen Fehler 2. Art möglichst gleichmäßig minimiert wird. Bei einfacher Hypothese und einfacher Alternative (also $\#\mathfrak{P} = 2$) kann man dieses Optimierungsproblem lösen:



Satz 7.13 (Das Neyman-Pearson-Lemma)
Es sei $\mathfrak{P} = \{P_0, P_1\},\ \alpha \in (0,1)$. Sei $p_i(x) := P_i(\{x\})$.
Dann existieren $c \geq 0$ und $\gamma \in [0,1]$ mit $P_0(\{x \in \mathfrak{X} : p_1(x) > c \cdot p_0(x)\}) + \gamma \cdot P_0(\{x \in \mathfrak{X} : p_1(x) = c \cdot p_0(x)\}) = \alpha$, und der Test $\varphi(x) = \begin{cases} 1 & , \quad p_1(x) > c \cdot p_0(x) \\ \gamma & , \quad p_1(x) = c \cdot p_0(x) \\ 0 & , \quad p_1(x) < c \cdot p_0(x) \end{cases}$ ist ein Test zum Niveau α für die Hypothese $H : P = P_0$, der unter allen solchen Tests die kleinste W. für einen Fehler 2. Art hat.

Beweis: Für alle $t \geq 0$ sei $\alpha_l(t) := P_0(p_1 \geq t \cdot p_0)$, $\alpha_r(t) := P_0(p_1 > t \cdot p_0)$.
Klar: $\alpha_l(0) = 1,\ \alpha_l(t) \geq \alpha_r(t) \quad \forall t \geq 0$, α_l und α_r sind (schwach) monoton fallend.

Sei $C := \{\frac{p_1(x)}{p_0(x)} : x \in \mathfrak{X},\ p_0(x) > 0\} = \{c_1, c_2, \dots\}$ (da "alles diskret").

$A_i := \{x \in \mathfrak{X} : \frac{p_1(x)}{p_0(x)} = c_i,\ p_0(x) > 0\}$.

Dann gilt: $\sum P_0(A_i) = 1$, also $\lim\limits_{t\to\infty} P_0(p_1 \leq t \cdot p_0) = \lim\limits_{t\to\infty} \sum\limits_{i \in \mathbb{N},\ c_i \leq t} P_0(A_i) = 1$ (Stetigkeit von unten von P_0)

Also: $\lim\limits_{t\to\infty} \alpha_r(t) = 0$. Setze nun $c := \inf\{t \geq 0 : \alpha_r(t) \leq \alpha\}$.
Dann gilt für alle $k \in \mathbb{N}$: $\alpha \geq \alpha_r(c + \frac{1}{k}) = \sum\limits_{c_i > c + \frac{1}{k}} P_0(A_i) \underset{k\to\infty}{\longrightarrow} \alpha_l(c)$, da P_0 stetig von unten.

$$\alpha_r(c - \tfrac{1}{k}) = \sum_{i \in \mathbb{N},\ c_i > c - \frac{1}{k}} P_0(A_i) \longrightarrow \alpha_l(c) \text{ (Stetigkeit von oben).}$$

Wäre $\alpha_l(c) < \alpha$, so gäbe es ein $k \in \mathbb{N}$ mit $\alpha_r(c - \frac{1}{k}) \leq \alpha$, im Widerspruch zur Definition von c. Also $\alpha_l(c) \geq \alpha \geq \alpha_r(c)$.
Mit $\gamma := \frac{\alpha - \alpha_r(c)}{\alpha_l(c) - \alpha_r(c)}$ gilt dann $\gamma \in [0,1]$ und es folgt $\underbrace{P_0(p_1 > c \cdot p_0) + \gamma P_0(p_1 = c \cdot p_0)}_{E_0\varphi} = \alpha_r(c) + \gamma(\alpha_l(c) - \alpha_r(c)) = \alpha$.

Damit ist der 1. Teil (Existenz) bewiesen.

Hieraus folgt auch $E_0\varphi = \alpha$.
Sei nun $\tilde{\varphi}$ irgendein Test zum Niveau α für $H : P = P_0$, sei $A := \{x \in \mathfrak{X} : \varphi(x) > \tilde{\varphi}(x)\}$, $B := \{x \in \mathfrak{X} : \varphi(x) < \tilde{\varphi}(x)\}$. Auf A ist $\varphi(x) > 0$, also $p_1(x) \geq c \cdot p_0(x)$, auf B ist $\varphi(x) < 1$, also $p_1(x) \leq c \cdot p_0(x)$. Damit $E_1\varphi(x) - E_1\tilde{\varphi}(x) = \sum_{x\in\mathfrak{X}} (\varphi(x) - \tilde{\varphi}(x))p_1(x) = \sum_{x\in A} (\varphi(x) - \tilde{\varphi}(x))p_1(x) + \sum_{x\in B} (\varphi(x) - \tilde{\varphi}(x))p_1(x) \geq \sum_{x\in A} (\varphi(x) - \tilde{\varphi}(x))cp_0(x) + \sum_{x\in B} (\varphi(x) - \tilde{\varphi}(x))cp_0(x) = c \sum_{x\in\mathfrak{X}} (\varphi(x) - \tilde{\varphi}(x))p_0(x) = c \cdot (\underbrace{E_0\varphi(x)}_{=\alpha} - \underbrace{E_0\tilde{\varphi}(x)}_{\leq\alpha}) \geq 0$.
In Worten: Der Neyman-Pearson-Test φ hat eine Ablehnungswahrscheinlichkeit, die, wenn A falsch ist, mindestens so groß ist, wie die von $\tilde{\varphi}$. □

Der optimale Test hängt also nun über das Verhältnis $\frac{p_1(x)}{p_0(x)}$ von x ab. Der Ablehnungsbereich entsteht dadurch, daß man x-Werte mit größtem solchen Verhältnis zusammenfaßt, soweit dies die Fehlerschranke α erlaubt.

Beispiel 7.14 Es sei wieder wie im Beispiel 7.12 $\mathfrak{X} = \{0,1\}^n$, $P_\theta(\{(x_1,\dots,x_n)\}) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$. Wir betrachten zunächst die Familie $\mathfrak{P} = \{P_{\theta_0}, P_{\theta_1}\}$ mit festen $0 < \theta_0 < \theta_1 < 1$. Sei wieder $T(x) = \sum_{i=1}^n x_i$. Als Verhältnis der Massenfunktionen erhält man $\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} = \frac{\theta_1{}^{T(x)}(1-\theta_1)^{n-T(x)}}{\theta_0{}^{T(x)}(1-\theta_0)^{n-T(x)}} = \underbrace{\left(\frac{\theta_1}{\theta_0}\right)}_{>1}^{T(x)} \underbrace{\left(\frac{1-\theta_1}{1-\theta_0}\right)}_{<1}^{n-T(x)}$.
Dies ist eine streng monoton wachsende Funktion von $T(x)$ (wegen $\theta_0 < \theta_1$),

$$\text{d.h. } \forall c \in \mathbb{R}\ \exists \tilde{c} \in \mathbb{R}\ \forall x \in \mathfrak{X}\ p_{\theta_1}(x) \begin{array}{c}>\\=\\<\end{array} cp_{\theta_0}(x) \iff T(x) \begin{array}{c}>\\=\\<\end{array} \tilde{c}.$$

Nach dem Neyman-Pearson-Lemma ist der beste Test für θ_0 gegen θ_1 von der Form

$$(1) \quad \varphi(x) = \begin{cases} 1 & , & > \\ \gamma & , \quad \sum_{i=1}^n x_i & = \tilde{c} \\ 0 & , & < \end{cases},$$

wobei γ und $\tilde{c}$ bestimmt werden aus (2) $P_{\theta_0}(\sum x_i > \tilde{c}) + \gamma P_{\theta_0}(\sum x_i = \tilde{c}) = \alpha$.
Beachte nun: in (1) und (2) tritt θ_1 nicht mehr auf, nur $\theta_1 > \theta_0$ wurde zur Herleitung verwendet: $H : \theta = \theta_0$ gegen $K : \theta = \hat{\theta}_1$ ($\hat{\theta}_1 > \theta_0$) führt auf denselben Test. φ minimiert also die Fehlerwahrscheinlichkeit 2. Art auf $\theta_1 > \theta_0$, d.h. φ ist gleichmäßig bester Test zum Niveau α für $H : \theta = \theta_0$ gegen $K : \theta > \theta_0$. Da $E_\theta\varphi$ eine monoton wachsende Funktion von θ ist, ist φ auch Test für $H : \theta \leq \theta_0$ gegen $K : \theta > \theta_0$, sogar gleichmäßig bester.

Bemerkung 7.15

(i) In der Parxis wird selten randomisiert.

(ii) Bei nicht-randomisierten Tests mit einer Testgröße T, d.h. bei Ablehnungsbereichen der Form $\{x : T(x) \geq c\}$, wird häufig der p-Wert angegeben; dies ist der kleinste Wert α, der noch eine Verwerfung von H ergibt. Hat man beispielsweise in Beispiel 7.12 9-mal "Kopf" erhalten, so würde man bei $\alpha > 0.0108$ verwerfen und bei $\alpha < 0.0108$ nicht verwerfen, d.h. der p-Wert ist 0.0108.

(iii) Tests sind "unsymmetrisch": man hat nur für eine Fehlerart eine obere Schranke (nämlich das Signifikanzniveau, für das irrtümliche Ablehnen der Hypothese). Dies ist bei der Wahl der Hypothese zu berücksichtigen. Das Nicht-Verwerfen einer Hypothese bedeutet nicht, daß diese Hypothese als richtig bewiesen wurde.

7.4 Konfidenzbereiche

Definition 7.16

Es sei $\mathfrak{P} = \{P_\theta : \theta \in \Theta\}$ eine Familie von WMaßen auf $\mathfrak{X}$. Eine Abbildung $C : \mathfrak{X} \to \mathbb{P}(\Theta)$ heißt Konfidenzbereich zum Konfidenzniveau $1-\alpha$ (auch $100(1-\alpha)\%$ Konfidenzgebiet), wenn gilt: $P(C(\mathfrak{X}) \ni \theta) \geq 1-\alpha$ für alle $\theta \in \Theta$.

Ist im Falle $\Theta \subseteq \mathbb{R}$ $C(x)$ ein Intervall für alle $x \in \mathfrak{X}$, so nennt man $C(x)$ ein Konfidenzintervall. Im Falle $C(x) = (-\infty, c(x)]$ heißt $c(x)$ eine obere Konfidenzschranke etc. Für α sind die Werte 0.1, 0.05, 0.01, 0.001 gebräuchlich.

Beispiel 7.17

Wie in Beispiel 7.4 werden einer Urne mit N Kugeln n Kugeln mit zurücklegen entnommen, Y sei das Maximum der gezogenen Werte.
Dann gilt für alle $n \in \mathbb{N}$: $P_N(Y \leq N \leq \lfloor Y\alpha^{-\frac{1}{n}}\rfloor) = 1 - P_N(N \leq \lfloor Y\alpha^{-\frac{1}{n}}\rfloor) \geq P_N(Y > N\alpha^{\frac{1}{n}})$

$$= 1 - P_N(X_i \leq N\alpha^{\frac{1}{n}} \text{ für alle } i = 1,\dots,n) = 1 - \left(\tfrac{N\alpha^{\frac{1}{n}}}{N}\right)^n = 1-\alpha.$$

Also: $\{Y, Y+1, \dots, \lfloor Y\alpha^{-\frac{1}{n}}\rfloor\}$ ist ein $100(1-\alpha)\%$ Konfidenzgebiet für N.

Bemerkung 7.18

Hat man in der Situation von 7.17 Kugeln mit den Nummern 4, 14, 3, 22, 14 gezogen, so ergibt sich $\{22, 23, \dots, 40\}$ als zugehöriges 95% Konfidenzintervall. Ein weit verbreiteter Fehler ist es, hier den Schluß zu ziehen, daß N mit W. 0.95 zwischen 22 und 40 liegt: beachte, daß $C(x)$ Realisierung einer (intervallwertigen) Zufallsgröße $C(X)$ ist, die W.

0.95 bezieht sich auf $C(X)$. Analog ist beim Werfen eines fairen Würfels die W. dafür, daß die Augenzahl X den Wert 1 annimmt gleich $\frac{1}{6}$. Wird das Experiment ausgeführt, und ist das Resultat der Wert 3, so heißt dies nicht, daß 3 mit W. $\frac{1}{6}$ gleich 1 ist. Der analoge Fall bei der Testtheorie besteht in der Behauptung, H sei mit Wahrscheinlichkeit $1-\alpha$ falsch.

Zwischen Ablehnungbereichen von Tests einfacher Hypothesen und Konfidenzbereichen besteht ein nützlicher Zusammenhang:

Satz 7.19 Für jedes $\theta_0 \in \Theta$ sei $A(\theta_0) \subseteq \mathfrak{X}$ der Ablehnungsbereich eines nicht-randomisierten Tests zum Niveau α von $H: \theta = \theta_0$. Dann ist $C(x) := \{\theta \in \Theta : A(\theta) \not\ni x\}$ ein Konfidenzbereich zum Niveau $1-\alpha$ (dies ist auch umkehrbar.)
Beweis: $P_\theta(\{x \in \mathfrak{X} : C(x) \ni \theta\}) = P_\theta(\{x \in \mathfrak{X} : A(\theta) \not\ni x\}) = 1 - P_\theta(A(\theta)) \geq 1-\alpha$. □

7.5 Statistische Anwendungen der Normalapproximation

7.5.1 Konfidenzintervalle für Wahrscheinlichkeiten

Es seien $X_1, X_2, \ldots$ unabhängig und $Bin(1,\theta)$-verteilt mit unbekanntem $\theta \in (0,1)$. Wir verwenden $\overline{X_n} := \frac{1}{n}\sum_{i=1}^n X_i$ als Schätzer fur θ (siehe Beispiel 7.3 und 7.12). Nach dem Satz von Moivre-Laplace gilt mit $S_n = \sum_{i=1}^n X_i = n\overline{X_n}$

$P(a \leq \frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leq b) \overset{n\to\infty}{\longrightarrow} \Phi(b) - \Phi(a)$, wobei Φ wieder die Verteilungsfunktion zu $N(0,1)$ bezeichnet.

Ist u_α das α-Quantil zu $N(0,1)$ (also $\Phi(u_\alpha) = \alpha$), so folgt mit $b := u_{1-\frac{\alpha}{2}},\ a := -b$:

$P(-u_{1-\frac{\alpha}{2}} \leq \frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leq u_{1-\frac{\alpha}{2}}) \approx 1-\alpha$ (bei großem n.)

Wegen $\theta(1-\theta) \leq \frac{1}{4}$ ergibt sich:
$\overline{X_n} - \frac{u_{1-\frac{\alpha}{2}}}{2\sqrt{n}} \leq \theta \leq \overline{X_n} + \frac{u_{1+\frac{\alpha}{2}}}{2\sqrt{n}}$, also ergibt sich $[\overline{X_n} - \frac{1}{2\sqrt{n}}u_{1-\frac{\alpha}{2}}, \overline{X_n} + \frac{1}{2\sqrt{n}}u_{1-\frac{\alpha}{2}}]$ also asymptotisches, konservatives $100(1-\alpha)\%$ Konfidenzintervall für θ.
In jedem Fall erhält man Intervalle, deren Länge mit $\frac{1}{\sqrt{n}}$ fällt. Für eine weitere Dezimalstelle muß man den Stichprobenumfang also verhundertfachen.

Zahlenbeispiel: Soll bei einer Wahl ein Konfidenzintervall für die Anzahl der Stimmen einer Partei von der Form "Prozentsatz in der Stichprobe $\pm 1\%$" auf dem Niveau 0.95 erhalten werden, so muß $\frac{1}{2\sqrt{n}}\underbrace{u_{0.975}}_{\approx 1.96} \leq 0.01$ gelten, also $n \geq 9604$.
Für $\pm 0.1\%$ erhält man $n \geq 960400$.

7.5.2 Kritische Bereiche für zweiseitige Tests

In der in Absatz 7.5.1 betrachteten Situation soll nun $H: \theta = \theta_0$ ($\theta_0 \in (0,1)$ bekannt, fest) getestet werden. Es liegt nahe, die Hypothese dann zu verwerfen, wenn $\overline{X_n}$ zu weit von θ_0 abweicht, also einen kritischen Bereich von der Form $|\overline{X_n} - \theta_0| \geq c$ zu wählen, wobei c von n, θ_0 und α abhängt. Die Normalapproximation erlaubt eine näherungsweise Bestimmung von c: $P_{\theta_0}(|\overline{X_n} - \theta_0| \geq c) = 1 - P(-\frac{c\sqrt{n}}{\sqrt{\theta_0(1-\theta_0)}} < \frac{S_n - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} < \frac{c\sqrt{n}}{\sqrt{\theta_0(1-\theta_0)}}) \approx 1 - \Phi(\frac{c\sqrt{n}}{\sqrt{\theta_0(1-\theta_0)}}) + \Phi(-\frac{c\sqrt{n}}{\sqrt{\theta_0(1-\theta_0)}}) = 2\Phi(-\frac{c\sqrt{n}}{\sqrt{\theta_0(1-\theta_0)}})$
(wir unterstellen hierbei, daß n so groß ist, daß W. von der Form $P_{\theta_0}(|\overline{X_n} - \theta_0| = c)$ vernachlässigt werden können (vgl. Aufgabe 39)). Man wird also bei vorgegebenem Signifikanzniveau α als kritische Schranke den Wert $c = \frac{1}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\sqrt{\theta_0(1-\theta_0)}$ wählen, denn dann gilt:

$$P(\mathit{Ablehnung}) \approx 2\Phi\left(\frac{\left(-\frac{1}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\sqrt{\theta_0(1-\theta_0)}\right)\sqrt{n}}{\sqrt{\theta_0(1-\theta_0)}}\right) = 2\Phi(-u_{1-\frac{\alpha}{2}}) = 2(1-\Phi(u_{1-\frac{\alpha}{2}})) = \alpha.$$

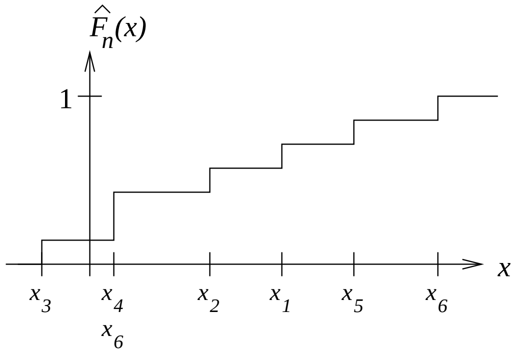
Was läßt sich über den Fehler 2. Art aussagen?
Sei $\theta \neq \theta_0$; $\epsilon := \frac{1}{2}|\theta - \theta_0|$. Aus $|\overline{X_n} - \theta| \leq \epsilon$ folgt dann $|\overline{X_n} - \theta_0| \geq |\theta - \theta_0| - |\overline{X_n} - \theta| \geq \epsilon \geq \frac{1}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\sqrt{\theta_0(1-\theta_0)}$ für n groß genug (so daß die Normalapproximation "legal" ist).
Mit dem schwachen Gesetz der großen Zahlen folgt nun:
$1 = \lim_{n\to\infty} P_\theta(|\overline{X_n} - \theta| \leq \epsilon) \leq \liminf_{n\to\infty} P_\theta(|\overline{X_n} - \theta_0| \geq \frac{1}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\sqrt{\theta_0(1-\theta_0)}) \leq 1$, also $P_\theta(H$ wird verworfen$) = 1 \quad \forall \theta \neq \theta_0$.
Ist also die Hypothese falsch, so geht die W. für eine richtige Entscheidung bei wachsendem Stichprobenumfang gegen 1. Man nennt Tests mit dieser Eigenschaft konsistent.

7.6 Ausblick: Bootstrap-Konfidenzintervalle

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus einer Verteilung mit unbekannter Verteilungsfunktion F. Wie schätzt man F? Für jedes $x \in \mathbb{R}$ ist $F(x) \quad (= P(X_i \leq x))$ eine W., die man durch die zugehörige relative Häufigkeit

$\frac{1}{n}\#\{1 \leq i \leq n : X_i \leq x\} = \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,x]}(X_i)$ schätzen kann.

Dies führt auf die <u>empirische Verteilungsfunktion</u> $\hat{F}_n : \mathbb{R} \to \mathbb{R}$; $\hat{F}_n(x) := \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,x]}(X_i)$ als Schätzer für F.



$\hat{F}_n$ ist eine zufällige Verteilungsfunktion, die der Gleichverteilung auf den Werten der Stichprobe. Die Konvergenz von $\hat{F}_n$ gegen F mit $n \to \infty$ (in einem starken Sinne) ist die Aussage des <u>Satzes von Glivenko-Cantelli</u> (in der älteren Literatur: "Hauptsatz der Stochastik").
Interessiert man sich für einen Parameter θ der Verteilung, $\theta = \Psi(F)$ (wobei Ψ bekannt ist), so liegt es nahe, θ durch $\hat{\theta}_n := \Psi(\hat{F}_n)$ zu schätzen (Einsetzprinzip). Ist θ beispielsweise der zu F gehörende Erwartungswert, so ist $\hat{\theta}_n$ der zu $\hat{F}_n$ gehörende Erwartungswert, also: $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ (Stichprobenmittelwert als Schätzer für den Erwartungswert).
Analog kann man Quantile $F^{-1}(\alpha)$ durch die Stichprobenquantile $\hat{F}_n^{-1}(\alpha)$ schätzen.

Der Bootstrap ist eine (auf dem Einsetzprinzip beruhende) Methode zum Erhalt von Konfidenzintervallen für einen Parameter θ, der durch $\hat{\theta}$ geschätzt wird.
<u>Vorüberlegung</u>: Ist die Verteilungsfunktion R von $\hat{\theta} - \theta$ bekannt, also auch Quantile r_α, so gilt:
$P(r_{\frac{\alpha}{2}} \leq \hat{\theta} - \theta \leq r_{1-\frac{\alpha}{2}}) = P(\hat{\theta} - \theta \leq r_{1-\frac{\alpha}{2}}) - P(\hat{\theta} - \theta \leq r_{\frac{\alpha}{2}}) = R(r_{1-\frac{\alpha}{2}}) - R(r_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = \alpha.$
Wegen $\hat{\theta} - \theta \leq r_{1-\frac{\alpha}{2}} \iff \theta \geq \hat{\theta} - r_{1-\frac{\alpha}{2}}$ folgt, daß $[\hat{\theta} - r_{1-\frac{\alpha}{2}}, \hat{\theta} - r_{\frac{\alpha}{2}})$ ein $100(1-\alpha)\%$ Konfidenzintervall für θ ist.
R ist natürlich unbekannt. <u>Idee</u>: Schätze R.
Hierzu werden künstliche Stichproben (resample) aus den Originaldaten $x_1, \ldots, x_n$ konstruiert.

$$\left.\begin{array}{ccc} x_{11}^* & ,\ldots, & x_{1n}^* \\ x_{21}^* & ,\ldots, & x_{2n}^* \\ \vdots & \vdots & \vdots \\ x_{B1}^* & ,\ldots, & x_{Bn}^* \end{array}\right\}$$

insgesamt B Stück, unabhängig, ziehen mit zurücklegen

B ist groß (10000 oder 100000)

23	10	0	1.1	2	5	$\frac{41.1}{6}$
2	5	5	1.1	1.1	0	$\frac{14.2}{6}$
0	0	1.1	23	10	10	$\frac{44.1}{6}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Dies geschieht auf einem Computer mit einem Pseudozufallszahlengenerator.
Bei jeder Stichprobe wird der Wert $\eta_k := \theta_k^* - \hat{\theta}$ berechnet, $1 \leq k \leq B$, wobei θ_k^* den Schätzwert bezeichnet, der zum Resample $x_{k1}^*, \ldots, x_{kn}^*$ gehört. Die empirische Verteilungsfunktion $\hat{R}_B$ zu diesen Werten $\eta_1, \ldots, \eta_B$ wird als Schätzer für R verwendet, also $(\hat{R}_B)^{-1}(\alpha)$ als Schätzer für r_α. Insgesamt erhält man also das Bootstrap-Konfidenzintervall $[\hat{\theta} - (\hat{R}_b)^{-1}(1 - \frac{\alpha}{2}), \hat{\theta} - (\hat{R}_B)^{-1}(\frac{\alpha}{2}))$ zum (näherungsweise, es wurde ja geschätzt) Konfidenzniveau $1 - \alpha$ für θ.
Diese Methode ist (z.Z.) sehr populär, da sie mit minimalen Voraussetzungen auskommt und sehr allgemein verwendbar.
Funktioniert nicht immer: Wirkliches Verständnis erfordert "jede Menge Mathematik".

Index