Vorlesungsmitschrift:
“Stochastik I + II”

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

(SS 96 und WS 96/97)

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

1 Grundbegriffe, die Axiome von Kolmogorov **1**
1.1 Ergebnisraum und Ergebnisalgebra . . . 1
1.2 Wahrscheinlichkeit . . . 2

2 Bedingte Wahrscheinlichkeiten und Unabhängigkeit **5**

3 Laplace-Experimente **7**
3.1 Allgemeines . . . 7
3.2 Etwas Kombinatorik . . . 7
3.3 Lösung einiger typischer Probleme . . . 8
3.3.1 Paradox von de Méré . . . 8
3.3.2 Geburtstagsproblem . . . 9
3.3.3 Bridge . . . 9
3.3.4 Der zerstreute Postbote . . . 10

4 Diskrete WRäume und Zufallsgrößen **11**
4.1 Allgemeines . . . 11
4.2 Einige wichtige Verteilungen . . . 11
4.2.1 Binomial- und Bernoulli-Verteilung . . . 11
4.2.2 Poisson-Verteilung . . . 12
4.2.3 Geometrische Verteilung und negative Binomialverteilung . . . 12
4.2.4 Hypergeometrische Verteilung . . . 12
4.2.5 Multinomialverteilung . . . 13
4.3 Erwartungswert und Varianz von Zufallsgrößen . . . 13
4.4 Bedingte Verteilungen und Unabhängigkeit . . . 14
4.5 Reellwertige (diskrete) Zufallsgrößen . . . 15
4.6 Das schwache Gesetz der großen Zahlen . . . 17

5 Allgemeine Wahrscheinlichkeitsräume **19**
5.1 “You can’t always get what you want” . . . 19
5.2 Mengensysteme . . . 19
5.3 Zufallsgrößen und Verteilungen . . . 21
5.4 Reellwertige Zufallsgrößen . . . 22
5.5 Verteilungsfunktionen . . . 23
5.6 Einige wichtige Verteilungen mit Riemann-Dichten . . . 24
5.6.1 Gleich- bzw. Rechteck-Verteilung . . . 24
5.6.2 Gamma- und Exponential-Verteilung . . . 25
5.6.3 Normalverteilung . . . 25
5.7 Erwartungswerte . . . 25
5.8 Unabhängigkeit . . . 26

6 Verteilungskonvergenz und Normalapproximation **29**
6.1 Verteilungskonvergenz . . . 29
6.2 Normalapproximation bei Poisson-Verteilungen . . . 30
6.3 Normalapproximation bei der Binomialverteilung . . . 31

7 Grundbegriffe der Statistik **33**
7.1 Allgemeines . . . 33
7.2 Schätztheorie . . . 33
7.3 Tests . . . 36
7.4 Konfidenzbereiche . . . 38
7.5 Statistische Anwendungen der Normalapproximation . . . 39
7.5.1 Konfidenzintervalle für Wahrscheinlichkeiten . . . 39
7.5.2 Kritische Bereiche für zweiseitige Tests . . . 39
7.6 Ausblick: Bootstrap-Konfidenzintervalle . . . 39

8 Integral und Erwartungswert **41**
8.1 Einführung und Erinnerung . . . 41
8.2 Konstruktion des Lebesgue-Integrals . . . 41
8.3 Transformationsformel, Nullmengen . . . 43
8.4 Konvergenzsätze . . . 44
8.5 Maße und Dichten . . . 45
8.6 Erwartungswerte . . . 45
8.7 Ungleichungen . . . 46

9 Produktmaß und Unabhängigkeit **47**
9.1 Das Produkt von zwei Maßräumen . . . 47
9.2 Produkte von mehr als zwei Maßräumen . . . 48
9.3 Unabhängigkeit . . . 49

10 Gesetze der großen Zahlen **51**
10.1 Konvergenz fast-sicher und Konvergenz in Wahrscheinlichkeit . . . 51
10.2 Das schwache Gesetz der großen Zahlen . . . 51
10.3 Das starke Gesetz der großen Zahlen . . . 52

11 Charakteristische Funktion **55**
11.1 Grundlegendes . . . 55
11.2 Ableitungen und Momente . . . 56
11.3 Umkehrsätze . . . 56
11.4 Faltungen . . . 57

12 Verteilungskonvergenz **59**
12.1 Erinnerung, Zusammenhang zu anderen Konvergenzbegriffen . . . 59
12.2 Straffheit und charakteristische Funktionen . . . 60

13 Der Zentrale Grenzwertsatz (ZGWS) **63**
13.1 Identisch verteilte Summanden . . . 63
13.2 Der Satz von Lindeberg . . . 64
13.3 Anwendungen . . . 65
13.3.1 Ein Sammelproblem . . . 65
13.3.2 Rekorde . . . 66
13.3.3 Primteiler . . . 67
13.3.4 Maximum-Likelihood-Schätzer . . . 67

14 Zufallsvektoren **69**
14.1 Allgemeines . . . 69
14.2 Mehrdimensionale Normalverteilungen . . . 69
14.3 Grenzwertsätze . . . 71

15 Martingale **73**
15.1 Bedingte Erwartungswerte . . . 73
15.2 Martingale . . . 75
15.3 Stoppzeiten und Transformationen . . . 77
15.4 Konvergenzsätze . . . 79
15.4.1 Fast sichere Konvergenz . . . 79
15.4.2 Gleichgradige Integrierbarkeit . . . 79
15.4.3 Gleichgradig integrierbare Martingale . . . 81
15.4.4 Ein Rückwärtskonvergenzsatz . . . 82
15.5 Anwendungen . . . 82
15.5.1 Ein 0-1-Gesetz von Kolmogorov . . . 82

15.5.2 Das starke Gesetz der großen Zahlen 83
15.5.3 Der Satz von Radon-Nikodym 83
15.5.4 Verzweigungsprozesse 83

... 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.
$\overline{A}, A^c$: 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 $\Omega \neq \emptyset$ 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) $P(A) \geq 0 \qquad \forall A \in \mathfrak{A}$
(iii) $P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \qquad$ 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 $\text{Laplace} - \text{Experiment}$ 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) $P(\emptyset) = 0, \ P(A) \leq 1 \qquad \forall A \in \mathfrak{A}$
(ii) $P(A^c) = 1 - P(A) \qquad \forall A \in \mathfrak{A}$
(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) $P(A) \geq 0 \qquad \forall A \in \mathfrak{A}$

(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.

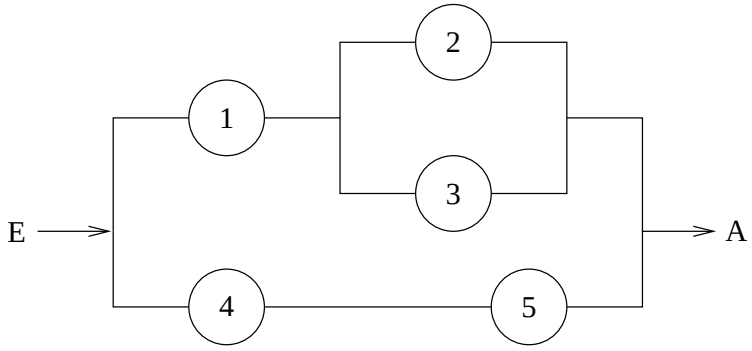
Beispiel 2.6 (paarweise unabhängigig $\not\Rightarrow$ unabhängig)
Eine faire Münze wird zweimal geworfen ($\rightarrow$ Beispiel 1.1.(i)).
Laplace-Experiment $\Omega = \{(0,0),(0,1),(1,0),(1,1)\}$

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: $\Omega := \bigtimes_{i=1}^{n} \Omega_i = \Omega_1 \times \ldots \times \Omega_n = \{\omega = (\omega_1, \ldots, \omega_n) : \omega_i \in \Omega_i \text{ für } i = 1, \ldots, n\}$. 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 $M_n = \{1, \ldots, n\}$
Regel 2 liefert $\#{M_n}^k = \#\{(i_1, \ldots, i_k) : 1 \leq i_j \leq n \ \forall j\} = n^k$
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 $(i_1, \ldots, i_k)$ 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 $(i_1, \ldots, i_k) \in {M_n}^k$ 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: $\#\{(i_1, \ldots, i_k) \in {M_n}^k : i_l \neq i_j \text{ für } l \neq j\} = \frac{n!}{(n-k)!}$.
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: $\#\{(i_1, \ldots, i_k) \in {M_n}^k : i_1 < i_2 < \ldots < i_k\} = \binom{n}{k}$.
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 $\#\{(i_1, \ldots, i_k) \in {M_n}^k : i_1 \leq \ldots \leq i_k\} = \binom{n+k-1}{k}$ k-Kombinationen mit Wiederholungen.
Beweis: Wir definieren eine bijektive Abbildung

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

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.
$A_1, \ldots, A_4$ sind unabhängig und haben W. $\frac{1}{6}$, also:
$P(A) = 1 - P(A^c) = 1 - P(A_1^c \cap \ldots \cap A_4^c) = 1 - P(A_1^c) \cdot \ldots \cdot P(A_4^c)$
$= 1 - (1 - P(A_1))(1 - P(A_2))(1 - P(A_3))(1 - P(A_4)) = 1 - (1 - \frac{1}{6})^4 \approx 0.518\ldots$

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

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 $\Omega := \{(i_1, \ldots, i_n) : 1 \leq i_1, \ldots, i_n \leq 365\} = \{1, \ldots, 365\}^n$ plausibel, wobei $i_j = k$ bedeutet, daß Person j am Tag k Geburtstag hat.
$A := \{(i_1, \ldots, i_n) \in \Omega : \exists l \neq j \text{ mit } i_l = i_j\}$.
Man hat $A^c = \{(i_1, \ldots i_n) \in \Omega : i_l \neq i_j \text{ für } l \neq j\}$. 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 $P(A) \approx 0.97\ldots$

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.
$\Omega' := \{(\omega_1, \ldots, \omega_{52}) \in \{1, \ldots, 52\}^{52} : \omega_i \neq \omega_j \text{ für } i \neq j\}$, also Menge der Permutationen von $\{1, \ldots, 52\}$. 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 $\Omega := \{(D_1, D_2, D_3, D_4) : D_i \subseteq \{1, \ldots, 52\}\ \forall i, \#D_i = 13, \text{ disjunkt}\}$ 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 $\{1, \ldots, 52\} \setminus D_1$, 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:

$$P(A) = P(A_1) + \ldots + P(A_4) = 4P(A_1) \approx 0.01056\ldots$$

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

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

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 $\Omega_n := \{(\omega_1, \ldots, \omega_n) \in \{1, \ldots n\}^n : \omega_i \neq \omega_j \text{ für } i \neq j\}$ 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:

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

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.
□

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

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:

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

$EY = \sum_{y \in \mathbb{R}} y p_Y(y) = \sum_{x \in \mathbb{R}} f(x) p_X(x)$, 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 $p_i := \frac{\Pi_i}{\sum_j \Pi_i}$, 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: $X \le |X|,\ -X \le |X| \implies EX \le E|X|,\ -EX = E(-X) \le E|X|$, also: $|EX| \le E|X|$. □

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 $\{0,\dots,4\} \times \{0,\dots,4\}$ 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: Die Tabelle enthält die Werte $20825 \cdot P(X = k, Y = l)$

	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

Aus den Werte n der Tabelle ergibt sich mit $P(X = i) = P(X = i, Y = 0) + P(X = i, Y = 1) + \dots + P(X = i, Y = 4)$ für $i = 0,\dots,4$, 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.:
$P(X = Y) = P(X = 0, Y = 0) + \dots + P(X = 4, Y = 4) = \frac{1}{20825}(1150 + 4225 + 468 + 0 + 0) = 0.280576\dots$
"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 bedingte Verteilung von Y unter $X = x$, Schreibweise $P^{Y|X=x}$ oder $\mathfrak{L}(Y|X = x)$.
Ist $S_2 = \mathbb{R}$ und $\sum_{y \in \mathbb{R}} |y| P^{Y|X=x}(\{y\}) < \infty$, so nennen wird

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

den bedingten Erwartungswert von Y unter $X = x$.
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: $P(X = k, Y = B) = \prod_{j \in B} p \prod_{j \notin B} (1 - p) = p^k (1 - p)^{n-k}$.

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 $\frac{1}{3}$.
Beachte: 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:

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

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 stochastisch unabhängig, 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:
$P(X_{i_1} = x_{i_1}, \ldots, X_{i_n} = x_{i_n}) = \prod_{j=1}^{n} P(X_{i_j} = x_{i_j})$ 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:

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

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 nicht.

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 $A_{xy} := \{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\}$, $x \in Bild(X)$, $y \in Bild(Y)$, 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 $(EXY)^2 \le EX^2\cdot EY^2$.

Beweis: Wegen $|(X\cdot Y)(\omega)| = |X(\omega)||Y(\omega)| \le X(\omega)^2 + Y(\omega)^2$ gilt:

$\sum_{\omega\in\Omega} |(XY)(\omega)|P(\{\omega\}) \le \underbrace{\sum X(\omega)^2P(\{\omega\})}_{EX^2<\infty} + \underbrace{\sum Y(\omega)^2P(\{\omega\})}_{EY^2<\infty}$, also existiert der Erwartungswert.

Für beliebiges $t\in\mathbb{R}$ existiert dann auch das zweite Moment zu $X+tY$, und ist nicht-negativ:

$0 \le E(X+tY)^2 = EX^2 + t^2EY^2 + 2tEXY \ \forall t\in\mathbb{R}$.

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 $\frac{1}{EY^2}(EX^2EY^2 - (EXY)^2)$.

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 $\sigma_X,\ \sigma_Y > 0$, 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) $-1 \le \varrho(X,Y) \le 1$

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) $\sigma_X^2\sigma_Y^2\varrho(X,Y)^2 = (E(E-EX)(Y-EY))^2 \le E(X-EY)^2E(Y-EY)^2 = var(X)var(Y)$

$\Longrightarrow \varrho(X,Y)^2 \le 1$

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 $X_1,\dots,X_n$ ZV mit existierendem zweiten Moment. Dann gilt:

$var(X_1+\dots+X_n) = \sum_{i=1}^n var(X_i) + \sum_{i\ne j} cov(X_i,X_j)$.

Sind $X_1,\dots,X_n$ 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:

$var(\sum_{i=1}^n X_i) = E(\sum_{i=1}^n X_i)^2 - (E(\sum_{i=1}^n X_i))^2 = \sum_{i,j=1}^n EX_iX_j - \sum_{i,j=1}^n EX_iEX_j$

$= \sum_{i=1}^n EX_i^2 - (EX_i)^2 + \sum_{i\ne j}(EX_iX_j - EX_iEX_j) = \sum_{i=1}^n var(X_i) + \sum_{i\ne j} cov(X_i,X_j)$.

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 $X_1,\dots,X_n$ 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 $S_n = X_1+\dots+X_n$:
$ES_n = \sum_{i=1}^n EX_i = np,\ var(S_n) = \sum_{i=1}^n var(X_i) = np(1-p)$(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 $\{1,\dots,n\}$ (siehe 3.3.4)
Sei $X_i(\omega) = \begin{cases} 1 & \omega_i = i \\ 0 & \omega_i \ne i \end{cases}$ $\forall\omega = (\omega_1,\dots,\omega_n)\in\Omega_n$ (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 $X_1, \ldots, X_n$ 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.: $\underbrace{r_k \geq 0}_{\text{trivial}}$, $\sum r_n = 1$

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

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

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

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

$X+Y$ 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 $n \in \mathbb{N}_0$, $k \in \{0, \ldots, n\}$ gilt:

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

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

4.6 Das schwache Gesetz der großen Zahlen

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

$$E\overline{X_n} = \tfrac{1}{n}\sum_{i=1}^{n} EX_i = \mu$$
$$var(\overline{X_n}) = \tfrac{1}{n^2} var(\sum_{i=1}^{n} X_i) = \tfrac{1}{n^2}\sum_{i=1}^{n} var(X_i) = \tfrac{\sigma^2}{n}$$

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 $p = 2$:
$P(|X - EX| \geq \alpha) = P(|Y| \geq \alpha) \leq \frac{1}{\alpha^2} EY^2 = \frac{1}{\alpha^2} E(X - EX)^2 = \frac{1}{\alpha^2} var(X)$ □

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
$P(|\overline{X_n} - \mu| > \epsilon) \leq \frac{1}{\epsilon^2} var(\overline{X_n}) \longrightarrow 0 \quad \forall \epsilon > 0$ (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 $[a,b) = \bigcap_{n=1}^{\infty} \underbrace{\bigcup_{m=1}^{\infty} (a - \frac{1}{n}, b - \frac{1}{m}]}_{(a-\frac{1}{n}, b)}$, und $\mathfrak{B} \subset \sigma(\mathfrak{E}_1)$ durch $(a,b] = \bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} [a + \frac{1}{n}, b + \frac{1}{m})$

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

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: $U = \bigcup_{\{(a,b) \in \mathbb{Q}^2 | (a,b] \subset U\}} (a,b]$. 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 $(a,b] = \bigcap_{n=1}^{\infty} (a, b + \frac{1}{n})$.

Dieser Satz impliziert, daß die Intervalle $[a,b)$, $(-\infty, a]$ Borel-Mengen sind, ebenso alle offenen und abgeschlossenen Mengen. Wegen $\{a\} = \bigcap_{n=1}^{\infty} (a - \frac{1}{n}, a]$ (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 $0 \le a < b < 1$.

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 $P_{\mathbb{R}}(B) := P(B \cap [0,1)) \quad \forall B \in \mathfrak{B}_{\mathbb{R}}$.
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 $P_A(B) := \underbrace{P(A \cap B)}_{=B} \quad \forall B \in \mathfrak{B}_A$.

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 $P(\{x\}) = \lim_{n \to \infty} P([x, x + \frac{1}{n})) = 0$ 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 $\mu(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ 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 $i \neq j \implies \bigcup_{i=1}^{\infty} A_i \in \mathfrak{D}$

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
$\delta(\mathfrak{E}) = \bigcap_{\substack{\mathfrak{D} \supset \mathfrak{E} \\ \mathfrak{D} \text{ Dynkin-System}}} \mathfrak{D}$ 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.: $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{D}$.

Setze $B_1 := A_1$, $B_n := A_n \cap A_1{}^c \cap \ldots \cap A_{n-1}^c$ $(n \geq 1)$. Durchschnittstabilität 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 ein WMaß auf $\mathfrak{B}_{[0,1)}$ mit der Eigenschaft (5.2); wir können also von der 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 Zufallsgröße (auf $(\Omega, \mathfrak{A}, P)$ und mit Werten in $(\Omega', \mathfrak{A}')$), wenn X $(\mathfrak{A}, \mathfrak{A}')$-meßbar 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 stetig, 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 Verteilung von X, 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; $X : \Omega \to \mathbb{R}$, $X(\omega) = \min\{\omega, 1-\omega\}$.
$P(X < \frac{1}{4}) = P(X^{-1}(\underbrace{(0, \frac{1}{4})}_{\in \mathfrak{B}})) = P((0, \frac{1}{4}) \cup (\frac{3}{4}, 1)) = \frac{1}{2}$.
$X^{-1}((0, \frac{1}{4})) = (0, \frac{1}{4}) \cup (\frac{3}{4}, 1) \in \mathfrak{A}$

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:
$X^{-1}(\Omega') = \Omega \in \mathfrak{A}$, 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) $\{\omega \in \Omega : \sup_{n\in\mathbb{N}} X_n(\omega) \le a\} = \bigcap_{n=1}^{\infty} \underbrace{\{\omega \in \Omega : X_n(\omega) \le a\}}_{\in \mathfrak{A}}$

Die Meßbarkeit der anderen Abbildungen folgt nun mit $\inf_{n\in\mathbb{N}} X_n = -\sup_{n\in\mathbb{N}}(-X_n)$, $\limsup_{n\to\infty} X_n = \inf_{n\in\mathbb{N}} \sup_{m\ge n} X_m$, $\liminf_{n\to\infty} X_n = -\limsup_{n\to\infty}(-X_n)$.

Konvergiert X_n mit $n \to \infty$ punktweise gegen X, so gilt $X = \limsup_{n\to\infty} X_n$, 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) $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$
(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 $\lim_{n\to\infty} x_n = -\infty$. Setze $y_n := \sup_{m\ge n} x_m$. 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 $\bigcap_{n=1}^{\infty} A_n = \emptyset$):
$0 \le F(x_n) \le F(y_n) = P((-\infty, y_n]) \longrightarrow 0$ für $n \to \infty$, 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)).

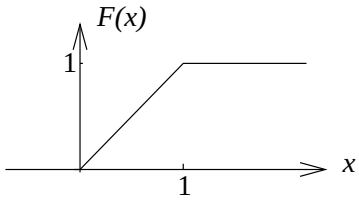
(iii) Ist $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ mit $x_n \ge x$, $x_n \to x$, so gilt für $y_n := \sup_{m\ge n} x_m$, $y_m \downarrow x$, also
$F(x) = P((-\infty, x]) \le P((-\infty, x_n]) \le P((-\infty, y_n]) \longrightarrow P((-\infty, x]) = F(x)$,
da die $(-\infty, y_n]_{n\in\mathbb{N}}$ 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.

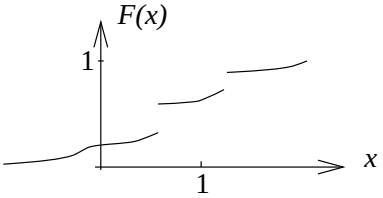
Definition 5.19 Es sei F eine Funktion mit den Eigenschaften (i)-(iii) aus Satz 5.18. Dann definieren wir die Quantilfunktion Q zu F durch $Q : (0,1) \to \mathbb{R}$, $Q(y) := \inf\{x \in \mathbb{R} : F(x) \ge y\}$ (wir schreiben auch F^{-1} für Q).

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.

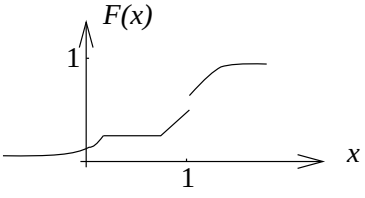
Im Falle $X \sim unif(0,1)$ erhält man: $F(X) = P(X \le x) = \begin{cases} 0 & , \quad x < 0 \\ x & , \quad 0 \le x \le 1 \\ 1 & , \quad x > 1 \end{cases}$



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

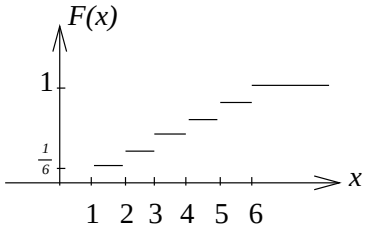
Beweis: "$\Longrightarrow$" folgt unmittelbar aus der Definition von F^{-1}.
Da außerdem $F(x) < y \implies F(x+\frac{1}{n}) < y$ für ein $n \in \mathbb{N}$, denn f ist stetig von rechts $\implies F^{-1}(y) \geq x + \frac{1}{n}$, denn F schwach monoton steigend $\implies F^{-1}(y) > x$ gilt, hat man auch die Gegenrichtung. □

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

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.

Beispiel: Würfelwurf

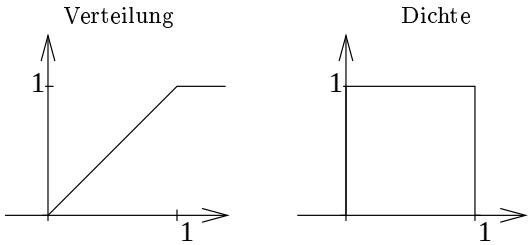


Ist $f : \mathbb{R} \to \mathbb{R}_+$ eine Funktion mit $\int_{-\infty}^{\infty} f(x)dx = 1$, so wird nach dem obigen Resultat durch $P((-\infty, x]) := \int_{-\infty}^{x} f(y)dy$ für alle $x \in \mathbb{R}$ ein WMaß auf $(\mathbb{R}, \mathfrak{B})$ definiert, das WMaß mit (Riemann-) Dichte f.

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

(Hierbei ist $\Gamma(z) := \int_0^\infty x^{z-1}e^{-x}dx$ 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 $\varphi_{\mu,\sigma^2} := \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2),\ x \in \mathbb{R}$.

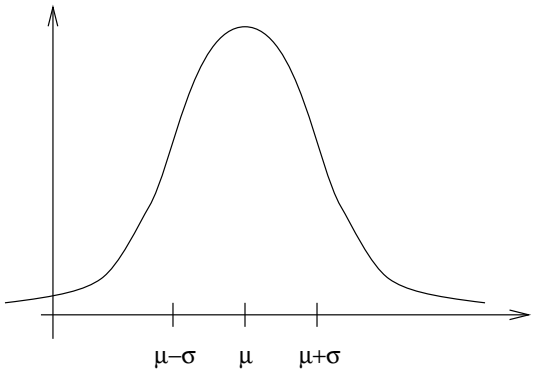
Die Parameter μ und σ^2 beschreiben die Lage und Breite von φ.
Im Falle $\mu = 0,\ \sigma^2 = 1$ spricht man von den Standardparametern, $N(0,1)$ ist die Standardverteilung.
Offensichtlich gilt: $\varphi_{\mu,\sigma^2} = \frac{1}{\sigma}\varphi_{0,1}(\frac{x-\mu}{\sigma}) \quad \forall x \in \mathbb{R}$.

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

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

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

(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:
$$(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}dx)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-\tfrac{1}{2}(x^2+y^2))dxdy = \int_0^{2\pi}\int_0^{\infty} \underbrace{r\exp(-\frac{1}{2}r^2)}_{-\frac{d}{dr}\exp(-\frac{1}{2}r^2)}\, drd\varphi = 2\pi.$$

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

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

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

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 $x \in \mathbb{R}$ $\lceil x \rceil := \min\{k \in \mathbb{Z} : k \ge x\},\ \lfloor x \rfloor := \max\{k \in \mathbb{Z} : k \le x\}$, 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
$\underline{X_n} \le X \le \overline{X_n},\ \overline{X_n} - \underline{X_n} \le 2^{-n}$.

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

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

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 $EX = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2)dx$

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

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 $\Omega = X^{-1}(\Omega')$, $X^{-1}(A^c) = (X^{-1}(A))^c$, $X^{-1}(\bigcup_{i=1}^{\infty} A_n) = \bigcup_{i=1}^{\infty} X^{-1}(A_n)$.

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 $P(\bigcap_{j\in J} E_j) = \prod_{j\in J} P(E_j)$ 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
$$P(A_{j_1} \cap A_{j_2} \cap \ldots \cap A_{j_n}) = P(A_{j_1}) \cdot P(A_{j_2}) \cdot \ldots \cdot P(A_{j_n}) \text{ für alle } A_{j_1} \in \mathfrak{A}_{j_1}, \ldots, A_{j_n} \in \mathfrak{A}_{j_n}. \quad \square$$

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 $\omega = \sum_{n=1}^{\infty} 2^{-n} X_n(\omega) = 0.X_1(\omega)X_2(\omega)\ldots$ ist "die" Binärdarstellung von ω.
Für alle $k_1, \ldots k_n \in \{0,1\}$ gilt: $P(X_1 = k_1, \ldots, X_n = k_n) = P(\{\omega \in \Omega : \sum_{l=1}^{n} 2^{-l}k_l \leq \omega < \sum_{l=1}^{n} 2^{-l}k_l + 2^{-n}\}) = 2^{-n}$ (das

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

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

$= 2^{-i_n} \#\{(k_1, \ldots, k_{i_n}) \in \{0,1\}^{i_n} : k_{i_j} = 1 \text{ für } j = 1, \ldots, n\} = 2^{-i_n} \cdot 2^{i_n - n} = 2^{-n}$, denn n Positionen sind festgelegt.
Insbesondere folgt $P(X_{i_j} = 1) = \frac{1}{2}$ und damit $P(X_{i_1} = 1, \ldots, X_{i_n} = 1) = P(X_{i_1} = 1) \cdot \ldots \cdot P(X_{i_n} = 1)$.
Da $\{X_i^{-1}(\{1\})\}$ ein $\cap$-stabiles Erzeugendensystem von $\sigma(X_i)$ ist, ist damit gezeigt, daß die ZV $X_1, X_2, \ldots$ 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 $k_1, k_2 \ldots$ 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 $X_1, \ldots, X_n$ seien unabhängig und $unif(0,1)$-verteilt. Wir interessieren uns für die Verteilung des Minimums $Y_n := \min\{X_1, \ldots, X_n\}$.

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

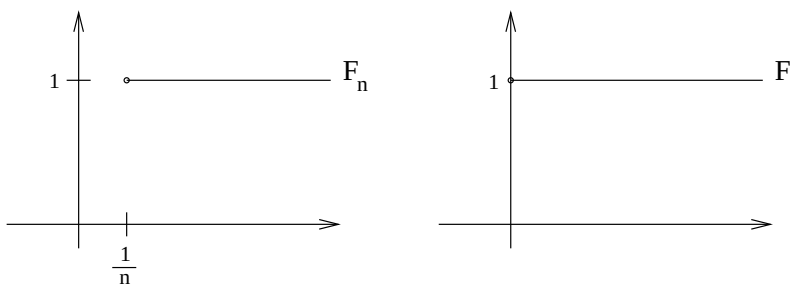
$$1 - P(X_1 > y) \cdot \ldots \cdot P(X_n > y) = 1 - \prod_{i=1}^{n}(1 - F_{X_i}(y)) = \begin{cases} 1-(1-y)^n & , \quad 0 \leq y \leq 1 \\ 0 & , \quad y < 0 \\ 1 & , \quad y > 1 \end{cases} .$$

Mit $n \to \infty$ erhalten wir den (punktweisen) Limes $\begin{cases} 1 & , \quad y > 0 \\ 0 & , \quad y \leq 0 \end{cases}$ (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:
$F_{nY_n}(y) = P(nY_n \leq y) = P(Y_n \leq \frac{y}{n}) = 1 - (1 - \frac{y}{n})^n \underset{n\to\infty}{\longrightarrow} 1 - e^{-y}$, 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 $k \in \mathbb{N}_0,\ \lambda > 0$
$g(k|\lambda) := \log P_\lambda(X=k) = -\lambda + k\log\lambda - \log k!$
$z(k,\lambda) := \frac{1}{\sqrt{\lambda}}(k-\lambda).$
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 $\log(1+x) = x + O(x^2)\quad (x\to 0)$ 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:
(*) $\quad \lim\limits_{\lambda\to\infty}\sqrt{2\pi\lambda}h(\lambda) = 1$
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 $\int\limits_{-\infty}^{\infty}\varphi(x)dx = 1$: $\quad \limsup\limits_{\lambda\to\infty}\sqrt{2\pi\lambda}h(\lambda) \leq 1.$
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

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

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.
$\frac{e^{-\lambda}\frac{\lambda^k}{k!}}{\frac{1}{\sqrt{2\pi\lambda}}\exp(-\frac{1}{2\lambda}(k-\lambda)^2)} \overset{\lambda\to\infty}{\longrightarrow} 1$ 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:

$$\lim_{\lambda\to\infty}P(a\leq\tfrac{X_\lambda-\lambda}{\sqrt{\lambda}}\leq b) = \lim_{\lambda\to\infty}\sum_{\lambda+a\sqrt{\lambda}\leq k\leq\lambda+b\sqrt{\lambda}}P(X_\lambda=k) = \lim_{\lambda\to\infty}\sum_{a\leq\frac{k-\lambda}{\sqrt{\lambda}}\leq b}\tfrac{1}{\sqrt{\lambda}}\varphi(\tfrac{k-\lambda}{\sqrt{\lambda}}) = \int_a^b\varphi(x)dx = \Phi(b)-\Phi(a),$$

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 $\lim_{\lambda\to\infty} P(\frac{X_\lambda-\lambda}{\sqrt{\lambda}}\leq b)=\Phi(b)$ 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:

$$\begin{aligned}\binom{n}{k}p^k(1-p)^{n-k}&\sim\frac{\varphi(k|\lambda,\lambda)\varphi(n-k|\mu,\mu)}{\varphi(n|\lambda+\mu,\lambda+\mu)}=\frac{\frac{1}{\sqrt{2\pi\lambda}}\exp(-\frac{1}{2\lambda}(k-\lambda)^2)\frac{1}{\sqrt{2\pi\mu}}\exp(-\frac{1}{2\mu}\exp(-\frac{1}{2\mu}(n-k-\mu)^2)}{\frac{1}{\sqrt{2\pi(\lambda+\mu)}}\exp(-\frac{1}{2(\lambda+\mu)}(n-\lambda-\mu)^2)}\\&=\frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{np\cdot n(1-p)}}\exp(-\tfrac{1}{2np}(k-np)^2-\tfrac{1}{2n(1-p)}(n-k-n(1-p))^2)=\frac{1}{\sqrt{2\pi np(1-p)}}\exp(-\tfrac{1}{2n}(\tfrac{1}{p}+\tfrac{1}{1-p})(k-np)^2)\\&=\varphi(k|np,np(1-p)).\end{aligned}$$

□

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: $\lim_{n\to\infty}P(S_n^*<y)=\Phi(y)$ 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 $EX_i{}^2<\infty$.

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)

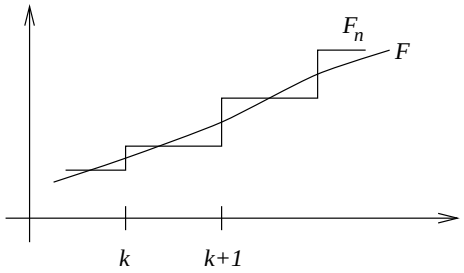
k	40	45	50	99
$P(X=k)$	0.01084	0.04847	0.07959	$0.7889\cdot10^{-28}$
Normalapprox.	0.01080	0.04839	0.07979	$0.1115\cdot10^{-21}$

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

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

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

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

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 $l(\hat{\theta}(x)|x) = \sup\{l(\theta|x) : \theta \in \Theta\} \quad \forall x$, 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: $P_N(\{x\}) \begin{matrix}<\\=\\>\end{matrix} P_{N-1}(\{x\}) \iff (N-M)(N-n) \begin{matrix}<\\=\\>\end{matrix} N(N-M-n+x) \iff nM \begin{matrix}<\\=\\>\end{matrix} Nx$ 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:

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

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

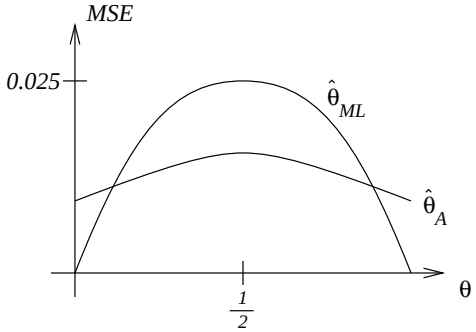
also $ES_n^2 = \frac{1}{n-1}(E(\sum_{i=1}^n X_i^2) - \frac{1}{n}(nEX_1^2 + n(n-1)(EX_1)^2)) = \frac{1}{n-1}((n-1)EX_1^2 - (n-1)(EX_1)^2) = var(X_1)$.

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 $p(x|\theta) := P_\theta(\{x\}) > 0 \quad \forall x \in \mathfrak{X},\ \theta \in \Theta$, 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 $I_\mathfrak{P}(\theta) = E_\theta(\sum_{i=1}^n \frac{\partial}{\partial\theta}\log l(\theta|X_i))^2 = \sum_{i=1}^n E_\theta(\frac{\partial}{\partial\theta}\log l(\theta|X_i))^2 + \sum_{i\neq j} E_\theta(\frac{\partial}{\partial\theta}\log l(\theta|X_i))(\frac{\partial}{\partial\theta}\log l(\theta|X_j)) =$

$n \cdot I_\Omega(\theta)$, 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 $I_\Omega(\theta) = (\frac{1}{1-\theta})^2(1-\theta) + (\frac{1}{\theta})^2\theta = \frac{1}{\theta(1-\theta)}$.
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 $(nI_\Omega(\theta))^{-1} = \frac{1}{n}\theta(1-\theta)$. 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

$n \in \mathbb{N},\ k \le n$) monoton wachsend (kann auch mit Hilfe der Ableitung gesehen werden). Die Testfunktion liefert also genau dann einen Test zum Niveau α, wenn $E_{\frac{1}{2}}\varphi(\mathfrak{X}) \le \alpha$ 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", $H : \theta \le \frac{1}{2}$, T: Anzahl der "Kopfwürfe", $T(x) = \sum_{i=1}^{n} x_i$, Testfunktion $\varphi(x) = \begin{cases} 1 & , \quad \sum_{i=1}^{n} x_i \ge k \\ 0 & , \quad \sum_{i=1}^{n} x_i < k \end{cases}$.

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

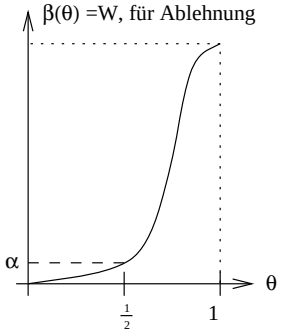
"Erweiterung": $\varphi(x) = \begin{cases} 1 & , \quad > \\ \gamma & , \quad \sum_{i=1}^{n} x_i = k \\ 0 & , \quad < \end{cases}$ (Randomisierung: bei $T(x) = k$ wird mit W. γ abgelehnt.)

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 $H : \theta \le \frac{1}{2}$ 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 $\varphi_c,\ \varphi_c = \begin{cases} 1 & , \quad T(x) > c \\ 0 & , \quad T(x) < c \end{cases}$. 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 $c \ge 0$ 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 $t \ge 0$ sei $\alpha_l(t) := P_0(p_1 \ge t \cdot p_0),\ \alpha_r(t) := P_0(p_1 > t \cdot p_0)$.
Klar: $\alpha_l(0) = 1,\ \alpha_l(t) \ge \alpha_r(t) \quad \forall t \ge 0,\ \alpha_l$ 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 $\lim_{t\to\infty} P_0(p_1 \le t \cdot p_0) = \lim_{t\to\infty} \sum_{i\in\mathbb{N},\ c_i \le t} P_0(A_i) = 1$ (Stetigkeit von unten von P_0)
Also: $\lim_{t\to\infty} \alpha_r(t) = 0$. Setze nun $c := \inf\{t \ge 0 : \alpha_r(t) \le \alpha\}$.
Dann gilt für alle $k \in \mathbb{N}$: $\alpha \ge \alpha_r(c + \frac{1}{k}) = \sum_{c_i > c+\frac{1}{k}} P_0(A_i) \underset{k\to\infty}{\longrightarrow} \alpha_l(c)$, da P_0 stetig von unten.
$\alpha_r(c - \frac{1}{k}) = \sum_{i\in\mathbb{N},\ c_i > c-\frac{1}{k}} P_0(A_i) \longrightarrow \alpha_l(c)$ (Stetigkeit von oben).
Wäre $\alpha_l(c) < \alpha$, so gäbe es ein $k \in \mathbb{N}$ mit $\alpha_r(c - \frac{1}{k}) \le \alpha$, im Widerspruch zur Definition von c. Also $\alpha_l(c) \ge \alpha \ge \alpha_r(c)$.
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 $E_1\varphi(x) - E_1\tilde{\varphi}(x) = \sum\limits_{x\in\mathfrak{X}} (\varphi(x) - \tilde{\varphi}(x))p_1(x) = \sum\limits_{x\in A} (\varphi(x) - \tilde{\varphi}(x))p_1(x) + \sum\limits_{x\in B} (\varphi(x) - \tilde{\varphi}(x))p_1(x) \geq \sum\limits_{x\in A} (\varphi(x) - \tilde{\varphi}(x))cp_0(x) + \sum\limits_{x\in B} (\varphi(x) - \tilde{\varphi}(x))cp_0(x) = c \sum\limits_{x\in\mathfrak{X}} (\varphi(x) - \tilde{\varphi}(x))p_0(x) = c \cdot (\underbrace{E_0\varphi(x)}_{=\alpha} - \underbrace{E_0\tilde{\varphi}(x)}_{\leq\alpha}) \geq 0.$

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 $\mathfrak{X} = \{0,1\}^n$, $P_\theta(\{(x_1,\ldots,x_n)\}) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$. Wir betrachten zunächst die Familie $\mathfrak{P} = \{P_{\theta_0}, P_{\theta_1}\}$ mit festen $0 < \theta_0 < \theta_1 < 1$. Sei wieder $T(x) = \sum\limits_{i=1}^{n} x_i$. Als Verhältnis der Massenfunktionen erhält man $\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} = \frac{{\theta_1}^{T(x)}(1-\theta_1)^{n-T(x)}}{{\theta_0}^{T(x)}(1-\theta_0)^{n-T(x)}} = \Big(\underbrace{\frac{\theta_1}{\theta_0}}_{>1}\Big)^{T(x)} \Big(\underbrace{\frac{1-\theta_1}{1-\theta_0}}_{<1}\Big)^{n-T(x)}$.

Dies ist eine streng monoton wachsende Funktion von $T(x)$ (wegen $\theta_0 < \theta_1$),

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

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

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

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 $n \in \mathbb{N}$: $P_N(Y \leq N \leq \lfloor Y\alpha^{-\frac{1}{n}} \rfloor) = 1 - P_N(N \leq \lfloor Y\alpha^{-\frac{1}{n}} \rfloor) \geq P_N(Y > N\alpha^{\frac{1}{n}})$

$= 1 - P_N(X_i \leq N\alpha^{\frac{1}{n}} \text{ für alle } i = 1,\ldots,n) = 1 - \big(\frac{N\alpha^{\frac{1}{n}}}{N}\big)^n = 1 - \alpha.$

Also: $\{Y, Y+1, \ldots, \lfloor Y\alpha^{-\frac{1}{n}} \rfloor\}$ 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 $\{22, 23, \ldots, 40\}$ 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 $A(\theta_0) \subseteq \mathcal{X}$ 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: $P_\theta(\{x \in \mathcal{X} : C(x) \ni \theta\}) = P_\theta(\{x \in \mathcal{X} : A(\theta) \not\ni x\}) = 1 - P_\theta(A(\theta)) \geq 1-\alpha$. □

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 $b := u_{1-\frac{\alpha}{2}}$, $a := -b$:

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

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 ±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 ±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:

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

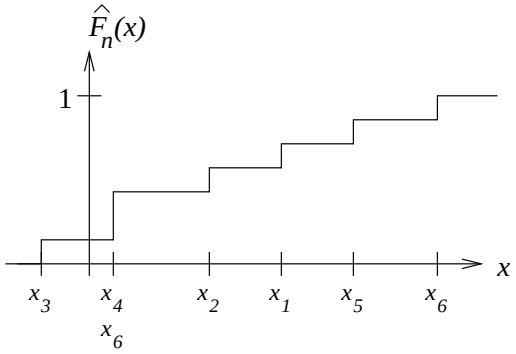
Was läßt sich über den Fehler 2. Art aussagen?
Sei $\theta \neq \theta_0$; $\epsilon := \frac{1}{2}|\theta - \theta_0|$. Aus $|\overline{X_n} - \theta| \leq \epsilon$ folgt dann $|\overline{X_n} - \theta_0| \geq |\theta - \theta_0| - |\overline{X_n} - \theta| \geq \epsilon \geq \frac{1}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\sqrt{\theta_0(1-\theta_0)}$ 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 $P_\theta(H \text{ wird verworfen}) = 1 \quad \forall \theta \neq \theta_0$.
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

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

Dies führt auf die empirische Verteilungsfunktion $\hat{F}_n : \mathbb{R} \to \mathbb{R};\ \hat{F}_n(x) := \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,x]}(X_i)$ 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 Satzes von Glivenko-Cantelli (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.
Vorüberlegung: Ist die Verteilungsfunktion R von $\hat{\theta} - \theta$ bekannt, also auch Quantile r_α, so gilt:
$P(r_{\frac{\alpha}{2}} \le \hat{\theta} - \theta \le r_{1-\frac{\alpha}{2}}) = P(\hat{\theta} - \theta \le r_{1-\frac{\alpha}{2}}) - P(\hat{\theta} - \theta \le r_{\frac{\alpha}{2}}) = R(r_{1-\frac{\alpha}{2}}) - R(r_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = \alpha$.
Wegen $\hat{\theta} - \theta \le r_{1-\frac{\alpha}{2}} \iff \theta \ge \hat{\theta} - r_{1-\frac{\alpha}{2}}$ 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. Idee: Schätze R.
Hierzu werden künstliche Stichproben (resample) aus den Originaldaten $x_1, \dots, x_n$ konstruiert.

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

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	$\frac{44.1}{6}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Dies geschieht auf einem Computer mit einem Pseudozufallszahlengenerator.
Bei jeder Stichprobe wird der Wert $\eta_k := \theta_k^* - \hat{\theta}$ berechnet, $1 \le k \le B$, wobei θ_k^* den Schätzwert bezeichnet, der zum Resample $x_{k1}^*, \dots, x_{kn}^*$ gehört. Die empirische Verteilungsfunktion $\hat{R}_B$ zu diesen Werten $\eta_1, \dots, \eta_B$ 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".

Kapitel 8

Integral und Erwartungswert

8.1 Einführung und Erinnerung

$(\Omega, \mathfrak{A}, P)$ WRaum, $X : \Omega \to \mathbb{R}$ eine Zufallsvariable (ZV) (d.h. X ist $(\mathfrak{A}, \mathfrak{B})$-meßbar).
EX war bisher nur für diskrete ZV definiert.
Allgemeine Situation wird mit dem Lebesgue-Integral behandelt.
Es sei $(\Omega, \mathfrak{A})$ ein meßbarer Raum und $\mu : \mathfrak{A} \to [0, \infty]$ ein Maß, d.h.

$\mu(\emptyset) = 0$

$\mu(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$, $A_1, A_2, \ldots \in \mathfrak{A}$ disjunkt (σ-Additivität).

Beachte: Wert ∞ ist zugelassen.
Rechenregeln: $a + \infty = \infty + a = \infty \qquad \forall a \in \mathbb{R} \cup \{\infty\}$
$a \cdot \infty = \infty \cdot a = \infty \qquad \forall a > 0$
$\infty - \infty$ nicht definiert.
Wichtiges Beispiel: Lebesgue-Maß l, festgelegt durch $l((a, b]) = b - a \quad \forall a, b \in \mathbb{R},\ a < b$.
WMaße sind Maße mit der Gesamtmasse 1 ($\mu(\Omega) = 1$)

Bemerkung 8.1

(i) "Allgemeine" Maße sind noch stets stetig von unten, aber i.A. nicht stetig von oben oder in $\emptyset$,
z.B. $l((-\infty, n]) = \infty \quad \forall n \in \mathbb{N}$, aber $l(\bigcap_{n=1}^{-\infty} (-\infty, n]) = l(\emptyset) = 0$

(ii) l und $2l$ stimmen auf dem $\cap$-stabilen Erzeuger $\{(-\infty, x] : x \in \mathbb{R}\}$ überein, sind aber nicht mehr gleich: Satz 5.9 (Eindeutigkeitssatz) muß modifiziert werden (Übungsaufgabe)

(iii) Beziehungen wie $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ sind bei allgemeinen Maßen i.A. falsch (sinnlos), da $\infty - \infty$ auftreten kann.

8.2 Konstruktion des Lebesgue-Integrals

$(\Omega, \mathfrak{A}, \mu)$ sei ein Maßraum. Für eine (möglichst große) Klasse von $(\mathfrak{A}, \mathfrak{B})$-meßbaren Funktionen $f : \Omega \to \mathbb{R}$ soll $\int f d\mu \quad (= \int f(\omega) \mu d(\omega) = \int f(\omega) d\mu(\omega))$ definiert werden. Wir beginnen mit der Forderung, daß für Indikatorfunktionen $f = 1_A \quad \int f d\mu = \mu(A)$ gelten soll.
Beachte: $f = 1_A$ meßbar $\iff A \in \mathfrak{A}$.

Soll $f \to \int f d\mu$ linear sein, so ist damit $\int f d\mu$ für endliche Linearkombinatinen von meßbaren Indikatorfunktionen mit nicht-negativen Koeffizienten festgelegt.
Wir schreiben $Prim_+(\Omega, \mathfrak{A})$ für diese Klasse ("primitive Funktionen").
Offensichtlich gilt: $Prim_+(\Omega, \mathfrak{A}) = \{f : \Omega \to \mathbb{R} : f\ (\mathfrak{A}, \mathfrak{B})\text{-meßbar}, \#f(\Omega) < \infty^1\}$.

Definition 8.2 ("Stufe 1") Für $f \in Prim_+(\Omega, \mathfrak{A})$ sei $\int f d\mu := \sum_{x \in f(\Omega)} x\mu \underbrace{(f^{-1}(\{x\}))}_{\{\omega \in \Omega : f(\omega) = x\}}$.

Hierbei vereinbaren wir $0 \cdot \infty = \infty \cdot 0 = 0$; insbesondere gilt auch für endliche Maße $\mu \int c d\mu = c\mu(\Omega),\ c \geq 0$ (Integrale konstanter Funktionen, $\int 0 d\mu = 0$)

Lemma 8.3 Es seien $f, g \in Prim_+(\Omega, \mathfrak{A})$ und $c \geq 0$. Dann gilt:
$cf \in Prim_+(\Omega, \mathfrak{A}),\ f + g \in Prim_+(\Omega, \mathfrak{A})$, sowie
$\int cf d\mu = c \int f d\mu,\ \int (f + g) d\mu = \int f d\mu + \int g d\mu$ und $f \leq g \Longrightarrow \int f d\mu \leq \int g d\mu$.
Beweis der Monotonie (Linearität in den Übungen): Sei $f(\Omega) = \{a_1, \ldots, a_n\}$ und $g(\Omega) = \{b_1, \ldots, b_m\}$. Man setze

[1] Anzahl der Elemente des Wertebereiches von f ist kleiner ∞

$A_i := f^{-1}(\{a_i\})$ und $B_j := g^{-1}(\{b_j\})$, so gilt: $f = \sum_{i=1}^{n} a_i \cdot 1_{A_i}$, $g = \sum_{j=1}^{m} b_j \cdot 1_{B_j}$.

Aus $f \le g$ folgt $A_i \cap B_j \neq \emptyset \implies a_i \le b_j$, also:

$\int f d\mu = \sum_{i=1}^{n} a_i \mu(A_i) =^2 \sum_{i=1}^{n} \sum_{j=1}^{m} a_i \mu(A_i \cap B_j) \le \sum_{i=1}^{n} \sum_{j=1}^{m} b_j \mu(A_i \cap B_j) = \sum_{j=1}^{m} b_j \mu(B_j) = \int g d\mu.$ □

Dieses Lemma zeigt, daß das bisher definierte Integral linear und isoton ist (auf $Prim_+$), insbesondere gilt für alle Funktionen $f = \sum_{i=1}^{n} a_i \cdot 1_{A_i}$, $a_i \ge 0$, $A_i \in \mathfrak{A}$ die Formel $\int f d\mu = \sum_{i=1}^{n} a_i \mu(A_i)$, auch wenn die A_i's nicht disjunkt oder die a_i's nicht verschieden sind (alle "Darstellungen" liefern dasselbe Integral).

Lemma 8.4 Zu jeder $(\mathfrak{A}, \mathfrak{B})$-meßbaren, nicht-negativen Funktion $f : \Omega \to \mathbb{R}$ existiert eine Folge $(f_n)_{n\in\mathbb{N}} \subset Prim_+(\Omega, \mathfrak{A})$ mit $f_n \uparrow f$.

Beweis: $f_n := \sum_{i=0}^{2^n-1} \frac{i}{2^n} 1_{A_{ni}}$; $A_{ni} := \{\omega : \frac{i}{2^n} \le f(\omega) < \frac{i+1}{2^n}\}$. □

Definition 8.5 ("Zweite Stufe") Das μ-Integral einer nicht-negativen, $(\mathfrak{A}, \mathfrak{B})$-meßbaren Funktion $f : \Omega \to \mathbb{R}^+$ wird definiert durch $\int f d\mu := \sup\{\int g d\mu : g \in Prim_+(\Omega, \mathfrak{A}), g \le f\}$.
Wegen $g \equiv 0 \in Prim_+$, $g \le f$ wird das Supremum über eine nicht-leere Menge gebildet.

Im Falle $f \in Prim_+$ taucht f selbst in dieser Menge auf; wegen der Monotonie auf $Prim_+$ (Lemma 8.4) impliziert dies, daß Def. 8.5 mit Def. 8.2 verträglich ist.

Lemma 8.6 Ist $(f_n)_{n\in\mathbb{N}} \subset Prim_+$ mit $f_n \uparrow f$, so gilt $\int f d\mu = \lim_{n\to\infty} \int f_n d\mu$.

Beweis: "$\ge$": folgt unmittelbar aus Def. 8.5.

"$\le$": Sei nun $g \in Prim_+(\Omega, \mathfrak{A})$; $g = \sum_{i=1}^{k} a_i \cdot 1_{A_i}$ mit $g(\Omega) = \{a_1, \ldots, a_k\}$, $g \le f$.

Sei $0 < \eta < 1$ und $B_n := \{\omega \in \Omega : \eta g(\omega) \le f_n(\omega)\}$.

Aus $\eta g < f$ auf $\{f > 0\}$ und $f_n = f$ auf $\{f = 0\}$, sowie $f_n \uparrow f$ folgt $B_n \uparrow \Omega$, also mit der Stetigkeit von unten von μ

$\eta \int g d\mu = \eta \lim_{n\to\infty} \sum_{i=1}^{k} a_i \mu(A_i \cap B_n) = \lim_{n\to\infty} \int \eta g 1_{B_n} d\mu \le \lim_{n\to\infty} \int f_n d\mu.$

Da die rechte Seite nicht von η abhängt, folgt hieraus mit $\eta \uparrow 1$ $\int g d\mu \le \lim_{n\to\infty} \int f_n d\mu$.

Bildet man das Supremum über die zugelassenen g, so folgt die andere Ungleichung. □

Lemma 8.7 Es seien $f, g : \Omega \to \mathbb{R}^+$ $(\mathfrak{A}, \mathfrak{B})$-meßbar und $c \ge 0$. Dann gilt:
$\int (cf) d\mu = c \int f d\mu$, $\int (f + g) d\mu = \int f d\mu + \int g d\mu$
$f \le g \implies \int f d\mu \le \int g d\mu$

Beweis: Lemma 8.4 liefert die Existenz von Folgen $(f_n)_{n\in\mathbb{N}}$, $(g_n)_{n\in\mathbb{N}} \subset Prim_+$ mit $f_n \uparrow f$, $g_n \uparrow g$.

Klar: $(f_n + g_n)_{n\in\mathbb{N}} \subset Prim_+$, $f_n + g_n \uparrow f + g$.

$\int (f + g) d\mu \overset{\text{Lemma 8.6}}{=} \lim_{n\to\infty} \int (f_n + g_n) d\mu = \lim_{n\to\infty} (\int f_n d\mu + \int g_n d\mu) = \lim_{n\to\infty} \int f_n d\mu + \lim_{n\to\infty} \int g_n d\mu$

$\overset{\text{2xLemma 8.6}}{=} \int f d\mu + \int g d\mu$

Analog: $\int (cf) d\mu = \lim_{n\to\infty} \int (cf_n) d\mu = c \lim_{n\to\infty} \int f_n d\mu = c \int f d\mu$

Isotonie folgt unmittelbar aus der Definition (es wird das Supremum über eine größere Menge gebildet.) □

Wir definieren Positiv- und Negativteil einer Funktion f durch $f^+ := \max\{f, 0\}$, $f^- := -\min\{f, 0\}$

Klar: $f = f^+ - f^-$, $|f| = f^+ + f^-$

Mit f sind auch f^+ und f^- $(\mathfrak{A}, \mathfrak{B})$-meßbar.

Definition 8.8 ("Stufe 3") Eine $(\mathfrak{A}, \mathfrak{B})$-meßbare Funktion $f : \Omega \to \mathbb{R}$ heißt integrierbar bezüglich μ, wenn $\int f^+ d\mu < \infty$ und $\int f^- d\mu < \infty$.
Das μ-Integral einer solchen Funktion wird definiert durch $\int f d\mu = \int f^+ d\mu - \int f^- d\mu$ (andere Schreibweisen: $\int f(\omega)\mu(d\omega)$, $\int f(\omega) d\mu(\omega)$, $\mu(f)$).
Die Menge aller μ-integrierbaren Funktionen auf $(\Omega, \mathfrak{A})$ wird mit $\mathfrak{L}^1(\Omega, \mathfrak{A}, \mu)$ bezeichnet.

Bemerkung 8.9 Gilt mindestens eine der Beziehung $\int f^+ d\mu < \infty$, $\int f^- d\mu < \infty$, so läßt sich $\int f d\mu$ noch sinnvoll als Differenz der beiden Werte definieren; solche Funktionen werden gelegentlich "quasi-integrierbar" oder "integrierbar im erweiterten Sinne" genannt.
Unmittelbar aus der Definition und $|f| = f^+ + f^-$ folgt:

$(*)$ f μ-integrierbar $\iff$ $|f|$ μ-integrierbar.

$(**)$ $|\int f d\mu| \le \int |f| d\mu$

Satz 8.10 Es seien $f, g : \Omega \to \mathbb{R}$ μ-integrierbar und $c \in \mathbb{R}$. Dann sind auch cf und $f + g$ integrierbar und es gilt
$\int (cf) d\mu = c \int f d\mu$, $\int (f + g) d\mu = \int f d\mu + \int g d\mu$
$f \le g \implies \int f d\mu \le \int g d\mu$.

[2] denn $B_1, \ldots, B_m$ bilden eine Partition von Ω

Beweis: Wegen $|f+g| \le |f|+|g|$ folgt die Integrierbarkeit von $f+g$ aus (*)
$(f+g)^+$ ist i.a. verschieden von $f^+ + g^+$.
Es gilt jedoch $(f+g)^+ - (f+g)^- = f+g = f^+ - f^- + g^+ - g^-$, also $(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+$.
Damit (nach Lemma 8.7) $\int (f+g)^+ d\mu + \int f^- d\mu + \int g^- d\mu = \int (f+g)^- d\mu + \int f^+ d\mu + \int g^+ d\mu$.
Also: $\int (f+g) d\mu = \int (f+g)^+ d\mu - \int (f+g)^- d\mu = \int f^+ d\mu + \int g^+ d\mu - \int f^- d\mu - \int g^- d\mu = \int f d\mu + \int g d\mu$.
Der Beweis $\int (cf) d\mu = c \int f d\mu$ ist einfach, wenn man $c \ge 0$ und $c < 0$ separat betrachtet.
Isotonie: Übungsaufgabe. □

Beispiel 8.11 (i) Bezeichnet δ_ω das Einpunktmaß in ω, also $\delta_\omega(A) = \begin{cases} 1 & , \quad \omega \in A \\ 0 & , \quad \text{sonst} \end{cases} = 1_A(\omega)$, so gilt $\int f d\delta_\omega = f(\omega)$ (Dirac-Maß)

(ii) Ein Maß μ von der Form $\mu = \sum_{i=1}^{\infty} \alpha_i \delta_{\omega_i}$ (mit $\alpha_i \ge 0 \ \ \forall i$) heißt diskret.

Eine Funktion $f : \Omega \to \mathbb{R}$ ist genau dann μ-integrierbar, wenn $\sum_{i=1}^{\infty} \alpha_i |f(\omega_i)| < \infty$, und man hat dann $\int f d\mu = \sum_{i=1}^{\infty} \alpha_i f(\omega_i)$ $\left(= \sum_{\omega \in \Omega} f(\omega) \mu(\{\omega\}) \right)$

Im Spezialfall $\Omega = \mathbb{N}$, $\mathfrak{A} = \mathfrak{P}(\mathbb{N})$, $\mu = \sum_{n=1}^{\infty} \delta_n$ (das abzählbare Maß; $\mu(A) = \#A$) erhält man Reihensummen als Integral.

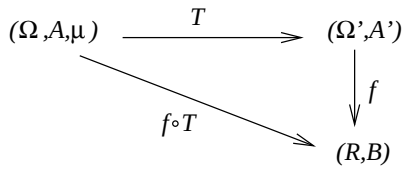
(iii) Es seien $\Omega = [a,b]$, $a < b$, $\mathfrak{A}$ die Spur von $\mathfrak{B}$ auf Ω und μ die Restriktion von l auf $\mathfrak{A}$. Ist $f : \Omega \to \mathbb{R}$ Borel-meßbar und Riemannintegrierbar (Konvergenz der Ober- und Untersummen, etc.), so ist f auch μ-integrierbar und es gilt $\int f d\mu = \int_a^b f(x) dx$.

Die Funktion $1_{\mathbb{Q} \cap [a,b]}$ ist ein Beispiel für eine μ-integrierbare Funktion, die nicht Riemannintegrierbar ist.
Grob: Das Integral bzgl. l (das Lebesgue-Integral) erweitert das (eigentliche) Riemann-Integral, so daß die bekannten Formeln für das Riemann-Integral weiterverwendet werden können (wir schreiben gelegentlich "dx" anstelle von "$l(dx)$".)

8.3 Transformationsformel, Nullmengen

$(\Omega, \mathfrak{A}, \mu)$ sei ein Maßraum, $(\Omega', \mathfrak{A}')$ ein meßbarer Raum, $T : \Omega \to \Omega'$ eine $(\mathfrak{A}, \mathfrak{A}')$-meßbare Funktion. Dann wird durch $\mu' : \mathfrak{A}' \to [0, \infty]$, $\mu'(A') := \mu(T^{-1}(A'))$ $\ \forall A' \in \mathfrak{A}'$ ein Maß auf $(\Omega', \mathfrak{A}')$ definiert, das Bildmaß von μ unter T (Schreibweise: μ^T.)
Dies verallgemeinert den aus Satz 5.4 bekannten Begriff der Verteilung einer Zufallsgröße.

Satz 8.12 (Integration bzgl. eines Bildmaßes, Transformationssatz)
Mit der oben gezeigten Beziehung gilt für alle $(\mathfrak{A}', \mathfrak{B})$-meßbaren $f : \Omega' \to \mathbb{R}$:
f ist genau dann μ^T-integrierbar, wenn $f \circ T$ μ-integrierbar ist und dann gilt:
$\int f d\mu^T = \int f \circ T d\mu$.



Beweis: (mit der "üblichen Maschinerie", "algebraische Induktion")
Im Falle $f = 1_A$ hat man $\int f d\mu^T = \mu(T^{-1}(A)) = \int 1_{T^{-1}(A)} d\mu = \int 1_A \circ T d\mu$.
Die Beh. ist also für Indikatorfunktionen richtig. Da das Integral linear ist, ist sie damit auch auf $Prim_+(\Omega', \mathfrak{A}')$ richtig.
Ist $f \ge 0$, so existiert nach Lemma 8.4 und 8.6 eine Folge $(f_n)_{n \in \mathbb{N}} \subset Prim_+(\Omega', \mathfrak{A}')$ mit $f_n \uparrow f$ und $\int f d\mu^T = \lim_{n \to \infty} \int f_n d\mu^T$.
Klar: $(f_n \circ T)_{n \in \mathbb{N}} \subset Prim_+(\Omega, \mathfrak{A})$, $f_n \circ T \uparrow f \circ T$.
Also folgt mit Lemma 8.6 $\int f d\mu^T = \lim_{n \to \infty} \int f_n d\mu^T = \lim_{n \to \infty} \int f_n \circ T d\mu = \int f \circ T d\mu$.
Ist schließlich $f : \Omega' \to \mathbb{R}$ eine beliebige $(\mathfrak{A}', \mathfrak{B})$-meßbare Abbildung, so erhält man mit dem bereits gezeigten:
$\int f^\pm d\mu^T < \infty \iff \int f^\pm \circ T d\mu < \infty$.
Da $(f \circ T)^\pm = f^\pm \circ T$, folgt die behauptete Integrabilitätsaussage; auch gilt dann
$\int f d\mu^T = \int f^+ d\mu^T - \int f^- d\mu^T = \int f^+ \circ T d\mu - \int f^- \circ T d\mu = \int (f \circ T)^+ d\mu - \int (f \circ T)^- d\mu = \int (f \circ T) d\mu$. □

Definition 8.13 $(\Omega, \mathfrak{A}, \mu)$ Maßraum. Für jedes $\omega \in \Omega$ sei $A(\omega)$ eine Aussage. Man sagt: "A gilt fast überall" (μ-f.ü.), wenn es eine μ-Nullmenge N (also ein $N \in \mathfrak{A}$ mit $\mu(N) = 0$) gibt mit: $A(\omega)$ ist wahr für alle $\omega \notin N$.

Satz 8.14 $(\Omega, \mathfrak{A}, \mu)$ Maßraum, $f, g : \Omega \to \mathbb{R}$ meßbar.
(i) Sei außerdem $f \ge 0$. Dann gilt: $\int f d\mu = 0 \iff f = 0$ μ-f.ü.
(ii) Ist f integrierbar und gilt $f = g$ μ-f.ü., so ist auch g integrierbar und es gilt $\int f d\mu = \int g d\mu$.
Beweis: (i) Setze $N = \{f \ne 0\}$.
- Angenommen es gilt $\int f d\mu = 0$. Setze $A_n = \{\omega \in \Omega : f(\omega) \ge \frac{1}{n}\}$. Dann $A_n \uparrow N$, also $\mu(N) = \lim_{n \to \infty} \mu(A_n)$.
  $0 = \int f d\mu \ge \int \frac{1}{n} 1_{A_n} d\mu = \frac{1}{n} \mu(A_n) \ge 0$, also $\mu(A_n) = 0$ $\forall n$, und damit $\mu(N) = 0$.
- Ist umgekehrt N eine μ-Nullmenge mit $f(\omega) = 0$ $\forall \omega \in N^c$, und $g = \sum a_i 1_{A_i} \in Prim_+$ mit $g \le f$, $\{a_1, \dots, a_n\} = g(\Omega)$, so folgt $A_i \subseteq N$ $\forall i$ mit $a_i > 0$ und damit $\int g d\mu = \sum_{a_i > 0} a_i \mu(A_i) = 0$.

Hieraus folgt $\int f d\mu = 0$ mit Def. 8.5 (Supremumsbildung)

(ii) Es seien zunächst $f, g \geq 0$, $N := \{f \neq g\}$. Dann gilt
$\int f d\mu = \int f 1_N d\mu + \int f 1_{N^c} d\mu = 0 + \int g 1_{N^c} d\mu = \int g 1_N d\mu + \int g 1_{N^c} d\mu = \int g d\mu$.
Insbesondere $\int f d\mu = \infty \iff \int g d\mu = \infty$, hieraus folgt (bei beliebigen f, g) die Integrierbarkeitsaussage wie im Beweis des Transformationssatzes.
Es seien nun $f, g : \Omega \to \mathbb{R}$ meßbar mit $f = g$ μ-f.ü. Wegen $\{f^+ = g^+\} \supseteq \{f = g\} \subseteq \{f^- = g^-\}$ gilt dann auch $f^+ = g^+$ μ-f.ü., also $\int f d\mu = \int f^+ d\mu - \int f^- d\mu = \int g^+ d\mu - \int g^- d\mu = \int g d\mu$. □

8.4 Konvergenzsätze

Das Maß-Integral hat bei der Behandlung von Funktionenfolgen gegenüber dem Riemann-Integral einige Vorzüge.
$(\Omega, \mathfrak{A}, \mu)$ Maßraum, $f, f_1, f_2, \ldots : \Omega \to \mathbb{R}$ seien meßbar.

Satz 8.15 (Satz von Beppo Levi, Satz von der monotonen Konvergenz)
Sind $f, f_1, f_2, \ldots$ nicht-negativ mit $f_n \uparrow f$, so gilt $\int f d\mu = \lim\limits_{n\to\infty} \int f_n d\mu$.

Beweis: Zu jedem f_n existiert eine Folge $(g_{nm})_{m\in\mathbb{N}} \subset Prim_+$ mit $g_{nm} \uparrow f_n$ für $m \to \infty$

Diagramm:

$$\begin{array}{ccccccccccc}
g_{11} & \leq & g_{12} & \leq & g_{13} & \leq & g_{14} & \leq & \ldots & \to f_1 \\
g_{21} & \leq & g_{22} & \leq & g_{23} & \leq & g_{24} & \leq & \ldots & \to f_2 \\
g_{31} & \leq & g_{32} & \leq & g_{33} & \leq & g_{34} & \leq & \ldots & \to f_3 \\
\vdots & & \vdots & & \vdots & & \vdots & & & \downarrow \\
 & & & & & & & & & f
\end{array}$$

Setze $h_m := \max\{g_{1m}, \ldots, g_{mm}\}$.
Klar: $h_m \in Prim_+ \ \forall m$, $h_m \uparrow$.
Für $n \leq m$ gilt $g_{nm} \leq h_m$, also folgt $f_n = \sup\limits_m g_{nm} = \sup\limits_{m\geq n} g_{nm} \leq \sup\limits_m h_m$.
Außerdem $h_m \leq f_m \leq f$, also insgesamt $h_m \uparrow f$.
Nun folgt mit Lemma 8.6 $\int f d\mu = \lim\limits_{m\to\infty} \int h_m d\mu \leq \lim\limits_{m\to\infty} \int f_m d\mu \leq \int f d\mu$. □

Satz 8.16 (Lemma von Fatou)
Gilt $f_n \geq 0 \ \forall n$, so folgt $\int \liminf\limits_{n\to\infty} f_n d\mu \leq \liminf\limits_{n\to\infty} \int f_n d\mu$.

Beweis: Setze $g_n := \inf\limits_{m\geq n} f_m$, dann gilt $g_n \uparrow f := \liminf\limits_{n\to\infty} f_n$.
$g_n \leq f_n$. Mit Satz 8.15 folgt $\int \liminf\limits_{n\to\infty} f_n d\mu = \lim\limits_{n\to\infty} \int g_n d\mu = \liminf\limits_{n\to\infty} \int g_n d\mu \leq \liminf\limits_{n\to\infty} \int f_n d\mu$. □

Satz 8.17 (Satz von der majorisierten (dominierten) Konvergenz, Satz von Lebesgue)
Es gelte $\lim\limits_{n\to\infty} f_n(\omega) = f(\omega)$ für alle $\omega \in \Omega$. Existiert dann eine μ-integrierbare Funktion $g : \Omega \to \mathbb{R}$ mit $|f_n(\omega)| \leq g(\omega)$ für alle $\omega \in \Omega$, $n \in \mathbb{N}$, so folgt $\int f d\mu = \lim\limits_{n\to\infty} \int f_n d\mu$. ($g$ heißt "integrierbare Majorante")

Beweis: Setze $g_n := |f_n - f|$, $h := |f| + g$.
Wegen $|h| \leq 2g$ ist h integrierbar. Außerdem $h - g_n = |f| + g - |f_n - f| \geq |f| + g - |f_n| - |f| = g - |f_n| \geq 0$.
Wegen $g_n \to 0$ gilt $h - g_n \to h$, also folgt mit Fatou $\int h d\mu \leq \liminf\limits_{n\to\infty} \int (h - g_n) d\mu = \int h d\mu - \limsup\limits_{n\to\infty} \int g_n d\mu$.
(beachte: hier geht Integrierbarkeit von g ein, sonst könnt hier u.U. $\infty - \infty$ stehen.)
also: $\limsup \int g_n d\mu = 0$.
Damit $\lim\limits_{n\to\infty} |f_n - f| d\mu = \lim\limits_{n\to\infty} \int g_n d\mu = 0$ und mit $|\int f_n d\mu - \int f d\mu| = |\int (f_n - f) d\mu| \leq \int |f_n - f| d\mu$ folgt hieraus $\int f_n d\mu \to \int f d\mu$. □

Bemerkung 8.18

(i) Man kann die Konstruktion des μ-Integrals auf $\overline{\mathbb{R}}$ ($:= [-\infty, \infty]$)-wertige Funktionen ausdehnen, muß dann aber bei den Linearitätsaussagen Einschränkungen machen ($f + g$ u.U. nicht überall definiert). Ist ein $\overline{\mathbb{R}}$-wertiges f in diesem Sinne integrierbar, so ist $N := \{\omega \in \Omega : |f(\omega)| = \infty\}$ eine μ-Nullmenge und es gilt $\int f d\mu = \int \underbrace{f 1_{N^c}}_{\mathbb{R}-\text{wertig}} d\mu$.

(ii) Die Erweiterung aus (i) wurde in den obigen Sätzen bereits implizit verwendet ($\liminf f_n$ beispielsweise kann durchaus den Wert ∞ annehmen.)

(iii) Zu den Sätzen 8.15-8.17 existieren auch "fast-überall"-Versionen. Hat man beispielsweise in Satz 8.17 nur $f_n \to f$ μ-f.ü., so gilt noch stets $\int f d\mu = \lim\limits_{n\to\infty} \int f_n d\mu$:

Ist N eine μ-Nullmenge, außerhalb dieser Konvergenz gilt, so hat man $f_n 1_{N^c}(\omega) \to f 1_{N^c}(\omega)$ für alle $\omega \in \Omega$.
Auf $f_n 1_{N^c}$ läßt sich Satz 8.17 anwenden:

$$\int f d\mu \overset{8.14\,(ii)}{=} \int f 1_{N^c} d\mu \overset{8.17}{=} \lim_{n\to\infty} \int f_n 1_{N^c} d\mu \overset{8.14\,(ii)}{=} \lim_{n\to\infty} \int f_n d\mu.$$

8.5 Maße und Dichten

Abkürzende Schreibweise: $\int\limits_A f d\mu = \int f 1_A d\mu$.

Lemma 8.19 $(\Omega, \mathfrak{A}, \mu)$ Maßraum, $f : \Omega \to \mathbb{R}_+$ meßbar. Dann wird durch $\nu : \mathfrak{A} \to [0, \infty]$, $\nu(A) := \int\limits_A f d\mu$ ein Maß auf $(\Omega, \mathfrak{A})$ definiert.
Beweis: $\nu(\emptyset) = \int\limits_\emptyset f d\mu = \int f 1_\emptyset d\mu = \int 0 d\mu = 0$.

Seien $(A_n)_{n\in\mathbb{N}}$ paarweise disjunkt.

$$\nu(\sum_{n=1}^{\infty} A_n) = \int f 1_A d\mu \overset{3}{=} \lim_{n\to\infty} \int f 1_{\bigcup_{i=1}^{n} A_i} d\mu \overset{4}{=} \lim_{n\to\infty} \sum_{i=1}^{n} \int f 1_{A_i} d\mu \overset{5}{=} \lim_{n\to\infty} \sum_{i=1}^{n} \nu(A_i) = \sum_{i=1}^{\infty} \nu(A_i). \qquad \square$$

Definition 8.20 μ, ν seien Maße auf $(\Omega, \mathfrak{A})$. Existiert eine nicht-negative, meßbare Funktion $f : \Omega \to \mathbb{R}$ mit $\nu(A) = \int\limits_A f d\mu$ für alle $A \in \mathfrak{A}$, so heißt f eine μ-Dichte von ν. Schreibweise: $f = \frac{d\nu}{d\mu}$.

Aus Satz 8.14 (ii) folgt, daß mit f auch jedes meßbare g mit $f = g$ μ-f.ü. eine μ-Dichte von ν ist. Dichten sind i.A. nicht eindeutig.
Hat ν eine Dichte f bezüglich μ, so folgt mit Satz 8.14 (i) $\mu(N) = 0 \implies f \cdot 1_N = 0$ μ-f.ü. $\implies \nu(N) = 0$.
Gilt dies, so sagt man, daß μ ν dominiert und schreibt $\mu >> \nu$. Der Satz von Radon-Nikodym besagt, daß die Eigenschaft bei σ-endlichen Maßen bereits die Existenz einer Dichte impliziert.

Beispiel 8.21 Der Fall $(\Omega, \mathfrak{A}, \mu) = (\mathbb{R}, \mathfrak{B}, l)$ ist von besonderem Interesse. Ist $f : \mathbb{R} \to \mathbb{R}$ eine Borel-meßbare, nicht-negative Funktion mit l-Integral 1, so wird durch $P : \mathfrak{B} \to [0, 1]$, $P(A) := \int\limits_A f(x)dx$ ein WMaß definiert, das WMaß mit (Lebesgue-) Dichte f.
Ist X eine Zufallsvariable (meßbare Abbildung) mit Verteilung (Bildmaß) P, so heißt f eine WDichte von X.

Satz 8.22 μ, ν Maße auf $(\Omega, \mathfrak{A})$, ν habe μ-Dichte f. Dann gilt für alle $(\mathfrak{A}, \mathfrak{B})$-meßbaren $g : \Omega \to \mathbb{R}$:
g ist genau dann ν-integrierbar, wenn $g \circ f$ μ-integrierbar ist, und dann: $\int g d\nu = \int g \circ f d\mu$.
Beweis: mit der "üblichen Maschinerie", Übungsaufgabe. $\square$

8.6 Erwartungswerte

$(\Omega, \mathfrak{A}, P)$ WRaum, $X : \Omega \to \mathbb{R}$ ZV.

Definition 8.23

(i) Ist X P-integrierbar, dann heißt $EX := \int X dP$ der Erwartungswert von X. Gilt $E|X|^k \leq \infty$, so heißt EX^k das k-te Moment zu X (im Falle $E|X|^k = \infty$ sagen wir, daß das k-te Moment nicht existiert.)
(ii) Existiert zu X das zweite Moment, so heißt $var(X) := E(X - EX)^2$ die Varianz von X und $\sigma(X) := \sqrt{var(X)}$ die Standardabweichung.

Satz 8.24 (Wir setzen voraus, daß die beteiligten Erwartungswerte, etc. existieren.)
(i) $E(X + Y) = EX + EY$, $E(c \cdot X) = c \cdot EX$
(ii) $X \leq Y \implies EX \leq EY$
(iii) Ist X diskret verteilt, d.h. gibt es eine abzählbare Menge $A \subset \mathbb{R}$ mit $P(X \in A) = 1$, so gilt für alle Borel-meßbaren $g : \mathbb{R} \to \mathbb{R}$ $Eg(X) = \sum\limits_{x\in A} g(x) \cdot P(X = x)$.
(iv) Ist X absolut stetig verteilt, d.h. X hat eine Dichte f, so gilt für alle Borel-meßbaren $g : \mathbb{R} \to \mathbb{R}$ $Eg(X) = \int g(x)f(x)dx$.
Beweis: (i) und (ii): Satz 8.10
(iii) $P^X = \sum\limits_{x\in A} P(X = x)\delta_x$, Satz 8.12
(iv) Aus Satz 8.12, Satz 8.22 folgt $Eg(X) = \int\limits_\Omega g(X(\omega))P(d\omega) = \int\limits_{\mathbb{R}} g(x)P^X(dx) = \int\limits_{\mathbb{R}} g(x)f(x)dx$. $\square$

Offensichtlich gilt weiterhin: $var(a + X) = var(X) \quad \forall a \in \mathbb{R}$
$var(aX) = a^2 var(X) \quad \forall a \in \mathbb{R}$
$var(X) = EX^2 - (EX)^2$

Beispiel 8.25 (siehe auch Beispiel 5.25)
$X \sim N(0, 1)$, d.h. X hat Dichte $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$.
Mit Satz 8.24(iv) folgt $EX = \int x \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})dx = 0$, $EX^2 = \int x^2 \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})dx = 1$, also $var(X) = 1$.

---

[3] mit $A := \bigcup_{i=1}^{\infty} A_i$
[4] Mit dem Satz von der monotonen Konvergenz für $(0 \leq) f 1_{\bigcup_{i=1}^{n} A_i} \uparrow f 1_A$
[5] da A_i's disjunkt

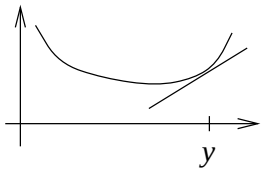
Bekanntlich: $Y := \mu + \sigma X \sim N(\mu, \sigma^2)$, also $EY = E(\mu + \sigma X) = \mu + \sigma EX = \mu$,
$var(Y) = var(\mu + \sigma X) = var(\sigma X) = \sigma^2 var(X) = \sigma^2$.
Interpretation der Parameter einer Normalverteilung: $\mu \rightsquigarrow$ Erwartungswert, $\sigma^2 \rightsquigarrow$ Varianz.

8.7 Ungleichungen

Satz 8.26 $(\Omega, \mathfrak{A}, P)$ WRaum, $X : \Omega \to \mathbb{R}$ ZV.
(i) Markovsche Ungleichung
$P(|X| \geq a) \leq \frac{1}{a^k} E|X|^k$ für alle $a > 0$.
(ii) Ungleichung von Chebychev
Existiert zu X die Varianz, so gilt $P(|X - EX| \geq a) \leq \frac{1}{a^2} var(X)$.

Beweis:
(i) Definiere $Y : \Omega \to \mathbb{R}$ durch $Y(\omega) = \begin{cases} 0 & , \quad |X(\omega)| < a \\ a & , \quad |X(\omega)| \geq a \end{cases}$
Offensichtlich gilt dann $|Y| \leq |X|$, also $|Y|^k \leq |X|^k$ und es folgt $a^k P(|X| \geq a) = a^k P(|Y| = a) = E|Y|^k \leq E|X|^k$.
(ii) Wende Teil (i) auf $X - EX$, $k = 2$ an. □

Es sei $I \subset \mathbb{R}$ ein offenes Intervall und $\varphi : I \to \mathbb{R}$ eine konvexe Funktion,
d.h. $\varphi(\alpha x + (1-\alpha)y) \leq \alpha\varphi(x) + (1-\alpha)\varphi(y)$.



Lemma 8.27 φ, I wie oben. Dann existiert zu jedem $y \in I$ ein $m \in \mathbb{R}$ mit $\varphi(x) \geq \varphi(y) + m(x - y)$ für alle $x \in I$
(anschaulich: durch jeden Punkt des Graphen geht eine Gerade (nicht notwendig eine Tangente, siehe $|\ldots|$), die unterhalb des Graphen liegt.

Satz 8.28 (Jensensche Ungleichung)
$I \subset \mathbb{R}$ sei ein offenes Intervall, $\varphi : I \to \mathbb{R}$ konvex, X ZV mit $E(X) < \infty$, $E|\varphi(X)| < \infty$, $P(X \in I) = 1$.
Dann gilt: $EX \in I$ und $\varphi(EX) \leq E\varphi(X)$. (Wichtige Spezialfälle: $|EX| \leq E|X|$, $(EX)^2 \leq EX^2$.)
Beweis: Ist $I = (-\infty, \infty)$, so gilt automatisch $EX \in I$.
Ist $X < a$ P-f.s., so gilt $EX \leq Ea = a$.
Wegen $a - X \geq 0$ würde $E(a - X) = 0$ auf $P(X = a) = 1$ führen, also $EX < a$ (analog: $X > a$ P-f.s. $\implies EX > a$.)
Also: $P(X \in I) = 1 \implies EX \in I$.
Nach Lemma 8.27 existiert ein $m \in \mathbb{R}$ mit $\varphi(X) \geq \varphi(EX) + m(X - EX)$. Nehme nun den Erwartungswert davon. □

$(\Omega, \mathfrak{A}, \mu)$ Maßraum, $\mathfrak{L}^p(\Omega, \mathfrak{A}, \mu) := \{f : \Omega \to \mathbb{R} : \int |f|^p d\mu < \infty\}$, $||f||_p := (\int |f|^p d\mu)^{\frac{1}{p}}$.
Klar: $||f||_p \geq 0$, $||af||_p = |a| \cdot ||f||_p$.

Satz 8.29
(i) (Höldersche Ungleichung)
Hat man $f \in \mathfrak{L}^p$, $g \in \mathfrak{L}^q$, wobei $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, so folgt $f \cdot g \in \mathfrak{L}^1$ und $||f \cdot g||_1 \leq ||f||_p ||g||_q$.
(ii) (Minkowskische Ungleichung)
$f, g \in \mathfrak{L}^p$, $p \geq 1 \implies f + g \in \mathfrak{L}^p$, $||f + g||_p \leq ||f||_p + ||g||_p$.

Beweis: (i) Man kann f durch $|f|$ und g durch $|g|$ ersetzen, also $f, g \geq 0$ annehmen. Sei $c := \int f^p d\mu$.
Bei $c = 0$ hat man $f = 0$ μ-f.ü. und die Ungleichung ist trivial.
Sei also $c > 0$. Sei P das Maß mit μ-Dichte $\frac{1}{c} f^p$.
Sei $h(\omega) := \begin{cases} \frac{g(\omega)}{f(\omega)^{p-1}} & , \quad f(\omega) > 0 \\ 0 & , \quad \text{sonst.} \end{cases}$.
Dann $||f||_p ||g||_q \geq ||f||_p ||g 1_{\{f>0\}}||_q = c^{\frac{1}{p}} (\int h^q g^{q(p-1)} d\mu)^{\frac{1}{q}} = c^{\frac{1}{p}} (c \int h^q dP)^{\frac{1}{q}} \overset{\text{Jensen}}{\geq} c^{\frac{1}{p}+\frac{1}{q}} \int h dP = \int g f d\mu = ||fg||_1$.
(ii) Wegen $|f + g| \leq |f| + |g|$ gilt $||f + g||_p \leq || \, |f| + |g| \, ||_p$, es reicht also wieder, die Ungleichung für $f, g \geq 0$ zu beweisen. Im Falle $p = 1$ hat man $||f + g||_1 = \int (f + g) d\mu = \int f d\mu + \int g d\mu = ||f||_1 + ||g||_1$.
Sei $p > 1$: $(f + g)^p \leq (2 \cdot \max\{f, g\})^p = 2^p \max\{f^p, g^p\} \leq 2^p (f^p + g^p)$. Dies zeigt, daß mit f und g auch $f + g$ in $\mathfrak{L}^p$ liegt. Mit $q := (1 - \frac{1}{p})^{-1}$ liefert Hölder: $||f + g||_p^p = \int (f + g)^p d\mu = \int f (f + g)^{p-1} d\mu + \int g (f + g)^{p-1} d\mu \leq (||f||_p + ||g||_p) ||(f + g)^{p-1}||_q$.
$||(f + g)^{p-1}||_q = (\int (f + g)^{(p-1)q} d\mu)^{\frac{1}{q}} = ||f + g||_p^{(p-1)}$.
Im Falle $||f + g||_p = 0$ ist nichts zu zeigen.
Im Falle $||f + g||_p > 0$ folgt die behauptete Ungleichung nach Division durch $||f + g||_p^{p-1}$ $(< \infty)$. □

Spezialfall von Hölder für $p = q = 2$. $|\int f g d\mu| \leq (\int f^2 d\mu)^{\frac{1}{2}} (\int g^2 d\mu)^{\frac{1}{2}}$. Dies ist die Cauchy-Schwarz-Ungleichung.

Spezialfall: $\Omega = \{1, \ldots, n\}$, $\mathfrak{A} = \mathbb{P}(\Omega)$, $\mu = \sum_{i=1}^{n} \delta_i$ $(\mu(A) = \#A$, "abzählendes Maß", "Zählmaß").

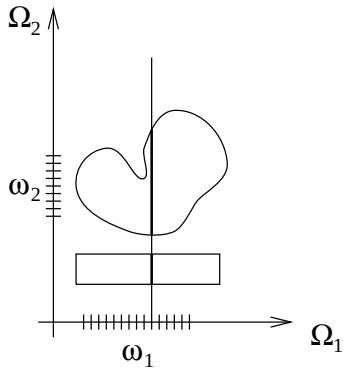
Dann: $\int a d\mu = \sum_{i=1}^{n} a_i$, $a = (a_1, \ldots a_n)$, also wird Cauchy-Schwarz zu $|\sum_{i=1}^{n} a_i b_i| \leq (\sum_{i=1}^{n} a_i^2)^{\frac{1}{2}} (\sum_{i=1}^{n} b_i^2)^{\frac{1}{2}}$.

Kapitel 9

Produktmaß und Unabhängigkeit

9.1 Das Produkt von zwei Maßräumen

Für $i = 1, 2$ sei $(\Omega_i, \mathfrak{A}_i, \mu_i)$ ein Maßraum. Sei $\Omega := \Omega_1 \times \Omega_2$.



Definition 9.1 Das Produkt $\mathfrak{A} = \mathfrak{A}_1 \otimes \mathfrak{A}_2$ der σ-Algebren $\mathfrak{A}_1$ und $\mathfrak{A}_2$ über Ω ist die von der Rechteckmengen $A_1 \times A_2,\ A_i \in \mathfrak{A}_i$ erzeugte σ-Algebra.

Lemma 9.2 Für alle $A \in \mathfrak{A},\ \omega_1 \in \Omega_1,\ \omega_2 \in \Omega_2$ gilt: $A_{\omega_1} := \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\} \in \mathfrak{A}_2,\ A_{\omega_2} := \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A\} \in \mathfrak{A}_1$. Man nennt A_{ω_1} den ω_1-Schnitt aus A.

Beweis: Sei $\omega_1 \in \Omega_1$. Wegen $\Omega_{\omega_1} = \Omega_2$ gilt $\Omega \in \mathfrak{A}' := \{A \in \mathfrak{A} : A_{\omega_1} \in \mathfrak{A}_2\}$.

$(\Omega \setminus A)_{\omega_1} = \{\omega_2 : (\omega_1, \omega_2) \notin A\} = (\{\omega_2 : (\omega_1, \omega_2) \in A\})^c = (A_{\omega_1})^c$ liefert "Komplementstabilität" von $\mathfrak{A}'$.

$(\bigcup_{n=1}^{\infty} A_n)_{\omega_1} = \bigcup_{n=1}^{\infty} (A_n)_{\omega_1}$. Damit: abzählbare Vereinigungen führen nicht aus $\mathfrak{A}$' heraus.

Insgesamt: $\mathfrak{A}$' ist eine σ-Algebra.

$$(A_1 \times A_2)_{\omega_1} = \begin{cases} A_2 & , \quad \omega_1 \in A_1 \\ \emptyset & , \quad \text{sonst} \end{cases}.$$

Also enthält $\mathfrak{A}'$ ein Erzeugendensystem von $\mathfrak{A}$, und damit insgesamt: $\mathfrak{A}' \supset \mathfrak{A}$. □

Definition 9.3 Ein Maß μ auf dem meßbaren Raum $(\Omega, \mathfrak{A})$ heißt σ-endlich, wenn es eine Folge $(A_n)_{n \in \mathbb{N}} \subseteq \mathfrak{A}$ gibt mit $\mu(A_n) < \infty\ \forall n \in \mathbb{N},\ \bigcup_{n=1}^{\infty} A_n = \Omega$.

Beispielsweise ist das Lebesgue-Maß l auf $(\mathbb{R}, \mathfrak{B})$ σ-endlich: $l([-n, n]) = 2n,\ \bigcup_{n=1}^{\infty} [-n, n] = \mathbb{R}$.

Lemma 9.4 Die Maße μ_1 und μ_2 seien σ-endlich.
Dann gilt für alle $A \in \mathfrak{A}$: $\omega_1 \to \mu_2(A_{\omega_1})$ ist $(\mathfrak{A}_1, \overline{\mathfrak{B}})$-meßbar, $\omega_2 \to \mu_1(A_{\omega_2})$ ist $(\mathfrak{A}_2, \overline{\mathfrak{B}})$-meßbar.

Beweis: Da μ_2 σ-endlich ist, existiert eine Folge $(B_n)_{n \in \mathbb{N}} \subset \mathfrak{A}_2$ mit $B_n \uparrow \Omega_2,\ \mu_2(B_n) < \infty\ \forall n$.

Setze $f_A(\omega_1) := \mu_2(A_{\omega_1}),\ f_{A,n}(\omega_1) := \mu_2(A_{\omega_1} \cap B_n)$.

Sei $\mathfrak{D}$ die Menge aller $A \in \mathfrak{A}$ mit $(\mathfrak{A}_1, \mathfrak{B})$-meßbaren f_A. Man kann leicht zeigen: $\mathfrak{D}$ ist ein Dynkin-System, das den $\cap$-stabilen Erzeuger $\{A_1 \times A_2 : A_i \in \mathfrak{A}_i\}$ von $\mathfrak{A}_1 \otimes \mathfrak{A}_2$ enthält, also $\mathfrak{D} = \mathfrak{A}$.

Also: $f_{A,n}$ ist meßbar $\forall A \in \mathfrak{A},\ n \in \mathbb{N}$ und mit $f_A = \sup f_{A,n}$ folgt die Behauptung. □

Satz und Definition 9.5 Sind $\mu_1,\ \mu_2$ σ-endlich, so existiert genau ein Maß $\mu = \mu_1 \otimes \mu_2$ auf $\mathfrak{A}_1 \otimes \mathfrak{A}_2$, das Produktmaß von μ_1 und μ_2 mit der Eigenschaft $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \quad \forall A_i \in \mathfrak{A}_i$.
Weiter gilt für alle $A \in \mathfrak{A},\ \mu(A) = \int \mu_2(A_{\omega_1})\mu_1(d\omega_1) = \int \mu_1(A_{\omega_2})\mu_2(d\omega_2)$.

Beweis: $f_A(\omega_1) := \mu_2(A_{\omega_1})$. Sind $(A_n)_{n \in \mathbb{N}} \subset \mathfrak{A}$ paarweise disjunkt, $A = \sum_{n=1}^{\infty} A_n$, so gilt:

$$\int f_A d\mu_1 \overset{\mu_2 \text{ stetig von unten}}{=} \int \lim_{n \to \infty} f_{\bigcup_{i=1}^n A_i} d\mu_1 \overset{\text{monotone Konvergenz}}{=} \lim_{n \to \infty} \int f_{\bigcup_{i=1}^n A_i} d\mu_1 = \lim_{n \to \infty} \sum_{i=1}^{n} \int f_{A_i} d\mu_1 = \sum_{i=1}^{\infty} \int f_{A_i} d\mu_1.$$

Außerdem $\int f_\emptyset d\mu_1 = 0$. Also ist $\Pi : \mathfrak{A} \to [0, \infty],\ \Pi(A) := \int f_A d\mu_1$ ein Maß auf $\mathfrak{A}$. Für Π gilt:
$\Pi(A_1 \times A_2) = \int f_{A_1 \times A_2} d\mu_1 = \int \mu_2(A_2) 1_{A_1} d\mu_1 = \mu_2(A_2) \cdot \mu_1(A_1)$.

Analog wird durch $\Pi'(A) = \int \mu_1(A_{\omega_2})\mu_2(d\omega_2)$ ein Maß auf $\mathfrak{A}$ definiert, das auf den Rechteckmengen mit Π übereinstimmt. Das System der Rechteckmengen ist ein $\cap$-stabiler Erzeuger von $\mathfrak{A}$, Eindeutigkeit folgt mit Aufgabe 2 (hier geht die σ-Endlichkeit von $\mu_1,\ \mu_2$ ein.)

Mit μ_1 und μ_2 ist auch $\mu_1 \otimes \mu_2$ σ-endlich. Wie integriert man bzgl. $\mu_1 \otimes \mu_2$?

Ist $f : \Omega \to \mathbb{R}$ eine Abbildung, so schreiben wir f_{ω_1}, f_{ω_2} für die Abbildung
$f_{\omega_1} : \Omega_2 \to \mathbb{R}$, $f_{\omega_1}(\omega_2) := f(\omega_1, \omega_2)$, $f_{\omega_2} : \Omega_1 \to \mathbb{R}$, $f_{\omega_2}(\omega_1) := f(\omega_1, \omega_2)$.
Setzt man $f = 1_A$, so sieht man, das dies den Schnittbegriff bei Mengen verallgemeinert.

Lemma 9.6
Ist f $(\mathfrak{A}, \mathfrak{B})$-meßbar, so ist f_{ω_1} $(\mathfrak{A}_2, \mathfrak{B})$-meßbar für alle $\omega_1 \in \Omega_1$ und f_{ω_2} $(\mathfrak{A}_1, \mathfrak{B})$-meßbar für alle $\omega_2 \in \Omega_2$.

Beweis: $f_{\omega_1}^{-1}(B) = \{\omega_2 \in \Omega_2 : f_{\omega_1}(\omega_2) \in B\} = (\{\omega \in \Omega : f(\omega) \in B\})_{\omega_1} = \underset{\in \mathfrak{A}}{(f^{-1}(B))_{\omega_1}} \overset{\text{Lemma 9.2}}{\in} \mathfrak{A}_2$. □

Satz 9.7 (Fubini I) μ_1, μ_2 seien σ-endlich, $f : \Omega \to \mathbb{R}$ nicht-negativ und $(\mathfrak{A}, \mathfrak{B})$-meßbar.
Dann ist $\omega_1 \to \int f_{\omega_1} d\mu_2$ $(\mathfrak{A}_1, \overline{\mathfrak{B}})$-meßbar, $\omega_2 \to \int f_{\omega_2} d\mu_1$ $(\mathfrak{A}_2, \overline{\mathfrak{B}})$-meßbar, und es gilt:
$\int f d(\mu_1 \otimes \mu_2) = \int (\int f_{\omega_2} d\mu_1) \mu_2(d\omega_2) = \int (\int f_{\omega_1} d\mu_2) \mu_1(d\omega_1)$.

Beweis: Im Falle $f = \sum_{i=1}^{n} a_i 1_{A_i}$ erhält man $\int f_{\omega_2} d\mu_1 = \sum_{i=1}^{n} a_i \int 1_{(A_i)_{\omega_2}} d\mu_1 = \sum_{i=1}^{n} a_i \mu_1((A_i)_{\omega_2})$, also die Meßbarkeit von $\omega_2 \to \int f_{\omega_2} d\mu_1$, und $\int (\int f_{\omega_2} d\mu_1) \mu_2(d\omega_2) = \sum_{i=1}^{n} a_i \int \mu_1((A_i)_{\omega_2}) \mu_2(d\omega_2) = \sum_{i=1}^{n} a_i \mu_1 \otimes \mu_2(A_i) = \int f d(\mu_1 \otimes \mu_2)$.

Ist $f \geq 0$ $(\mathfrak{A}, \mathfrak{B})$-meßbar, so existiert eine Folge $(f_n)_{n \in \mathbb{N}} \subset Prim_+(\Omega, \mathfrak{A})$ mit $f_n \uparrow f$, $\int f d\mu = \lim_{n \to \infty} \int f_n d\mu$.

Wegen $(f_n)_{\omega_2} \uparrow f_{\omega_2}$ und $(f_n)_{\omega_2} \in Prim_+(\Omega, \mathfrak{A})$, folgt mit monotoner Konvergenz $g_n(\omega_2) := \int (f_n)_{\omega_2} d\mu_1 \uparrow \int f_{\omega_2} d\mu_1 \quad \forall \omega_2 \in \Omega_2$.

Nach dem bereits bewiesenen Teil gilt $\int g_n d\mu_2 = \int f_n d\mu$.

Wieder monotone Konvergenz liefert $\int (\int f_{\omega_2} d\mu_1) \mu_2(d\omega_2) = \lim_{n \to \infty} \int g_n d\mu_2 = \lim_{n \to \infty} \int f_n d\mu = \int f d\mu$.

Wiederholt man dies mit ω_1 anstelle von ω_2, so folgt der Rest der Behauptung. □

(Marginale) Erweiterung des Integralbegriffs:
Ist $(\Omega, \mathfrak{A}, \mu)$ ein Maßraum, $A \in \mathfrak{A}$ mit $\mu(A^c) = 0$, $f : \Omega \to \mathbb{R}$, so nennen wir f $(\mathfrak{A}, \mathfrak{B})$-meßbar, μ-integrierbar, etc. ..., wenn dies auf $\tilde{f} : \Omega \to \mathbb{R}$, $\tilde{f}(\omega) := \begin{cases} f(\omega) & , \quad \omega \in A \\ 0 & , \quad \text{sonst} \end{cases}$ zutrifft.

Wir schreiben dann $\int f d\mu$ anstelle von $\int \tilde{f} d\mu$. Hierdurch wird die Integration von Funktionen möglich, die nur fast überall definiert sind (Rechenregeln bleiben erhalten.)

Satz 9.8 (Fubini II)
μ_1, μ_2 σ-endlich, $\mu := \mu_1 \otimes \mu_2$, $f : \Omega \to \mathbb{R}$ sei μ-integrierbar. Dann sind μ_1-fast-alle f_{ω_1} μ_2-integrierbar, μ_2-fast-alle f_{ω_2} μ_1-integrierbar und es gilt: $\int f d\mu = \int (\int f_{\omega_2} d\mu_1) \mu_2(d\omega_2) = \int (\int f_{\omega_1} d\mu_2) \mu_1(d\omega_1)$.

Beweis: Reduktion auf Fubini I mit $f = f^+ - f^-$, etc. ... □

Man kann Fubini-Formel auch in der Form $\int f d\mu = \iint f(\omega_1, \omega_2) \mu_1(d\omega_1) \mu_2(d\omega_2) = \iint f(\omega_1, \omega_2) \mu_2(d\omega_2) \mu_1(d\omega_1)$ schreiben. Also: Integration bzgl. des Produktmaßes kann nacheinander über die Einzelkomponenten ausgeführt werden; Die Reihenfolge ist dabei egal.

9.2 Produkte von mehr als zwei Maßräumen

$(\Omega_i, \mathfrak{A}_i, \mu_i)$ σ-endliche Maßräume für $i = 1, \ldots, n$.
$(\ldots((\mathfrak{A}_1 \otimes \mathfrak{A}_2) \otimes \mathfrak{A}_3) \otimes \ldots) \otimes \mathfrak{A}_n$ ist eine σ-Algebra auf $(\ldots((\Omega_1 \times \Omega_2) \times \Omega_3) \times \ldots) \times \Omega_n$.

Identifiziert man dies mit $\Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$, so erhält man, daß die Produkt-σ-Algebra von den Rechteckmengen $A_1 \times A_2 \times \ldots \times A_n$, $A_i \in \mathfrak{A}_i$ für $i = 1. \ldots, n$ erzeugt wird und schreibt kurz: $\mathfrak{A} := \mathfrak{A}_1 \otimes \ldots \otimes \mathfrak{A}_n$ ($\otimes$ ist assoziativ.) Mit Induktion erhält man die Existenz (und Eindeutigkeit) eines Maßes μ mit $\mu(A_1 \times \ldots \times A_n) = \mu_1(A_1) \cdot \ldots \cdot \mu_n(A_n)$.

Schreibweise: $\mathfrak{A} = \bigotimes_{i=1}^{n} \mathfrak{A}_i$, $\mu = \bigotimes_{i=1}^{n} \mu_i$.

Der Satz von Fubini gilt in der entsprechenden Form.

Beispiel 9.9 $(\Omega_i, \mathfrak{A}_i, \mu_i) = (\mathbb{R}, \mathfrak{B}, l)$ für $i = 1, \ldots, n$.
$\underset{i=1}{\overset{n}{\times}} \Omega_i = \mathbb{R}^n$, $\bigotimes_{i=1}^{n} \mathfrak{B}$ $(= \mathfrak{B}^{\otimes n})$ $=: \mathfrak{B}^n$ ist die σ-Algebra der n-dimensionalen Borel-Mengen und $l^n := l^{\otimes n}$ ist das n-dimensionale Lebesgue-Maß.
$\mathfrak{B}^n$ wird erzeugt von $I^n := \{(a, b] : a, b \in \mathbb{R}^n\}$, $(a, b] := \{(x_1, \ldots, x_n) : a_i < x_i \leq b_i \text{ für } i = 1, \ldots, n\}$.
WMaße P auf $(\mathbb{R}^n, \mathfrak{B}^n)$ lassen sich durch die Verteilungsfunktion $F : \mathbb{R}^n \to [0, 1]$, $F(x) := P((-\infty, x])$ charakterisieren.
Ist $(\Omega, \mathfrak{A}, P)$ ein WRaum und $X : \Omega \to \mathbb{R}^n$ $(\mathfrak{A}, \mathfrak{B}^n)$-meßbar, so heißt X (n-dimensionaler) Zufallsvektor, die Verteilungsfunktion zu $F_X(x) := P^X((-\infty, x]) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$ heißt auch Verteilungsfunktion zu X, usw.
Hat P^X eine Dichte f bzgl. l^n, so nennt man f eine Dichte zu X; X heißt dann absolutstetig verteilt.

Schwieriger: Produkte von unendlich vielen Maßräumen.
Einschränkung auf WRäume; Sei also $I \neq \emptyset$ und für jedes $i \in I$ sei $(\Omega_i, \mathfrak{A}_i, P_i)$ ein WRaum.
Für jedes $J \subset I$ wird durch $\Pi_J : \Omega \to \underset{i \in J}{\times} \Omega_i$, $\Pi_J(\omega) := \omega|_J$ die Projektion auf die J-Koordinaten definiert.
Die Produkt-σ-Algebra $\mathfrak{A} := \bigotimes_{i \in I} \mathfrak{A}_i$ auf Ω ist die von den Zylindermengen mit endlicher Basis $\Pi_J^{-1}(A_J)$, $A_J \in \bigotimes_{i \in J} \mathfrak{A}_i$, J

endlich, erzeugte σ-Algebra.
$\mathfrak{A}$ wird auch von den Mengen $\Pi_i{}^{-1}(A_i) := \Pi_{\{i\}}^{-1}(A_i),\ A_i \in \mathfrak{A}_i$ erzeugt. Dieses System ist aber nicht $\cap$-stabil. Die Produkt-σ-Algebren werden von den Projektionen erzeugt.

Satz 9.10 Auf $(\Omega, \mathfrak{A})$ existiert genau ein WMaß $P = \bigotimes_{i\in I} P_i$ mit $P^{\Pi_J} = \bigotimes_{i\in J} P_i$ für alle endlichen $J \subset I$.

Beweis: Ähnlich wie bei der Existenz von l, siehe Bauer. □

Münzwurf: $(\{0,1\}, \mathbb{P}(\{0,1\}), \frac{1}{2}(\delta_0 + \delta_1))$.
Der unendlich oft wiederholte Münzwurf: $(\{0,1\}, \mathbb{P}(\{0,1\}), \frac{1}{2}(\delta_0 + \delta_1))^{\mathbb{N}}$, $\Omega = \{0,1\}^{\mathbb{N}}$ Menge aller 0-1-Folgen.
Durch $\{0,1\}^{\mathbb{N}} \ni (\omega_i)_{i\in\mathbb{N}} \xrightarrow{T} \sum_{i=1}^{\infty} \omega_i \cdot 2^{-i} \in [0,1]$ wird eine $(\mathfrak{A}, \mathfrak{B}_{[0,1]})$-meßbare Abbildung mit der Eigenschaft $((\frac{1}{2}(\delta_0 + \delta_1))^{\otimes\mathbb{N}})^T = l|_{[0,1]}$ (siehe auch Bsp. 5.30) definiert.

9.3 Unabhängigkeit

Erinnerung (Def. 5.27):

- $\{\mathfrak{A}_i : i \in I\}$ heißt unabhängig, wenn $P(\bigcap_{i\in J} A_i) = \prod_{i\in J} P(A_i),\ A_i \in \mathfrak{A}_i,\ \forall J \subset I,\ J$ endlich, gilt.
- Ist $(\Omega_i, \mathfrak{A}_i)$ ein meßbarer Raum $\forall i \in I$ und $X_i : \Omega \to \Omega_i$ $(\mathfrak{A}, \mathfrak{A}_i)$-meßbar, so heißt $\{X_i : i \in I\}$ unabhängig, wenn die σ-Algebren $\sigma(X_i) = \{X_i{}^{-1}(A_i) : A_i \in \mathfrak{A}_i\}$ unabhängig sind.

Klar: Eine Familie von σ-Algebren/Zufallsgrößen ist genau dann unabhängig, wenn jede endliche Teilfamilie unabhängig ist. Beim Nachweis der Unabhängigkeit kann man sich auf $\cap$-stabile Erzeugendensysteme beschränken ($\rightsquigarrow$ Satz 5.28.)

Beispiel 9.11 ("∞-oft wiederholter Münzwurf")
Es sei $\Omega = \{0,1\}^{\mathbb{N}}$, $\mathfrak{A} = \mathbb{P}(\{0,1\})^{\otimes\mathbb{N}}$, $P = (\frac{1}{2}(\delta_0 + \delta_1))^{\otimes\mathbb{N}}$. Definiere $X_n : \Omega \to \mathbb{R}$ durch $X_n((\omega_i)_{i\in\mathbb{N}}) := \omega_n$ (Maßtheorie: Projektion auf die n-te Koordinate, Stochastik: Resultat des n-ten Wurfes). X_n ist nach Konstruktion meßbar.
Klar: $P(X_n = 0) = P(X_n = 1) = \frac{1}{2}$.
Ein $\cap$-stabiles Erzeugendensystem von $\mathfrak{A} = \sigma(X_i)$ ist $\mathfrak{E}_i := \{A_i\}$ mit $A_i = \{X_i = 1\}$.
Ist $J \subset \mathbb{N}$ endlich, $J = \{i_1, \ldots, i_n\}$, so gilt:

$$P(\bigcap_{i\in J} A_i) = P(X_{i_1} = 1, \ldots X_{i_n} = 1) = P^{\Pi_J}(\{1\} \times \ldots \times \{1\}) = (\tfrac{1}{2})^n = \prod_{j=1}^{n} P(X_{i_j} = 1).$$

Also: $\{X_n : n \in \mathbb{N}\}$ ist eine unabhängige Familie. Außerdem ist $X := (X_1, \ldots, X_n)$ ein n-dimensionaler Zufallsvektor ($\rightsquigarrow$ Aufgabe 15). Die Verteilung von X heißt gemeinsame Verteilungsfunktion/Dichte von $X_1, \ldots, X_n$.

Satz 9.12 (verallgemeinert Satz 4.18)
Es seien $X_1, \ldots, X_n$ ZV auf $(\Omega, \mathfrak{A}, P)$.

(i) Dann gilt: $X_1, \ldots, X_n$ sind unabhängig $\iff P^{(X_1,\ldots,X_n)} = P^{X_1} \otimes \ldots \otimes P^{X_n}$.

(ii) Es sei F_i die Verteilungsfunktion zu X_i $(1 \le i \le n)$, F die gemeinsame Verteilungsfunktion. Dann gilt:

$$X_1, \ldots, X_n \text{ unabhängig} \iff F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_i(x_i) \qquad \forall x \in \mathbb{R}^n.$$

(iii) Es sei f_i eine Dichte von X_i; $f(x) := \prod_{i=1}^{n} f_i(x)$ für $x = (x_1, \ldots, x_n)$.
Dann $X_1, \ldots, X_n$ sind unabhängig $\iff f$ ist gemeinsame Dichte.

Beweis: $J := \{(-\infty, a] : a \in \mathbb{R}\}$ ist ein $\cap$-stabiles Erzeugendensystem zu $\mathfrak{B}$, also ist $\{X_i{}^{-1}((-\infty, a]) : a \in \mathbb{R}\}$ ein $\cap$-stabiles Erzeugendensystem zu $\sigma(X_i)$.
(i) und (ii) folgen aus Satz 5.28 und der Definition des Produktmaßes.
Teil (iii) folgt mit Fubini I aus Teil (ii). □

Bekannt: Funktionen von unabhängigen Zufallsgrößen sind wieder unabhängig.

Satz 9.13 (Multiplikationssatz, verallgemeinert Satz 4.19)
Für unabhängige Zufallsvariablen X, Y gilt $E(XY) = (EX)(EY)$ (vorausgesetz die Erwartungswerte existieren).

Beweis: $EXY = \int X(\omega)Y(\omega)dP =^{1} \int xyP^{(X,Y)}(dx, dy) \overset{9.12(i)}{=} \int xyP^X \otimes P^Y(dx, dy) \overset{\text{Fubini II}}{=} \int (\int xyP^X(dx))P^Y(dy) = \int y(\int xP^X(dx))P^Y(dy) = EX \int yP^Y(dy) = (EX)(EY)$. □

In Verallgemeinerung von Def. 4.21 nennen wir den Erwartungswert
$cov(X,Y) = E((X - EX)(Y - EY))$ $(= E(XY) - (EX)(EY))$ die Kovarianz der ZV X und Y.
Im Falle $cov(X,Y) = 0$ heißen X und Y unkorreliert.
Im Falle $var(X) > 0,\ var(Y) > 0$ heißt $\rho(x,y) := \frac{cov(X,Y)}{\sqrt{var(X)\cdot var(Y)}}$ der Korrelationskoeffizient.

$$\text{Unabhängig} \overset{\Rightarrow}{\not\Leftarrow} \text{Unkorreliert.}$$

$$var(X_1 + \ldots + X_n) = \sum_{i=1}^{n} var(X_i) + \sum_{\substack{i,j=1 \\ i\neq j}}^{n} cov(X_i, X_j) \qquad \rightsquigarrow \text{Bienaymé.}$$

[1] Transformationsformel, Verlagerung der Integration auf den Bildbereich

Satz 9.14 (Das Borel-Cantelli-Lemma)

$(\Omega, \mathfrak{A}, P)$ WRaum, $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{A}$ eine Folge von Ereignissen; $\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$ (unendlich viele der A_k's treten ein, $\longrightarrow$ Bsp. 1.2(i)). Dann gilt:

(i) $\sum_{n=1}^{\infty} P(A_n) < \infty \quad \Longrightarrow \quad P(\limsup_{n\to\infty} A_n) = 0$

(ii) $\sum_{n=1}^{\infty} P(A_n) = \infty$ und $\{A_n | n \in \mathbb{N}\}$ unabhängig $\quad \Longrightarrow \quad P(\limsup_{n\to\infty} A_n) = 1.$

Beweis: (i) Aufgabe 5c) zur Stochastik I.

(ii) Sei $p_n := P(A_n)$. Für alle $n \in \mathbb{N}$ gilt (mit $1 - p \le \exp(-p) \qquad \forall p \in \mathbb{R}$):

$$0 \le P(\bigcap_{k=n}^{\infty} A_k^c) \overset{\text{Stetigkeit von unten}}{=} \lim_{N\to\infty} P(\bigcap_{k=n}^{N} A_k^c) = \lim_{N\to\infty} \prod_{k=n}^{N} (1 - p_k) \le \lim_{N\to\infty} \exp(-\sum_{k=n}^{N} p_k) = 0.$$

also folgt: $0 \le P((\limsup_{n\to\infty} A_n)^c) = P(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c) \le \sum_{n=1}^{\infty} P(\bigcap_{k=n}^{\infty} A_k^c) = 0.$ □

Kapitel 10

Gesetze der großen Zahlen

10.1 Konvergenz fast-sicher und Konvergenz in Wahrscheinlichkeit

Gesetz der großen Zahlen gibt es in "starker" und "schwacher" Form; sie beziehen sich auf unterschiedliche Konvergenzarten.

Definition 10.1 Es seien $X,\ X_1,\ X_2, \ldots$ ZV auf $(\Omega, \mathfrak{A}, P)$.

(i) X_n konvergiert P-fast-sicher gegen X, Schreibweise: $X_n \xrightarrow{\text{f.s.}} X$, wenn gilt:
$P(\{\omega \in \Omega : \lim\limits_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$.

(ii) X_n konvergiert in Wahrscheinlichkeit gegen X, Schreibweise $X_n \xrightarrow{P} X$, wenn gilt:
$\lim\limits_{n\to\infty} P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}) = 0$ für alle $\epsilon > 0$.

Satz 10.2 (Konvergenz fast-sicher impliziert Konvergenz in Wahrscheinlichkeit, die Umkehrung gilt nicht)

Beweis: $\xrightarrow{\text{f.s.}} \Longrightarrow \xrightarrow{P}$

Sei $N := \{\omega \in \Omega : X_n(\omega) \not\to X(\omega)\}$, nach Voraussetzung: $P(N) = 0$.
Sei $\epsilon > 0$ und setze $A_n := \{\omega \in \Omega : \sup\limits_{m\geq n} |X_m(\omega) - X(\omega)| \geq \epsilon\}$.
Klar: $A_n \supseteq A_{n+1}$. Für reelle Zahlenfolgen $(a_n)_{n\in\mathbb{N}}$ gilt $\lim\limits_{n\to\infty} a_n = a \iff \lim\limits_{n\to\infty} \sup\limits_{m\geq n} |a_m - a| = 0$, also:

$$A_n \downarrow N_\epsilon \subseteq \{\omega \in \Omega : \limsup |X_n(\omega) - X(\omega)| \geq \epsilon\} \subseteq N \qquad (N_\epsilon := \bigcap_{n=1}^{\infty} A_n)$$

Also: $0 \leq P(|X_n - X| \geq \epsilon) \leq P(A_n) \longrightarrow P(N_\epsilon) \leq P(N) = 0$.

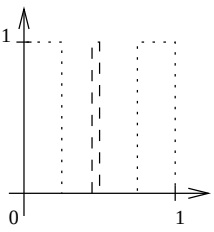
Gegenbeispiel zu $\xrightarrow{P} \Longrightarrow \xrightarrow{\text{f.s.}}$: $(\Omega, \mathfrak{A}, P) = ([0,1), \mathfrak{B}|_{[0,1)}, l|_{[0,1)})$.

Jedes $n \in \mathbb{N}$ läßt sich eindeutig in der Form $n = 2^m + k,\ m \in \mathbb{N}_0,\ k \in \{0, \ldots, 2^m - 1\}$ schreiben. Sei $X_n := 1_{[k\cdot 2^{-m}, (k+1)\cdot 2^{-m})},\ X = 0$.

Dann gilt $\forall \epsilon > 0 :\ P(|X_n - X| > \epsilon) = P([k \cdot 2^{-m}, (k+1) \cdot 2^{-m})) = 2^{-m} \overset{n\to\infty}{\longrightarrow} 0$, also $X_n \xrightarrow{P} X$.

Andererseits gibt es zu jedem $x \in [0,1)$ und $m \in \mathbb{N}$ ein $k \in \{0, \ldots, 2^m - 1\}$ mit $x \in [k \cdot 2^{-m}, (k+1) \cdot 2^{-m}),\ x \notin [j \cdot 2^{-m}, (j+1) \cdot 2^{-m})$ für $j \neq k$, d.h. die Folge $(X_n(x))_{n\in\mathbb{N}}$ nimmt ∞-oft den Wert 0 und ∞-oft den Wert 1 an.

Also: $P(\{\omega \in \Omega : X_n(\omega) \longrightarrow X(\omega)\}) = 0$. □



10.2 Das schwache Gesetz der großen Zahlen

Sammelbegriff für Resultate der Form $\frac{1}{a_n}(\sum\limits_{i=1}^{n} X_i - b_n) \xrightarrow{P} 0$, mit geeigneten $(a_n)_{n\in\mathbb{N}},\ (b_n)_{n\in\mathbb{N}} \subset \mathbb{R}$.

Ein einfaches Beispiel, das den diskreten Fall (Satz 4.28) verallgemeinert.

Satz 10.3 Es sei $EX_n{}^2 < \infty\ \forall n \in \mathbb{N}$, sowie $cov(X_i, X_j) = 0$ für $i \neq j$. Dann gilt:

$$\lim_{n\to\infty} \tfrac{1}{n^2} \sum_{i=1}^{n} var(X_i) = 0 \quad \Longrightarrow \quad \tfrac{1}{n} \sum_{i=1}^{n} (X_i - EX_i) \xrightarrow{P} 0.$$

Beweis: Sei $Y_n := \frac{1}{n} \sum\limits_{i=1}^{n} (X_i - EX_i)$.

Es gilt $EY_n = 0,\ var(Y_n) = \frac{1}{n^2} \sum\limits_{i=1}^{n} var(X_i) + \underbrace{\sum\limits_{i\neq j} cov(X_i, X_j)}_{=0}$

Mit der Ungleichung von Chebychev: $P(|Y_n| > \epsilon) \leq \frac{1}{\epsilon^2} var(Y_n) \overset{n\to\infty}{\longrightarrow} 0$. □

10.3 Das starke Gesetz der großen Zahlen

Sammelbegriff für Resultate der Form $\frac{1}{a_n}(\sum_{i=1}^{n} X_i - b_n) \xrightarrow{\text{f.s.}} 0$, wieder mit geeigneten $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}} \subset \mathbb{R}$.

Die Partialsummen $S_n := \sum_{i=1}^{n} X_i$ werden durch $(b_n)_{n\in\mathbb{N}}$ in der 'Lage' und durch $(a_n)_{n\in\mathbb{N}}$ in der 'Skala' verändert.

Wichtigstes Beispiel:

Satz 10.4 Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger und identisch verteilter ZV mit $E|X_1| < \infty$, so gilt:

$\overline{X_n}$ $(= \frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^{n} X_i) \xrightarrow{\text{f.s.}} EX_1$.

Beweis: Wir nehmen zunächst $X_n \geq 0$ $\forall n \in \mathbb{N}$ an.

$Y_k := X_k \cdot 1_{\{X_k \leq k\}}$ ('bei k abgeschnitten')

$S_n^* := \sum_{k=1}^{n} Y_k$. Mit monotoner Konvergenz folgt: $EY_k = EX_k \cdot 1_{\{X_k \leq k\}} = EX_1 \cdot 1_{\{X_1 \leq k\}} \xrightarrow{k\to\infty} EX_1$.

Aus der Analysis bekannt: $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ mit $a_n \to a \implies \frac{1}{n}\sum_{i=1}^{n} a_i \to a$.[1], d.h.

(1) $\lim_{n\to\infty} \frac{1}{n} ES_n^* = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} EY_k = EX_1$.

(2) Die Y-Variablen sind wieder unabhängig (aber i.A. nicht mehr identisch verteilt), also

$var(S_n^*) = \sum_{i=1}^{n} var(Y_i) \leq \sum_{i=1}^{n} {EY_i}^2 = \sum_{k=1}^{n} {EX_k}^2 \cdot 1_{\{X_k \leq k\}} \leq \sum_{k=1}^{n} {EX_k}^2 \cdot 1_{\{X_k \leq n\}} = n{EX_1}^2 \cdot 1_{[0,n]}(X_1)$.

Wähle $\alpha > 1$ und setze $m_n := \lfloor \alpha^n \rfloor$, $\Psi(x) := \sum_{n=1}^{\infty} \frac{1}{m_n} \cdot 1_{[0,m_n]}(x)$, $N(x) = \min\{n : m_n \geq x\}$.

Wegen $\frac{1}{m_n} = \frac{1}{\lfloor \alpha^n \rfloor} \leq \frac{2}{\alpha^n}$ und $\alpha^{N(x)} \geq \lfloor \alpha^{N(x)} \rfloor = m_{N(x)} \geq x$ folgt

(3) $\Psi(x) = \sum_{n=N(x)}^{\infty} \frac{1}{m_n} \leq 2 \sum_{n=N(x)}^{\infty} \alpha^{-n} = 2\alpha^{-N(x)} \frac{1}{1-\alpha^{-1}} \leq \frac{k}{x}$ mit $k := \frac{2\alpha}{\alpha-1}$.

Mit Chebychev folgt nun für alle $\epsilon > 0$: $\sum_{n=1}^{\infty} P(\frac{1}{m_n}|S_{m_n}^* - ES_{m_n}^*| \geq \epsilon) \overset{(2)}{\leq} \sum_{n=1}^{\infty} \frac{1}{\epsilon^2 m_n} {EX_1}^2 \cdot 1_{[0,m_n]}(X_1)$

$= \frac{{EX_1}^2}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{m_n} 1_{[0,m_n]}(X_1) = \frac{1}{\epsilon^2} {EX_1}^2 \Psi(X_1) \overset{(3)}{\leq} kEX_1 < \infty$.

Hieraus folgt wegen Aufgabe 18(ii) $\frac{1}{m_n}(S_{m_n}^* - ES_{m_n}^*) \longrightarrow 0$ P-f.s. und wegen (1) und Aufgabe 18(i)

(4) $\frac{1}{m_n} S_{m_n}^* \longrightarrow EX_1$ P-f.s.

Wie wollen "den $*$ loswerden."

$\sum_{n=1}^{\infty} \underbrace{P(X_n \neq Y_n)}_{P(X_n > n)} = \sum_{n=1}^{\infty} P(X_1 > n) \leq \int_0^{\infty} P(X_1 \geq x)dx = EX_1 < \infty$ ($\rightsquigarrow$ Aufgabe 14)

Mit Borel-Cantelli folgt, daß $N_0 := \{\omega \in \Omega : X_n(\omega) \neq Y_n(\omega)$ für unendlich viele $n\}$ eine P-Nullmenge ist. Für $\omega \in {X_0}^c$ existiert somit ein $k(\omega) < \infty$ mit $X_n(\omega) = Y_n(\omega)$ $\forall n \geq k(\omega)$. Auf ${N_0}^c$ gilt also $\frac{1}{n}(S_n(\omega) - S_n^*(\omega)) = \frac{1}{n}(\sum_{k=1}^{k(\omega)} X_k(\omega) - Y_k(\omega)) \overset{n\to\infty}{\longrightarrow} 0$, d.h. $\frac{1}{n}(S_n - S_n^*) \longrightarrow 0$ P.-f.s. und mit (4) folgt

(5) $\frac{1}{m_n} S_{m_n} \longrightarrow EX_1$ P-f.s.

Als nächstes: Einschränkung auf Teilfolgen "loswerden".

$\frac{m_n}{m_{n+1}} \frac{S_{m_n}}{m_n} \leq \frac{S_k}{k} \leq \frac{m_{n+1}}{m_n} \frac{S_{m_{n+1}}}{m_{n+1}}$ für $m_n \leq k \leq m_{n+1}$ $\quad((S_k)_{k\in\mathbb{N}}$ hat Zuwächse $\geq 0)$

Mit $\frac{m_n}{m_{n+1}} \overset{n\to\infty}{\longrightarrow} \frac{1}{\alpha}$ folgt $\quad \frac{1}{\alpha}EX_1 \leq \liminf_{k\to\infty} \frac{S_k}{k}$

$\limsup_{k\to\infty} \frac{S_k}{k} \leq \alpha EX_1$ $\quad P$-f.s. $\forall \alpha > 1$ wegen (5).

Sei N_α die zugehörige Ausnahme-Nullmenge. Außerhalb der Nullmenge $\bigcup_{j=1}^{\infty} N_{1+\frac{1}{j}}$ gilt dann (mit $\alpha \to 1$)

$EX_1 \leq \liminf \frac{S_k(\omega)}{k} \leq \limsup \frac{S_k(\omega)}{k} \leq EX_1$, also

(6) $\overline{X_n} = \frac{1}{n}S_n \longrightarrow EX_1$ P-f.s.

Es bleibt die Einschränkung auf $X_i \geq 0$ zu beseitigen:

$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} {X_i}^+ - \frac{1}{n}\sum_{i=1}^{n} {X_i}^- \overset{(6)}{\longrightarrow} {EX_1}^+ - {EX_1}^- = EX_1$. □

Beispiel 10.5 (siehe auch Beispiel 9.11)

$(\Omega, \mathfrak{A}, P) = (\{0,1\}, \mathbb{P}(\{0,1\}), \frac{1}{2}(\delta_0 + \delta_1))^{\otimes\mathbb{N}}$ Modell für ∞-oft wiederholten Wurf einer fairen Münze.

X_n : Projektion auf n-te Koordinate (Resultat von Wurf Nr. n)

$(X_n)_{n\in\mathbb{N}}$ genügt den Voraussetzungen von Satz 10.4, also $\frac{1}{n}\#\{1 \leq i \leq n : X_i = 1\}$[2] $\longrightarrow EX_1 = \frac{1}{2}$ P-f.s.

---

[1] César-Mittelwerte haben den gleichen limes.

[2] $\frac{1}{n}S_n$, Anzahl der Kopfwürfe in den ersten n Versuchen.

In maßtheoretischer Sprechweise: Das Produktmaß P ist auf die Menge der 0-1-Folgen konzentriert, bei denen die relative Häufigkeit von Kopf gegen $\frac{1}{2}$ konvergiert.
In wtheoretischer Sprechweise: Die W. dafür, daß die relative Häufigkeit des Ereignisses Kopf gegen seine W. konvergiert, ist 1.
(Läßt sich auf $(1-p)\delta_0 + p\delta_1$ verallgemeinern, d.h. obige Aussagen gelten für Ereignisse mit beliebiger W. p.)

Beispiel 10.6 (siehe auch Aufgabe 28 zur Stochastik I)
Es sei $p \in \mathbb{N},\ p > 1$. Zu jedem $x \in [0,1)$ existiert eine eindeutige Entwicklung zur Basis p,

$$x = \sum_{k=1}^{\infty} a_k p^{-k},\ a_k \in \{0, \dots, p-1\},\ \#\{k : a_k \neq p-1\} = \infty.$$

Es sei B die Menge der zur Basis p normalen Zahlen, $B := \{x \in [0,1) : \lim_{n\to\infty} \frac{1}{n}\#\{1 \le k \le n : a_k = q\} = \frac{1}{p}$ für $q = 0, \dots, p-1\}$ (grob: alle möglichen Ziffern kommen gleichoft vor.)
$B \in \mathfrak{B}$ folgt mit Aufgabe 28 zur Stochastik I, also $l(B) = ?$
$(\Omega, \mathfrak{A}, P) = ([0,1), \mathfrak{B}|_{[0,1)}, l|_{[0,1)})$. Für jedes $n \in \mathbb{N},\ \omega \in \Omega$ sei $X_n(\omega)$ die n-te Ziffer der Entwicklung von ω. Dann ist $(X_n)_{n\in\mathbb{N}}$ eine iid Folge mit $P(X_n = q) = \frac{1}{p}$ für $q = 0, \dots, p-1$.
Sei $q \in \{0, \dots, p-1\},\ Y_n := 1_{\{q\}}(X_n)$. Dann ist $(Y_n)_{n\in\mathbb{N}}$ wieder eine iid Folge und es gilt:

$$P(B_q) = 1 \text{ mit } B_q = \{\omega \in \Omega : \lim_{n\to\infty} \underbrace{\frac{1}{n}\sum_{k=1}^{n} Y_k(\omega)}_{\frac{1}{n}S_n} = \underbrace{\frac{1}{p}}_{EY_1}\}$$

Wegen $\sum_{k=1}^{n} Y_k(\omega) = \#\{1 \le k \le n : a_k = q\}$ gilt $B = \bigcap_{q=0}^{p-1} B_q$, also $P(B) = 1$. Beachte: p war beliebig!
Daher: Satz von Borel: l-fast alle Zahlen sind normal in allen Basen $p > 1$.

Kapitel 11

Charakteristische Funktion

11.1 Grundlegendes

$(\Omega, \mathfrak{A}, P)$, $f : \Omega \to \mathbb{C}$, $f = g + ih$ (g, h $\mathbb{R}$-wertig)
$\int f d\mu := \int g d\mu + i \int h d\mu$, wenn g, h μ-integrierbar
Linearität, etc. gilt weiter, auch $Re(\int f d\mu) = \int Re(f) d\mu$.

Lemma 11.1 $|\int f d\mu| \leq \int |f| d\mu$.
Beweis: Annahme $\int f d\mu \neq 0$ (sonst ist nichts zu zeigen)
$Re(\overline{\int f d\mu} \cdot f) \leq |\overline{\int f d\mu} \cdot f| = |\int f d\mu| \cdot |f|$.
$|\int f d\mu|^2 = Re(\overline{\int f d\mu} \cdot \int f d\mu) = \int Re(\overline{\int f d\mu} \cdot f) d\mu \leq \int |\int f d\mu| \cdot |f| d\mu$
$\overset{|\int f d\mu| \neq 0}{\Longrightarrow} |\int f d\mu| \leq \int |f| d\mu$. □

Definition 11.2 (Charakteristische Funktion)
X ZV auf $(\Omega, \mathfrak{A}, P)$. Dann heißt $\varphi_X : \mathbb{R} \to \mathbb{C}$, $\varphi_X(\theta) = E \exp(i\theta X)$ die charakteristische Funhktion zu X.

$\varphi_X(\theta) = E \cos(\theta X) + iE \sin(\theta X)$. Dieser Erwartungswert existiert ($|\ldots| \leq 1$).
$E \exp(i\theta X) = \int_\Omega \exp(i\theta X(\omega)) P(d\omega) \overset{\text{Transformationssatz}}{=} \int_{\mathbb{R}} \exp(i\theta x) P^X(dx)$.

Also: Zufallsvariablen mit derselben Verteilung haben dieselbe charakteristische Funktion.

- Ist X diskret verteilt ($P(X \in A) = 1$ für eine abzählbare Menge $A \subset \mathbb{R}$), so gilt $\varphi_X(\theta) = \sum_{x \in A} \exp(i\theta x) P(X = x)$.
- Ist X absolutstetig verteilt mit Dichte f, so gilt $\varphi_X(\theta) = \int \exp(i\theta x) f(x) dx$. ($\varphi$ ist Fourier-Transformierte zu P^X).

Beispiel 11.3
(i) $X \sim Bin(n,p)$: $\varphi_X(\theta) = \sum_{k=0}^{n} \exp(i\theta k) \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (p \exp(i\theta))^k (1-p)^{n-k} = (1 - p + p \exp(i\theta))^n$.
(ii) $X \sim unif(0,1)$: $\varphi_X(\theta) = \int_0^1 \exp(i\theta x) dx = \begin{cases} \frac{1}{i\theta}(\exp(i\theta) - 1) & \text{für } \theta \neq 0 \\ 1 & \text{für } \theta = 0 \end{cases}$

Satz 11.4 Ist φ_X die charakteristische Funktion zu X, so gilt:
(i) $\varphi_X(0) = 1$
(ii) $|\varphi_X(\theta)| \leq 1 \quad \forall \theta \in \mathbb{R}$
(iii) φ_X ist gleichmäßig stetig auf $\mathbb{R}$
(iv) $\varphi_{(-X)} = \overline{\varphi_X}$
(v) $\varphi_{aX+b}(\theta) = \exp(i\theta b) \varphi_X(a\theta)$
Beweis: (i) $\varphi_X(0) = E \exp(0) = 1$
(ii) $|\varphi_X(\theta)| \overset{\text{Lemma 11.1}}{\leq} E|\exp(i\theta X)| = 1$.
(iii) $|\varphi_X(\theta + h) - \varphi_X(\theta)| = |E \exp(i(\theta + h)X) - \exp(i\theta X)| \leq E|\exp(i\theta X)(\exp(ihX) - 1)| \leq E|\exp(ihX) - 1|$
Sei nun $(h_n)_{n \in \mathbb{N}} \subset \mathbb{R}$ mit $h_n \to 0$. Dann gilt für alle $\omega \in \Omega$: $\cos(h_n X(\omega)) \to 1$, $\sin(h_n X(\omega)) \to 0$.
Alle diese Funktionen sind betragsmäßig durch die P-integrierbare Funktion $\omega \mapsto 1$ beschränkt, also folgt mit majorisierter Konvergenz: $\sup_{\theta \in \mathbb{R}} |\varphi_X(\theta + h_n) - \varphi_X(\theta)| \leq E|\cos(h_n X) - 1| + E|\sin(h_n X)| \to 0$.
(iv) $\varphi_{(-X)}(\theta) = E \exp(i\theta(-X)) = E \exp(i(-\theta)X) = E \cos(-\theta X) + iE \sin(-\theta X) = E \cos(\theta X) - iE \sin(\theta X) = \overline{\varphi_X(\theta)}$
(v) $\varphi_{aX+b}(\theta) = E \exp(i\theta(aX + b)) = \exp(i\theta b) E \exp(i\theta(aX)) = \exp(i\theta b) E \exp(i(\theta a)X) = \exp(i\theta b) \varphi_X(a\theta)$. □

11.2 Ableitungen und Momente

Lemma 11.5 Für alle $n \in \mathbb{N}_0$, $y \in \mathbb{R}$ gilt: $|e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!}| \le \min\{\frac{|y|^{n+1}}{(n+1)!}, \frac{2|y|^n}{n!}\}$.

Beweis: Es reicht, den Fall $y \ge 0$ zu betrachten ($|\overline{z}| = |z|$.) Zunächst:

(1) $e^{iy} = \sum_{k=0}^{n} \frac{(iy)^k}{k!} + \frac{i^{n+1}}{n!} \int_0^y (y-s)^n e^{is} ds$

Bei $n = 0$: $1 + i \int_0^y e^{is} ds = e^{iy}$. Partielle Integration ergibt:

(2) $\int_0^y (y-s)^n e^{is} ds = \frac{y^{n+1}}{n+1} + \frac{i}{n+1} \int_0^y (y-s)^{n+1} e^{is} ds$ liefert Schritt von n nach $n+1$.

$|\int_0^y (y-s)^n e^{is} ds| \le \int_0^y (y-s)^n ds = \frac{1}{n+1} y^{n+1}$.

Mit (1) folgt die Oberschranke $\frac{y^{n+1}}{(n+1)!}$.

In (2) erhält man mit $n-1$ anstelle von n: $\frac{i}{n} \int_0^y (y-s)^n e^{is} ds = \int_0^y (y-s)^{n-1} e^{is} ds - \frac{y_n}{n}$.

Also liefert (1) auch $e^{iy} = \sum_{k=0}^{n} \frac{(iy)^k}{k!} + \frac{i^n}{(n-1)!} \int_0^y (y-s)^{n-1} (e^{is} - 1) ds$ und damit

$|e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!}| \le \frac{1}{(n-1)!} \int_0^y (y-s)^{n-1} \underbrace{|e^{is} - 1|}_{\le 2} ds \le \frac{2}{(n-1)!} \frac{y^n}{n}$. □

Satz 11.6 Im Falle $E|X|^k < \infty$ ist φ_X k-mal differenzierbar und es gilt $\varphi_X^{(k)}(\theta) = E(iX)^k \exp(i\theta X)$, insbesondere $\varphi_X^{(k)}(0) = i^k EX^k$. Außerdem ist $\varphi_X^{(k)}$ gleichmäßig stetig und beschränkt.

Beweis: $\frac{\varphi(\theta+h) - \varphi(\theta)}{h} = E \exp(i\theta X)(\frac{\exp(ihX) - 1}{h})$

Lemma 11.5 ergibt bei $n = 1$: $|\exp(ihx) - 1 - ihx| \le \min\{\frac{h^2 x^2}{2}, 2|hx|\}$, also

$|\frac{\exp(ihx) - 1}{h}| \le 3|x|$, $\frac{\exp(ihX) - 1}{h} \to iX$ (mit $h \to 0$.)

Gilt nun $E|X| \le \infty$, so läßt sich der Satz von der majorisierten Konvergenz anwenden und liefert

$\varphi'(\theta) = \lim_{h \to 0} \frac{\varphi(\theta+h) - \varphi(\theta)}{h} = E \exp(i\theta X)(iX)$.

Hat man bereits "$E|X|^k < \infty \implies \varphi^{(k)}(\theta) = E(iX)^k \exp(i\theta X)$" gezeigt, so folgt bei $E|X|^{k+1} < \infty$ mit denselben Argumenten: $\varphi^{(k+1)}(\theta) = \lim_{h \to 0} \frac{1}{h}(\varphi^{(k)}(\theta + h) - \varphi^{(k)}(\theta)) = \lim_{h \to 0} E(iX)^k \exp(i\theta X)(\frac{\exp(ihX) - 1}{h}) = E(iX)^{k+1} \exp(i\theta X)$.

Rest wie im Beweis zu Satz 11.4 (iii). □

Satz 11.7 Für alle $\theta \in \mathbb{R}$ mit $\lim_{n \to \infty} \frac{|\theta|^n}{n!} E|X|^n = 0$ gilt $\varphi_X(\theta) = \sum_{k=0}^{\infty} \frac{(i\theta)^k}{k!} EX^k$.

Beweis: Lemma 11.5 liefert nach Integration ($y \rightsquigarrow \theta X$): $|\varphi_X(\theta) - \sum_{k=0}^{n} \frac{(i\theta)^k}{k!} EX^k| \le 2 \frac{|\theta|^n}{n!} E|X|^n$. □

Beispiel 11.8

$X \sim N(0,1)$. Alle Momente zu X existieren. Aus Symmetriegründen hat man $EX^{2k+1} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2k+1} \exp(-\frac{x^2}{2}) dx = 0$.

Partielle Intergration liefert $EX^{2k} = (2k-1) EX^{2k-2}$.

$\frac{\theta^{2n}}{(2n)!} EX^{2n} = \theta^{2n} \frac{(2n-1)(2n-3) \cdot \ldots \cdot 3 \cdot 1}{(2n)!} = \frac{\theta^{2n}}{2^n \cdot n!}$

$\varphi_X(\theta) = \sum_{k=0}^{\infty} \frac{(i\theta)^{2k}}{(2k)!} (2k-1) \cdot \ldots \cdot 5 \cdot 3 \cdot 1 = \sum_{k=0}^{\infty} \frac{(-\theta^2)^k}{2^k k!} = \exp(-\frac{\theta^2}{2})$.

$X \sim N(\mu, \sigma^2) \implies \varphi_X(\theta) = \exp(i\mu\theta - \frac{1}{2}\theta^2\sigma^2)$

11.3 Umkehrsätze

Satz 11.9 Es sei X ZV mit charakteristischer Funktion φ. Dann gilt für alle a, b mit $-\infty < a < b < \infty$:

$\frac{1}{2}P(X = a) + P(a < X < b) + \frac{1}{2}P(X = b) = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{\exp(-i\theta a) - \exp(-i\theta b)}{i\theta} \varphi(\theta) d\theta$.

Beweis: $S(y) = \int_0^y \frac{1}{x} \sin(x) dx$, $I(T) = \frac{1}{2\pi} \int_{-T}^{T} \frac{\exp(-i\theta a) - \exp(-i\theta b)}{i\theta} \varphi(\theta) d\theta$

$\Psi : \mathbb{R} \times [-T, T] \to \mathbb{C}$, $\Psi(\theta, x) := \begin{cases} \frac{1}{i\theta}(\exp(-i\theta(a-x)) - \exp(-i\theta(b-x))) & , \quad \theta \ne 0 \\ b - a & , \quad \theta = 0 \end{cases}$

Mit Lemma 11.5 (für $n = 1$) folgt, daß Ψ stetig ist und $|\Psi| \le b - a$ gilt.

Insbesondere: Ψ ist $P^X \otimes l|_{[-T,T]}$-integrierbar.

Also läßt sich Fubini II anwenden: $I(T) = \frac{1}{2\pi} \int_{-T}^{T} \frac{1}{i\theta}(\exp(-i\theta a) - \exp(-i\theta b))(\int \exp(i\theta x) P^X(dx)) d\theta =$

$\frac{1}{2\pi}\int(\int_{-T}^{T}\frac{1}{i\theta}(\exp(-i\theta(a-x))-\exp(-i\theta(b-x)))d\theta)P^X(dx)$

Mit $\Psi_{a,b,T}(x)=2\int_0^T\frac{1}{\theta}\sin((x-a)\theta)d\theta-2\int_0^T\frac{1}{\theta}\sin((x-b)\theta)d\theta$ hat man $I(T)=\frac{1}{2\pi}\int\Psi_{a,b,T}(x)P^X(dx)$,

$\int_0^T\frac{1}{\theta}\sin(c\theta)d\theta=\mathrm{sgn}(c)S(T|c|)$, $\quad\mathrm{sgn}(c)=\begin{cases}1 & , \quad >\\ 0 & , \quad c=0\\ -1 & , \quad <\end{cases}$, also

$\Psi_{a,b,T}(x)=2\mathrm{sgn}(x-a)S(T|x-a|)-2\mathrm{sgn}(x-b)S(T|x-b|)$ (Bekannt: $S(y)\to\frac{\pi}{2}$ mit $y\to\infty$)

$$\Psi_{a,b,T}(x)\overset{T\to\infty}{\longrightarrow}\Psi_{a,b}=\begin{cases}0 & , \quad x<a\\ \pi & , \quad x=a\\ 2\pi & , \quad a<x<b\\ \pi & , \quad x=b\\ 0 & , \quad x>b\end{cases}$$

Da S als Funktion auf $[0,\infty)$ stetig ist und mit $x\to\infty$ gegen einen endlichen Grenzwert strebt, gilt $\sup_{x\geq 0}|S(x)|<\infty$.

Man hat also zu $(\Psi_{a,b,T})_{T>0}$ eine integrierbare Majorante, d.h. es folgt mit Lebesgue (majorisierte Konvergenz)
$\lim_{T\to\infty}I(T)=\frac{1}{2\pi}\int\Psi_{a,b}(x)P^X(dx)=\frac{1}{2\pi}(\pi\cdot P(X=a \text{ oder } X=b)+2\pi\cdot P(a<x<b))$. □

Korollar 11.10

Sind X und Y ZV mit derselben charakteristischen Funktion, so haben X und Y dieselbe Verteilung.

Beweis: Man sieht leicht, daß für jede abzählbare Menge D das System $\mathfrak{I}_D:=\{(a,b]:a\leq b,\ a,b\notin D\}$ ein $\cap$-stabiles Erzeugendensystem von $\mathfrak{B}$ ist. Setzt man $D:=\{a\in\mathbb{R}:P(X=a)>0 \text{ oder } P(Y=a)>0\}$, so folgt mit Satz 11.9, daß P^X und P^Y auf $\mathfrak{I}_D$ übereinstimmen.

Sei $A(X):=\{x:P(X=x)>0\}$ die Menge der Atome von X.

Wegen $\#\{x\in\mathbb{R}:P(X=x)\geq\frac{1}{n}\}\leq n$, $A(X)=\bigcup_{n=1}^{\infty}\{x:P(X=x)\geq\frac{1}{n}\}$ ist $A(X)$ und damit $D=A(X)\cup A(Y)$ abzählbar. □

Satz 11.11

Es sei X eine ZV mit charakteristischer Funktion φ.

Gilt $\int|\varphi(\theta)|d\theta<\infty$, so hat X eine stetige Dichte f, die gegeben wird durch $f(x)=\frac{1}{2\pi}\int\exp(-i\theta x)\varphi(\theta)d\theta\qquad\forall x\in\mathbb{R}$.

Beweis: Lemma 11.5 liefert $\quad(*)\quad|\frac{1}{i\theta}(\exp(-i\theta a)-\exp(-i\theta b))|\leq|b-a|$.

Ist φ_X l-integrierbar, so ist $|b-a|\cdot|\varphi_X|$ eine integrierbare Majorante für den Grenzübergang in Satz 11.9, es gilt also $\frac{1}{2}P(X=a)+P(a<X<b)+\frac{1}{2}P(X=b)=\frac{1}{2\pi}\int\frac{1}{i\theta}(\exp(-i\theta a)-\exp(-i\theta b))\varphi(\theta)d\theta$.

Mit $(*)$ folgt $P(a<X<b)\leq\frac{1}{2\pi}|b-a|\int|\varphi(\theta)|d\theta$, insbesondere $P(X=x)=\lim_{n\to\infty}P(x-\frac{1}{n}<X<x+\frac{1}{n})=0$, d.h. X hat keine Atome. Sei F die Verteilungsfunktion zu X.

Dann: $F(b)-F(a)=\frac{1}{2\pi}\int\frac{1}{i\theta}(\exp(-i\theta a)-\exp(-i\theta b))\varphi_X(\theta)d\theta\qquad\forall a<b$.

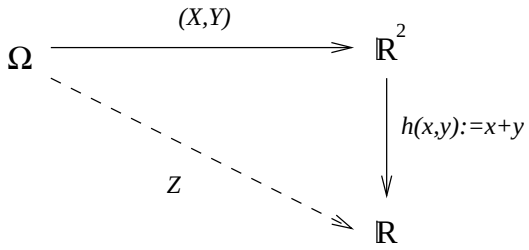
Wegen $(*)$ läßt sich wieder majorisierte Konvergenz anwenden:

$\lim_{h\to 0}\frac{F(x+h)-F(x)}{h}=\frac{1}{2\pi}\int\exp(-i\theta x)\lim_{h\to 0}(\frac{1-\exp(-i\theta h)}{i\theta h})\varphi(\theta)d\theta=\frac{1}{2\pi}\int\exp(-i\theta x)\varphi(\theta)d\theta$. □

11.4 Faltungen

X, Y seien unabhängige ZV auf $(\Omega,\mathfrak{A},P)$, $Z:=X+Y$.

Nach Satz 9.12(i): Der Zufallsvektor $(X,Y):\Omega\to\mathbb{R}^2$ hat die Verteilung $P^X\otimes P^Y$, Z ist Funktion von (X,Y).
P^Z ist das Bildmaß von $P^X\otimes P^Y$ unter h.



Definition 11.12 (Faltung)

Es seien μ_1, μ_2 WMaße auf $(\mathbb{R},\mathfrak{B})$. Dann heißt das Bildmaß von $\mu_1\otimes\mu_2$ unter der Abbildung $(x,y)\mapsto x+y$ die Faltung $\mu_1*\mu_2$ von μ_1 und μ_2.

Satz 11.13

Sind X und Y unabhängige ZV auf $(\Omega,\mathfrak{A},P)$, so gilt $P^{X+Y}=P^X*P^Y$.

Satz 11.14

(i) Für alle $A\in\mathfrak{B}$ gilt $\mu_1*\mu_2(A)=\int\mu_1(A-y)\mu_2(dy)=\int\mu_2(A-x)\mu_1(dx)$, wobei $A-x:=\{y-x:y\in A\}$, etc.

(ii) Sind A_X, A_Y abzählbar mit $P(X\in A_X)=P(Y\in A_Y)=1$, so gilt $P(Z\in A_Z)=1$ mit
$A_Z=\{x+y:x\in A_X,y\in A_Y\}$, A_Z abzählbar, und
$P(Z=z)=\sum_{x\in A_X}P(X=x)P(Y=z-x)=\sum_{y\in A_Y}P(X=z-y)P(Y=y)$.

(iii) X, Y seien unabhängige ZV. Sind f_X, f_Y Dichten von X und Y, so ist auch $Z=X+Y$ absolutstetig verteilt, mit Dichte $f_Z(z)=\int f_X(z-y)f_Y(y)dy=\int f_X(x)f_Y(z-x)dx$.

Beweis: (i) Sei $B:=\{(x,y)\in\mathbb{R}^2:x+y\in A\}\qquad B_y=\{x\in\mathbb{R}:(x,y)\in B\}=\{x\in\mathbb{R}:x+y\in A\}=A-y$

Dann gilt $\mu_1*\mu_2(A)=\mu_1\otimes\mu_2(B)\overset{\text{Satz 9.5}}{=}\int\mu_1(B_y)\mu_2(dy)=\int\mu_1(A-y)\mu_2(dy)$.

(ii) $P(Z=z)=\sum_{x\in A_X} P(Z=z,X=x)=\sum_{x\in A_X} P(Y=z-x,X=x)=\sum_{x\in A_X} P(Y=z-x)P(X=x).$[1]

(Diese Aussage zeigt, daß hier der in der Stochastik I eingeführte Faltungsbegriff ($\rightarrow$ Satz und Definition 4.25) verallgemeinert wird.)

(iii) Wir definieren $f:\mathbb{R}\to\mathbb{R}$ durch $f(z):=\int f_Y(z-x)f_X(x)dx$.

$$\int_{(-\infty,a]} f(z)dz=\int(\int 1_{(-\infty,a]}(z)f_Y(z-x)f_X(x)dx)dz \overset{\text{Fubini}}{=}_{1} \int(\underbrace{\int 1_{(-\infty,a]}(z)f_Y(z-x)dz}_{\text{Substitution } z'=z-x})f_X(x)dx=$$

$$\int(\int 1_{(-\infty,a-x]}(z')f_Y(z')dz')P^X(dx)=\int P(\underbrace{Y\le a-x}_{(-\infty,a]-x})P^X(dx) \overset{\text{Teil (i)}}{=} P^X * P^Y((-\infty,a]) \overset{\text{Satz}}{=}_{11.12} P(Z\le a). \quad \square$$

Beispiel 11.15 X, Y unabhängig, $X\sim N(0,\sigma^2)$, $Y\sim N(0,\tau^2)$. Die Verteilung von $X+Y$ ist gesucht.
Satz 11.14 (iii) liefert:

$f_{X+Y}(z) = \int \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{1}{2\sigma^2}x^2)\frac{1}{\sqrt{2\pi}\tau}\exp(-\frac{1}{2\tau^2}(z-x)^2)dx \overset{\text{quadratische Ergänzung}}{=} \frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}}\frac{1}{\sqrt{2\pi}}\int\exp(-\frac{1}{2}(u-z\frac{\sigma}{\tau\sqrt{\sigma^2+\tau^2}})^2 - \frac{z^2}{2(\sigma^2+\tau^2)})du = \frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}}\exp(-\frac{1}{2(\sigma^2+\tau^2)}z^2)$, also folgt $Z\sim N(0,\sigma^2+\tau^2)$.

Im allgemeinen Fall, $X\sim N(\mu,\sigma^2)$, $Y\sim N(\lambda,\tau^2)$, liefert Anwendung auf $X-\mu$ und $Y-\lambda$:

X, Y unabhängig, $X\sim N(\mu,\sigma^2)$, $Y\sim N(\nu,\tau^2) \implies X+Y\sim N(\mu+\nu,\sigma^2+\tau^2)$.

Satz 11.16 Sind X, Y unabhängige ZV mit charakteristischen Funktionen φ_X und φ_Y, so gilt für die charakteristische Funktion φ_{X+Y} zu $X+Y$: $\varphi_{X+Y}(\theta)=\varphi_X(\theta)\cdot\varphi_Y(\theta) \qquad \forall\theta\in\mathbb{R}$.

Beweis: Sei $\theta\in\mathbb{R}$. Mit X und Y sind auch $\exp(i\theta X)$ und $\exp(i\theta Y)$ unabhängig, also folgt:

$$\varphi_{X+Y}(\theta)=E\exp(i\theta(X+Y))=E(\exp(i\theta X)\exp(i\theta Y)) \overset{\text{Multiplikationsregel}}{=} (E\exp(i\theta X))\cdot(E\exp(i\theta Y))=\varphi_X(\theta)\cdot\varphi_Y(\theta). \quad \square$$

Beispiel 11.17 X, Y unabhängig, $X\sim N(\mu,\sigma^2)$, $Y\sim N(\nu,\tau^2)$.

$\varphi_X(\theta)=\exp(i\mu\theta-\frac{1}{2}\sigma^2\theta^2)$, $\varphi_Y(\theta)=\exp(i\nu\theta-\frac{1}{2}\tau^2\theta^2)$ (nach Beispiel 11.8)

Satz 11.16 liefert $\varphi_{X+Y}(\theta)=\exp(i(\mu+\nu)\theta-\frac{1}{2}(\sigma^2+\tau^2)\theta^2)$.

Dies ist die charakteristische Funktion (nach Beispiel 11.8) zu $N(\mu+\nu,\sigma^2+\tau^2)$, also mit Korollar 11.10

$X+Y\sim N(\mu+\nu,\sigma^2+\tau^2)$ (vergleiche mit Beispiel 11.15.)

Kapitel 12

Verteilungskonvergenz

12.1 Erinnerung, Zusammenhang zu anderen Konvergenzbegriffen

(Definition 6.2) Sind $P, P_1, P_2, \ldots$ WMaße auf $(\mathbb{R}, \mathfrak{B})$ mit Verteilungsfunktion $F, F_1, F_2, \ldots$, so konvergiert P_n schwach gegen P, $P_n \xrightarrow{W} P$, wenn gilt: $\lim_{n\to\infty} F_n(x) = F(x)$ für alle Stetigkeitspunkte x von F.

Sind $X, X_1, X_2, \ldots$ ZV auf (u.U. verschiedenen) WRäumen $(\Omega, \mathfrak{A}, P), (\Omega_1, \mathfrak{A}_1, P_1), \ldots$ so konvergiert X_n gegen X in Verteilung, $X_n \xrightarrow{D} X$, wenn ${P_n}^{X_n} \xrightarrow{W} P^X$, d.h. $\lim_{n\to\infty} P_n(X_n \le x) = P(X \le x)$ für alle x mit $P(X = x) = 0$.

Satz 12.1 Sind $X, X_1, X_2, \ldots$ ZV auf demselben WRaum $(\Omega, \mathfrak{A}, P)$, so gilt:

$$X_n \xrightarrow{\text{f.s.}} X \implies X_n \xrightarrow{D} X, \quad X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X.$$

Beweis: Da "$\xrightarrow{\text{f.s.}} \implies \xrightarrow{P}$" bekannt ist, reicht es, die zweite Aussage zu beweisen.

Es gelte also $X_n \xrightarrow{P} X$, $F, F_1, F_2, \ldots$ seien die zugehörigen Verteilungsfunktionen.

$P(X_n \le x) \le P(X \le x + \delta) + P(|X_n - X| > \delta)$ (denn $X_n \le x \implies X \le x + \delta$ oder $|X_n - X| > \delta$)

$P(X_n \le x) \ge P(X \le x - \delta) - P(|X_n - X| > \delta)$

Läßt man nun zunächst $n \to \infty$ gehen und dann $\delta \downarrow 0$, so folgt

$F(x-) \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le F(x)$, also $F_n(x) \longrightarrow F(x)$ für alle Stetigkeitspunkte x von F (in denen ja $F(x-) = F(x)$ gelten muß.) □

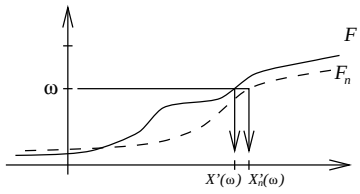
Schwache Konvergenz ist <u>wirklich</u> schwächer als $\xrightarrow{P}$:

$$X \sim N(0,1)\ X_n = \begin{cases} X & , \quad n \text{ gerade} \\ -X & , \quad n \text{ ungerade} \end{cases}$$

Satz 12.2 (Shorohod)

Es seien $X, X_1, X_2, \ldots$ ZV mit $X_n \xrightarrow{D} X$. Dann existieren ein WRaum $(\Omega, \mathfrak{A}, P)$ und hierauf ZV $X', X_1', X_2', \ldots$ mit $X' \overset{D}{=} X$,[1] $X_n' \overset{D}{=} X_n \quad \forall n \in \mathbb{N}$, $X_n' \xrightarrow{\text{f.s.}} X'$.

Beweis: Es seien wieder $F, F_1, F_2, \ldots$ die Verteilungsfunktionen zu $X, X_1, X_2, \ldots$.



Sei $(\Omega, \mathfrak{A}, P) := ([0,1), \mathfrak{B}|_{[0,1)}, l|_{[0,1)})$. $X' := F^{-1}$, $X_n' := {F_n}^{-1}$

(Definition 5.19) $F^{-1}(x) = \inf\{y : F(y) \ge x\}$ "Quantiltransformation"

Dann gilt: $X' \overset{D}{=} X$, $X_n' \overset{D}{=} X_n$ ($\rightsquigarrow$ Beweis zu Satz 5.21)

Es bleibt zu zeigen:

(1) $\lim X_n'(\omega) = X'(\omega)$ für P-fast alle $\omega \in \Omega$.

Sei $\omega \in (0,1)$. Zu gegebenem $\epsilon > 0$ wähle ein $x \in \mathbb{R}$ mit $X'(\omega) - \epsilon < x < X'(\omega)$ und $P(X = x) = 0$.

(X hat nur höchstens abzählbar viele Atome, also läßt sich in $(X'(\omega) - \epsilon, X'(\omega))$ ein solches x finden.)

Man hat $y \le F(x) \iff F^{-1}(y) \le x$ (Lemma 5.20), also $\omega \le F(x) \iff X'(\omega) \le x$.

Da F in x stetig ist, gilt $F_n(x) \longrightarrow F(x)$ mit $n \to \infty$, also existiert ein n_0 derart, daß $F_n(x) \le \omega$ für alle $n \ge n_0$ gilt, denn $F(x) < \omega$. Damit $X_n'(\omega) > x$ für alle $n \ge n_0$, denn: $\omega \le F_n(x) \iff X_n'(\omega) \le x$, also $\liminf_{n\to\infty} X_n'(\omega) \ge x \ge X'(\omega) - \epsilon$.

Hieraus folgt mit $\epsilon \downarrow 0$ $\quad \liminf_{n\to\infty} X_n'(\omega) \ge X'(\omega)$.

Eine ähnliche Argumentation zeigt $\limsup_{n\to\infty} X_n'(\omega) \le X'(\omega') \quad \forall \omega' > \omega$.

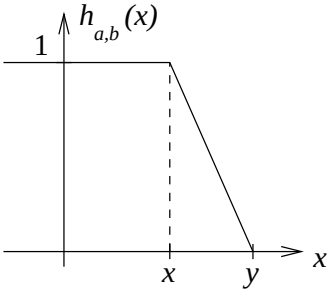
In den Punkten ω, in denen X' von rechts stetig ist, und das sind fast alle, kann man nun $\omega' \downarrow \omega$ ausführen und erhält insgesamt $X'(\omega) \le \liminf X_n'(\omega)$, $\limsup X_n'(\omega) \le X'(\omega)$. □

[1] $\mathfrak{L}(X') = \mathfrak{L}(X)$, X' hat dieselbe Verteilung wie X

Satz 12.3 Es sei $C_b(\mathbb{R})$ die Menge der stetigen und beschränkten Funktionen $h : \mathbb{R} \to \mathbb{R}$.
Dann gilt: $X_n \xrightarrow{D} X \iff Eh(X_n) \to Eh(X) \quad \forall h \in C_b(\mathbb{R})$.

Beweis: "$\Longrightarrow$": Nach Satz 12.2 existieren auf einem geeigneten WRaum $(\Omega, \mathfrak{A}, P)$ ZV $X', X_1', X_2', \ldots$ mit $X' \stackrel{D}{=} X$, $X_n' \stackrel{D}{=} X_n \quad \forall n$ und $X_n' \to X'$ fast sicher.
Da h stetig ist, gilt dann auch $h(X_n') \to h(X')$ f.s..
Da h beschränkt ist, kann der Satz von der majorisierten Konvergenz angewendet werden und liefert
$Eh(X_n) = Eh(X_n') \to Eh(X') = Eh(X)$

"$\Longleftarrow$": Für alle $a, b \in \mathbb{R}$ mit $a < b$ sei $h_{a,b}(x) = \begin{cases} 1 & , \quad x \le a \\ \frac{b-x}{b-a} & , \quad a < x < b \\ 0 & , \quad x \ge b \end{cases}$



Es seien F, F_k $(k \in \mathbb{N})$ die Verteilungsfunktionen zu X, X_k.
Dann gilt für alle $y > x$
$F_n(x) = E1_{(-\infty,x]}(X_n) \le Eh_{x,y}(X_n) \to Eh_{x,y}(X) \le E1_{(-\infty,y]}(X) = F(y)$.
Also: $\limsup_{n\to\infty} F_n(x) \le F(y) \quad \forall y > x$, mit $y \downarrow x$: $\limsup_{n\to\infty} F_n(x) \le F(x+) = F(x)$.
Analog erhält man für $y < x$: $F_n(x) \ge Eh_{y,x}(X_n) \to Eh_{y,x}(X) \ge F(y)$,
also mit $y \uparrow x$: $\liminf_{n\to\infty} F_n(x) \ge F(x-)$.
In den Stetigkeitspunkten gilt: $F(x-) = F(x)$, also $F_n(x) \to F(x)$. □

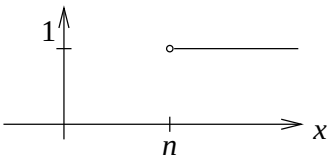
Satz 12.4 ("Continuous Mapping Theorem")
Es seien $X, X_1, X_2, \ldots$ ZV mit $X_n \xrightarrow{D} X$. Weiter sei $f : \mathbb{R} \to \mathbb{R}$ eine Borel-meßbare Abbildung mit der Eigenschaft $P(X \in \{x \in \mathbb{R} : \ f \text{ nicht stetig in } x\}) = 0$. Dann gilt $f(X_n) \xrightarrow{D} f(X)$.
Beweis: Übungsaufgabe. □

12.2 Straffheit und charakteristische Funktionen

Aus der Analysis:

Satz 12.5 (Helly)
Zu jeder Folge $(F_n)_{n\in\mathbb{N}}$ von Verteilungsfunktionen existiert eine Teilfolge $(F_{n_k})_{k\in\mathbb{N}}$ sowie eine schwach monoton steigende, rechtsstetige Funktion $G : \mathbb{R} \to [0,1]$ mit $\lim_{k\to\infty} F_{n_k}(x) = G(x)$ in allen Stetigkeitspunkten x von G.
Grobe Beweisskizze: Für jedes $x \in \mathbb{R}$ hat $(F_n(x))_{n\in\mathbb{N}}$ als beschränkte Folge reeller Zahlen einen Häufungspunkt. Ist $(r_k)_{k\in\mathbb{N}}$ eine Abzählung von $\mathbb{Q}$, so wähle nacheinander Teil-Teil-Folgen $(F_{n_{k,j}})_{j\in\mathbb{N}}$ mit $F_{n_{k,j}}(r_k) \stackrel{j\to\infty}{\longrightarrow} G_0(r_k)$ (Def. von G_0). Betrachte dann die Diagonalfolge $F_{n_{j,j}}(x) \to G_0(x) \quad \forall x \in \mathbb{Q}$. Setze $G(x) := \inf\{G_0(q) : q \in \mathbb{Q},\ q \ge x\}$.
Rest $\epsilon - \delta$.

Beachte: Das G aus Satz 12.5 ist nicht unbedingt eine Verteilungsfunktion.
Im Falle $F_n = 1_{(n,\infty)}$ hat man $G \equiv 0$ (Masse verschwindet im unendlichen.)



Definition 12.6 Eine Familie $\mathfrak{P}$ von WMaßen heißt straff (englisch: tight), wenn zu jedem $\epsilon > 0$ ein kompaktes Intervall $[a,b]$ existiert mit $P([a,b]) \ge 1 - \epsilon \quad \forall P \in \mathfrak{P}$.

Man sieht leicht: $\mathfrak{P}_1$ und $\mathfrak{P}_2$ straff $\Longrightarrow$ $\mathfrak{P}_1 \cup \mathfrak{P}_2$ straff.
$\mathfrak{P}$ straff, $\mathfrak{P}_0 \subset \mathfrak{P}$ $\Longrightarrow$ $\mathfrak{P}_0$ straff.
$\#\mathfrak{P} = 1$ $\Longrightarrow$ $\mathfrak{P}$ straff.

Satz 12.7 Ist $\{P_n : n \in \mathbb{N}\}$ eine straffe Familie von WMaßen auf $(\mathbb{R}, \mathfrak{B})$, so existiert eine Teilfolge $(P_{n_k})_{k\in\mathbb{N}}$ und ein WMaß P mit $P_{n_k} \xrightarrow{W} P$, $k \to \infty$.
Beweis: Sei F_n die Verteilungsfunktion von P_n. Der Satz von Helly liefert $(n_k)_{k\in\mathbb{N}}$ und eine schwach monoton steigende, rechtsstetige Funktion G mit Werten in $[0,1]$ derart, daß $F_{n_k}(x) \to G(x) \quad \forall x$, G stetig in x.
Noch zu zeigen: $\lim_{x\to-\infty} G(x) = 0$, $\lim_{x\to\infty} G(x) = 1$ (Dann ist G Verteilungsfunktion, und das zugehörige P tut's.)
Es sei $\epsilon > 0$. Da $\{P_n : n \in \mathbb{N}\}$ straff ist, existieren a, b mit $P_n([a+1, b]) \ge 1 - \epsilon \quad \forall n \in \mathbb{N}$.
Insbesondere gilt: $F_n(a) \le \epsilon \quad \forall n \in \mathbb{N}$. Da G höchstens abzählbar viele Unstetigkeitsstellen hat
($\rightsquigarrow$ Beweis zu Korollar 11.10), existiert $c < a$ mit G stetig in c.
Es folgt: $G(c) = \lim_{k\to\infty} F_{n_k}(c) \le \epsilon$, also $G(x) \le \epsilon \quad \forall x \le c$.
Damit ist gezeigt: $\forall \epsilon > 0 \ \exists c \in \mathbb{R} \ \forall x \le c : 0 \le G(x) \le \epsilon$.
Das ist die Aussage $\lim_{x\to-\infty} G(x) = 0$ in $\epsilon - \delta$-Schreibweise.
Bew. zu $\lim_{x\to\infty} G(x) = 1$ ganz analog. □

Satz 12.8 (Stetigkeitssatz für charakteristische Funktionen)
Es seien $X, X_1, \ldots$ ZV mit charakteristischen Funktionen $\varphi, \varphi_1, \ldots$. Dann gilt $X_n \xrightarrow{D} X \iff \varphi_n(\theta) \to \varphi(\theta) \quad \forall \theta \in \mathbb{R}$.
Beweis:
"$\Longrightarrow$": Sei $\theta \in \mathbb{R}$. Die Funktionen $x \to \cos\theta x,\ x \to \sin\theta x$ sind stetig und beschränkt, also folgt mit Satz 12.3:
$\varphi_n(\theta) = E\exp(i\theta X_n) = E\cos(\theta X_n) + iE\sin(\theta X_n) \overset{n\to\infty}{\longrightarrow} E\cos(\theta X) + iE\sin(\theta X) = E\exp(i\theta X) = \varphi(\theta)$.
"$\Longleftarrow$": Wir zeigen zunächst: $\{P^{X_n} : n \in \mathbb{N}\}$ ist straff.

Fubini II (für $\mathbb{C}$-wertig Funktionen) liefert für alle $\delta > 0$: $\frac{1}{\delta}\int_{-\delta}^{\delta} \underbrace{1}_{\varphi_n(0)} - \varphi_n(\theta)d\theta = \int(\frac{1}{\delta}\int_{-\delta}^{\delta}(1-\exp(i\theta x))d\theta)P^{X_n}(dx) =$

$$2\int \underbrace{(1 - \frac{\sin(\delta x)}{\delta x})}_{\geq 0} P^{X_n}(dx) \geq 2 \int_{|x|\geq\frac{2}{\delta}} \underbrace{(1 - \frac{1}{|\delta x|})}_{\geq\frac{1}{2} \text{ auf } |x|\geq\frac{2}{\delta}} P^{X_n}(dx) \geq P^{X_n}([-\tfrac{2}{\delta}, \tfrac{2}{\delta}]^c)$$

Sei nun $\epsilon > 0$. Da $\varphi(0) = 1$ und φ stetig in 0 (Satz 11.4), existiert ein $\delta > 0$ mit $|1 - \varphi(\theta)| < \frac{\epsilon}{4}$ für $-\delta \leq \theta \leq \delta$, d.h. $|\frac{1}{\delta}\int_{-\delta}^{\delta}(1 - \varphi(\theta))d\theta| \leq \frac{1}{\delta}2\delta\frac{\epsilon}{4} = \frac{\epsilon}{2}$.

Da alle φ_n's betragsmäßig durch 1 beschränkt sind und über ein endliches Intervall integriert wird, kann majorisierte Konvergenz angewendet werden: $\int_{-\delta}^{\delta}(1 - \varphi_n(\theta))d\theta \to \int_{-\delta}^{\delta}(1 - \varphi(\theta))d\theta$, also existiert ein $n_0 \in \mathbb{N}$ mit $\frac{1}{\delta}\int_{-\delta}^{\delta}(1 - \varphi_n(\theta))d\theta \leq \epsilon \quad \forall n \geq n_0$.

Die Abschätzung liefert also: $P(X_n \in [-\frac{2}{\delta}, \frac{2}{\delta}]) \geq 1 - \epsilon \quad \forall n \geq n_0$.
Zu zeigen: $\forall \epsilon > 0\ \exists a \in \mathbb{R}\ \forall n \in \mathbb{N} : P^{X_n}([-a, a]) \geq 1 - \epsilon$.
Zu jedem der (endlich vielen) $n = 1, \ldots, n_0 - 1$ existiert ein $a_n \in \mathbb{R}$ mit $P^{X_n}([-a_n, a_n]) \geq 1 - \epsilon$, denn $P^{X_n}([-m, m]) \overset{m\to\infty}{\longrightarrow} 1$.
Also folgt mit $a := \max\{a_1, \ldots, a_{n_0-1}, \frac{2}{\delta}\}$: $P^{X_n}([-a, a]) \geq 1 - \epsilon$ für alle $n \in \mathbb{N}$.

Angenommen, $X_n \xrightarrow{D} X$ gilt nicht. Dann existiert ein $x \in \mathbb{R}$ mit $P(X = x) = 0$ und $P(X_n \leq x) \not\to P(X \leq x)$.
Es gibt dann ein $\epsilon > 0$ und eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ mit $(*)$: $|P(X_{n_k} \leq x) - P(X \leq x)| \geq \epsilon$ für alle $k \in \mathbb{N}$.
Da $\{P^{X_{n_k}} : k \in \mathbb{N}\}$ wieder straff ist, existiert nach Satz 12.7 eine Teil-Teilfolge $(X_{n_{k_j}})_{j\in\mathbb{N}}$ und ein WMaß P_0 mit $P^{X_{n_{k_j}}} \xrightarrow{W} P_0$. Aus der Voraussetzung $\varphi_n \to \varphi$ und dem bereits bewiesenen "$\Longrightarrow$"-Teil folgt $\varphi_0 = \varphi$ (wobei φ_0 die charakteristische Funktion zu P_0 bezeichnet.)
Der Eindeutigkeitssatz liefert $P_0 = P^X$, also $X_{n_{k_j}} \xrightarrow{D} X$, und damit $P(X_{n_{k_j}} \leq x) \to P(X \leq x)$.
Dies ist ein Widerspruch zu $(*)$. □

Nocheinmal: Verteilungskonvergenz bezieht sich auf Verteilungen. $Y_n \xrightarrow{D} Y$, $Y \sim N(0,1)$ steht für $P^{Y_n} \xrightarrow{W} N(0,1)$, nach Satz 12.8 und Beispiel 11.8 reicht hierfür der Nachweis von $E\exp(i\theta Y_n) \to \exp(-\frac{1}{2}\theta^2) \quad \forall \theta \in \mathbb{R}$.

Kapitel 13

Der Zentrale Grenzwertsatz (ZGWS)

13.1 Identisch verteilte Summanden

Es wird das aus §6 (Stochastik 1) bekannte Material (kumulative Form der Normalapproximation) verallgemeinert.

Lemma 13.1 Es seien $z_1, \ldots, z_n, w_1, \ldots, w_n \in \{z \in \mathbb{C} : |z| \leq 1\}$. Dann gilt: $|\prod_{k=1}^{n} z_k - \prod_{k=1}^{n} w_k| \leq \sum_{k=1}^{n} |z_k - w_k|$.

Beweis: Ersetzt man in z-Produkt nacheinander die z-Werte durch w-Werte, so folgt mit der Dreiecksungleichung:
$|z_1 \cdot \ldots \cdot z_n - w_1 \cdot \ldots \cdot w_n| \leq |z_1 \cdot \ldots \cdot z_n - w_1 \cdot z_2 \cdot \ldots \cdot z_n| + |w_1 \cdot z_2 \cdot \ldots \cdot z_n - w_1 \cdot w_2 \cdot z_3 \cdot \ldots \cdot z_n| + \ldots +$
$|w_1 \cdot \ldots \cdot w_{n-1} \cdot z_n - w_1 \cdot \ldots \cdot w_n| \leq |z_1 - w_1| + |z_2 - w_2| + \ldots + |z_n - w_n|$. □

Satz 13.2
Es sei $X_1, X_2, \ldots$ eine Folge von unabhängigen, identisch verteilten ZV mit endlicher, positiver Varianz σ^2 und Erwartungswert μ. Dann gilt für $\overline{X_n} := \frac{1}{n}\sum_{i=1}^{n} X_i$ $\sqrt{n}(\overline{X_n} - \mu) \xrightarrow{D} Z$, $Z \sim N(0, \sigma^2)$ (alternativ: $\mathfrak{L}(\sqrt{n}(\overline{X_n} - \mu)) \xrightarrow{W} N(0, \sigma^2)$.)

Beweis: O.B.d.A. $\mu = 0$ (sonst ersetze X_i durch $X_i - \mu$).
Aus $\mu = 0$ und $EX_n^2 = var(X_n) = \sigma^2$ folgt mit Satz 11.6 $\varphi(\theta) = 1 - \frac{1}{2}\theta^2\sigma^2 + \theta^2 r(\theta)$, wobei $\lim_{\theta \to 0} r(\theta) = 0$.

Sei φ_n die charakteristische Funktion zu $\sqrt{n}\overline{X_n}$ $(= \sum_{i=1}^{n} \frac{X_i}{\sqrt{n}})$. Satz 11.4(i), Satz 11.16 $\Longrightarrow$ $\varphi_n(\theta) = (\varphi(\frac{\theta}{\sqrt{n}}))^n$

Lemma 13.1 liefert: für jedes $\theta \in \mathbb{R}$ existiert ein $n_0 \in \mathbb{N}$, so daß $\forall n \geq n_0 :$ $|\varphi_n(\theta) - (1 - \frac{1}{2n}\sigma^2\theta^2)^n| \leq \theta^2 r(\frac{\theta}{\sqrt{n}})$,
also $\lim_{n\to\infty} \varphi_n(\theta) = \lim_{n\to\infty} (1 - \frac{1}{2n}\sigma^2\theta^2)^n = \exp(-\frac{1}{2}\sigma^2\theta^2)$ $\forall \theta \in \mathbb{R}$.

Die rechte Seite ist die charakteristische Funktion zu $N(0, \sigma^2)$, also folgt die Behauptung mit dem Stetigkeitssatz. □

Bemerkung 13.3 In der Literatur wird der ZGWS häufig auch in der Form (vgl. Bemerkung 6.8)
$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{D} Z$, $Z \sim N(0,1)$ angegeben; es werden also die Partialsummen $S_n = \sum_{i=1}^{n} X_n$ anstelle der Mittelwerte betrachtet.
Wegen $\frac{1}{\sqrt{n}\sigma}(S_n - n\mu) = \frac{1}{\sigma}\sqrt{n}(\overline{X_n} - \mu)$ und der Eigenschaft "$Z_n \xrightarrow{D} Z$, $c \in \mathbb{R} \Longrightarrow cZ_n \xrightarrow{D} cZ$"[1] sind beide Formen äquivalent. Diese Grenzwertaussagen führen auf die folgenden, für die Praxis sehr wichtigen Approximationen: $N(\mu, \frac{\sigma^2}{n})$ für $\mathfrak{L}(\overline{X_n})$ und $N(n\mu, n\sigma^2)$ für $\mathfrak{L}(S_n)$.

Zusammenfassung: Ist eine ZV S die Summe einer großen Anzahl unabhängiger, identisch verteilter ZV, so ist S in etwa normalverteilt, oder: Die Verteilung von S kann durch die Normalverteilung mit dem gleichen Erwartungswert und der gleicher Varianz wie S approximiert werden.
In Absatz 13.2 werden wir sehen, daß "identisch verteilt" abgeschwächt werden kann.

Beispiel 13.4 Es seien $X_1, X_2, \ldots$ die zufälligen Lebensdauern (in Stunden) eines bestimmten Typs Glühbirnen. Wir nehmen an, daß diese unabhängig und exponentialverteilt mit Parameter $\frac{1}{100}$ sind (d.h. die mittlere Lebensdauer einer solchen Glühbirne beträgt 100 Stunden.)
Wieviele Glühbirnen sollte man kaufen, wenn man eine Gesamtbrenndauer von 1000 Stunden benötigt? Antwort: unendlich viele.
Modifikation: Wieviele sollte man kaufen, wenn man mit W. 0.99 sicher sein will, daß sie insgesamt mindestens 1000 Stunden brennen?

Es gilt; $EX_i = 100$, $var(X_i) = (EX_i)^2 = 10000$ (dies erhält man mit einer ähnlichen Rechnung wie in Beispiel 8.25)
Gesucht ist ein (möglichst kleines) n mit $P(S_n \geq 1000) \geq 0.99$
Man hat $P(S_n \geq 1000) \geq 0.99 \iff P(\frac{S_n - ES_n}{\sqrt{var(S_n)}} \geq \frac{1000 - n \cdot 100}{\sqrt{n}\sqrt{10000}}) \geq 0.99$.

[1] (Aufgabe 34)

Für eine $N(0,1)$-verteilte ZV Z gilt $P(Z \geq -2.326) \approx 0.99$, der ZGWS führt also auf die Näherung $\frac{10-n}{\sqrt{n}} \leq -2.326$, d.h. $n \geq 20.54$. Man kauft also 21 Glühbirnen.
In diesem Beispiel kann man den exakten Wert ausrechnen und erhält, daß 19 Glühbirnen gereicht hätten.

In Satz 13.2: $X_1, X_2, \ldots$ unabhängig und identisch verteilt, $\mu = EX_i$, $\sigma := var(X_i) < \infty$, $\mathfrak{L}(\frac{\sqrt{n}}{\sigma}(\overline{X_n} - \mu)) \xrightarrow{W} N(0, \sigma^2)$.
Aus der Stochastik I bekannt: Satz von Moivre-Laplace (Ist Spezialfall hiervon mit $P(X_i = 1) = 1 - P(X_i = 0) = p$).

13.2 Der Satz von Lindeberg

Satz 13.5 (Lindeberg)
Für jedes $n \in \mathbb{N}$ seien $X_{n1}, \ldots, X_{nr_n}$ unabhängige ZV auf einem WRaum $(\Omega_n, \mathfrak{A}_n, P_n)$ mit endlicher Varianz σ_{nk}^2 und Erwartungswert μ_{nk}, $1 \leq k \leq r_n$, $s_n^2 := \sum_{k=1}^{r_n} \sigma_{nk}^2 > 0$. Ist dann die Lindeberg-Bedingung

$$(*) \qquad \boxed{\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{r_n} \int\limits_{|X_{nk}-\mu_{nk}|>\epsilon \cdot s_n} (X_{nk} - \mu_{nk})^2 dP_n = 0 \quad \forall \epsilon > 0}$$

erfüllt, so gilt $\frac{1}{s_n} \sum_{k=1}^{r_n} (X_{nk} - \mu_{nk}) \xrightarrow{D} Z$, $Z \sim N(0,1)$.

Bemerkung 13.6
(i) Ist $Y_1, Y_2, \ldots$ eine iid Folge auf $(\Omega, \mathfrak{A}, P)$ mit endlicher Varianz $\sigma^2 > 0$, so genügt die Familie $\{X_{nk} : n \in \mathbb{N}\}$ (mit $r_n = n$), $X_{nk} := Y_k$, $(\Omega_n, \mathfrak{A}_n, P_n) = (\Omega, \mathfrak{A}, P)$ den Bedingungen von Satz 13.5:
$s_n^2 = n\sigma^2$
$\frac{1}{s_n^2} \sum_{k=1}^{r_n} \int\limits_{|X_{nk}-\mu_{nk}|>\epsilon s_n} (X_{nk} - \mu_{nk})^2 dP_n = \frac{1}{n\sigma^2} n \int\limits_{|Y_1-\mu_1|>\epsilon\sqrt{n}\sigma} (Y_1 - \mu_1)^2 dP \longrightarrow 0$ wegen majorisierter Konvergenz,
denn $(Y_1 - \mu_1)^2$ ist wegen $var(Y_1) < \infty$ eine integrierbare Majorante.
(ii) Beim Beweis wird deutlich werden, daß durch die Lindeberg-Bedingung ein dominanter Einfluß einzelner X_{nk}'s auf die Summe $X_{n1} + \ldots + X_{nr_n}$ ausgeschlossen wird, eine Bedingung dieser Art ist auch notwendig.

Beweis: Wir können $\mu_{nk} = 0$, $s_n = 1$ annehmen – ersetze sonst X_{nk} durch $X'_{nk} := \frac{1}{s_n}(X_{nk} - \mu_{nk})$.
Mit $n = 2$ liefert Lemma 11.5 $|\exp(i\theta x) - (1 + i\theta x - \frac{1}{2}\theta^2 x^2)| \leq \min\{|\theta x|^2, |\theta x|^3\}$.
Für die charakteristische Funktion φ_{nk} zu X_{nk} bedeutet dies $|\varphi_{nk}(\theta) - (1 - \frac{1}{2}\theta^2\sigma_{nk}^2)| \leq E\min\{|\theta X_{nk}|^2, |\theta X_{nk}|^3\}$.
Sei $\epsilon > 0$. Dann gilt:
$E\min\{\cdot\} \leq \int\limits_{|X_{nk}|<\epsilon} |\theta X_{nk}|^3 dP_n + \int\limits_{|X_{nk}|\geq\epsilon} |\theta X_{nk}|^2 dP_n \leq \epsilon\theta^3\sigma_{nk}^2 + \theta^2 \int\limits_{|X_{nk}|\geq\epsilon} X_{nk}^2 dP_n$.
Summiere über k: $\sum \sigma_{nk}^2 = 1$, also

$$(1) \quad \lim_{n\to\infty} \sum_{k=1}^{r_n} |\varphi_{nk}(\theta) - (1 - \tfrac{1}{2}\theta^2\sigma_{nk}^2)| \leq \epsilon\theta^3 + \theta^2 \underbrace{\overbrace{\sum_{k=1}^{r_n} \ldots}^{\text{wie in } (*)}}_{\to 0 \text{ mit } n\to\infty} \xrightarrow{\epsilon\to 0} 0.$$

Wegen $s_n = 1$ ist $\prod_{k=1}^{r_n} \varphi_{nk}$ die charakteristische Funktion zu $\frac{1}{s_n} S_n$, $S_n := X_{n1} + \ldots + X_{nr_n}$.
Wir behaupten nun:

$$(2) \quad \lim_{n\to\infty} |\prod_{k=1}^{r_n} \varphi_{nk}(\theta) - \prod_{k=1}^{r_n} (1 - \tfrac{1}{2}\theta^2\sigma_{nk}^2)| = 0 \qquad \forall \theta \in \mathbb{R}.$$

Für alle $\epsilon > 0$ gilt $\sigma_{nk}^2 \leq \epsilon^2 + \int\limits_{|X_{nk}|>\epsilon} X_{nk}^2 dP_n$, also folgt mit der Lindeberg-Bedingung:

$$(0 \leq) \max\{\sigma_{nk}^2 : 1 \leq k \leq r_n\} \leq \epsilon^2 + \underbrace{\sum_{k=1}^{r_n} \int\limits_{|X_{nk}|>\epsilon} X_{nk}^2 dP_n}_{\to 0 \text{ mit } n\to\infty}, \text{ also}$$

$$(3) \quad \lim_{n\to\infty} \max\{\sigma_{nk}^2 : 1 \leq k \leq r_n\} = 0$$

Für alle $\theta \in \mathbb{R}$ existiert also ein n_0 derart, daß für alle $n \geq n_0$
$(4) \quad |1 - \frac{1}{2}\theta^2\sigma_{nk}^2| \leq 1$ für $k = 1, \ldots, r_n$.
Also kann Lemma 13.1 verwendet werden
$|\prod_{k=1}^{r_n} \varphi_{nk}(\theta) - \prod_{k=1}^{r_n} (1 - \frac{1}{2}\sigma_{nk}^2\theta^2)| \leq \sum_{k=1}^{r_n} |\varphi_{nk}(\theta) - (1 - \frac{1}{2}\sigma_{nk}^2\theta^2)|$ für alle $n \geq n_0$, $k = 1, \ldots, r_n$.
Also folgt (2) mit (1).
Es bleibt zu zeigen:

$$(5) \quad \lim_{n\to\infty} |\underbrace{\prod_{k=1}^{r_k} \exp(-\frac{1}{2}\theta^2\sigma_{nk}^2)}_{\exp(-\frac{1}{2}\theta^2) \text{ char. Fkt. zu } N(0,1)} - \prod_{k=1}^{r_n} (1 - \tfrac{1}{2}\theta^2\sigma_{nk}^2)| \stackrel{!}{=} 0 \qquad \forall \theta \in \mathbb{R}$$

Wegen Lemma 13.1 reicht es,

$$(6) \quad \lim_{n\to\infty} \sum_{k=1}^{r_k} |\exp(-\tfrac{1}{2}\theta^2\sigma_{nk}^2) - 1 + \tfrac{1}{2}\theta^2\sigma_{nk}^2| = 0 \quad \text{zu zeigen.}$$

Für reelle x mit $|x| \le \frac{1}{2}$ gilt: $|\exp(x) - 1 - x| \le \frac{1}{2}\sum_{j=2}^{\infty}|x|^j$ (geometrische Reihe)

Dies führt auf $\sum_{k=1}^{r_n}|\exp(-\frac{1}{2}\theta^2\sigma_{nk}^2) - 1 + \frac{1}{2}\theta^2\sigma_{nk}^2| \le \frac{1}{2}\theta^4\sum_{k=1}^{r_k}\sigma_{nk}^4$

Wegen $\sum_{k=1}^{r_k}\sigma_{nk}^4 \le \max\{\sigma_{nk}^2 : 1 \le k \le r_n\}\sum_{k=1}^{r_n}\sigma_{nk}^2$ folgt hieraus (6) mit Hilfe von (3). □

Bemerkung 13.7 Der ZGWS hat eine lange "Verbesserungsgeschichte" hinter sich, die Lindeberg-Bedingung ist in gewisser Weise das Endresultat. Gelegentlich ist die Lyapunov-Bedingung

$$\boxed{\lim_{n\to\infty}\frac{1}{s_n^{2+\delta}}\sum_{k=1}^{r_n}E|X_{nk}-\mu_{nk}|^{2+\delta} = 0}$$

für ein $\delta > 0$ einfacher zu verwenden. Daß diese die Lindeberg-Bedingung impliziert folgt aus

$$\int_{|X_{nk}-\mu_{nk}|>\epsilon s_n}(X_{nk}-\mu_{nk})^2 dP_n \le \int_{|X_{nk}-\mu_{nk}|>\epsilon s_n}\frac{|X_{nk}-\mu_{nk}|^{2+\delta}}{\epsilon^\delta s_n^\delta}dP_n \le \frac{1}{\epsilon^\delta}\frac{1}{s_n^\delta}E|X_{nk}-\mu_{nk}|^{2+\delta}.$$

Zusammenfassung: Die Summe vieler unabhängiger, individuell vernachlässigbarer ZV ist näherungsweise normalverteilt.

13.3 Anwendungen

13.3.1 Ein Sammelproblem

Einer Urne mit n Kugeln werden solange Kugeln zufällig und mit Zurücklegen entnommen, bis insgesamt r_n verschiedene Kugeln gezogen worden sind. S_n sei die Anzahl der hierfür notwendigen Züge.
Dies kann als Modell für die folgende Situation dienen: Jedem Paket Haferflocken liegt eines von n verschiedenen Bildchen bei. Wieviele Pakete muß man kaufen, bis man r_n verschiedene Bilder hat?

($\rightsquigarrow$ G. Polya: Eine Wahrscheinlichkeitsaufgabe zur Kundenwerbung, ZAMM 1930)

Angenommen $r_n = \lceil \rho n\rceil$ mit $0 < \rho < 1$ (d.h. mindestens $100\cdot\rho\%$ der Bilder sollen gesammelt werden).
Bei n Kugeln werden die wiederholten Entnahmen von Kugeln beschrieben mit Hilfe des WRaums $(\Omega_n, \mathfrak{A}_n, P_n) = (\{1,\dots,n\}, \mathbb{P}(\{1,\dots,n\}), unif(\{1,\dots,n\}))^{\mathbb{N}}$ (siehe §9).
Sei X_{nk} die Anzahl der Züge, die nötig sind, um eine neue Kugel zu erhalten, wenn man bereits $k-1$ verschiedene Kugeln entnommen (und wieder zurückgelegt) hat. Dann ist X_{nk} geometrisch verteilt ($\rightsquigarrow$ Absatz 4.2.3) mit Parameter $\frac{n-(k-1)}{n}$ (klar: $X_1 = 1$). Außerdem ist X_{nk} unabhängig von $X_{n1},\dots,X_{n,k-1}$, hieraus folgt, daß $X_{n1},\dots,X_{nr_n}$ ($r_n := \lceil\rho n\rceil$) unabhängig sind.
Nach Konstruktion: $S_n = X_{n1} + X_{n2} + \dots + X_{nr_n}$.
Bekannt: Y geometrisch verteilt mit Parameter $p \implies EY = \frac{1}{p},\ var(Y) = \frac{q}{p^2},\ q := 1-p$.

Auch: $EY^4 = \sum_{k=1}^{\infty}k^4P(Y=k) = p\sum_{k=1}^{\infty}k^4q^{k-1} \le p\sum_{k=1}^{\infty}(k+3)(k+2)(k+1)kq^{k-1} = p\left(\frac{d^4}{(dx)^4}\underbrace{\sum_{k=0}^{\infty}x^k}_{\frac{1}{1-x}}\right)\Big|_{x=q} = \frac{24}{p^4}$.

Für den Erwartungswert und die Varianz von S_n ergibt sich damit

$\mu_n = ES_n = \sum_{k=1}^{r_n}\frac{n}{n-k+1},\ s_n^2 = var(S_n) = \sum_{k=1}^{r_n}\frac{1-\frac{n-k+1}{n}}{(\frac{n-k+1}{n})^2} = n\sum_{k=1}^{r_n}\frac{k-1}{(n-k+1)^2}$.

Als nächstes: Lyapunov-Bedingung mit $\delta = 2$ erfüllt: Mit Y wie oben hat man

$E(Y-\frac{1}{p})^4 \le E(2\max\{Y,\frac{1}{p}\})^4 = 16E\max\{Y^4,(\frac{1}{p})^4\} \le 16E(Y^4 + \frac{1}{p^4}) \le \frac{400}{p^4}$.

Damit $\sum_{k=1}^{r_n}E(X_{nk}-EX_{nk})^4 \le 400\sum_{k=1}^{r_n}\frac{n^4}{(n-k+1)^4} \le 400\lceil\rho n\rceil\frac{1}{(1-\rho)^4} = O(n)$.

Wegen $s_n^2 \ge n\frac{1}{n^2}\sum_{k=1}^{r_n}(k-1) \ge \frac{1}{2}\frac{1}{n}(\rho n-1)(\rho n-2) \sim cn,\ c>0$, also $\lim_{n\to\infty}\frac{1}{s_n^{2+2}}\sum_{k=1}^{r_n}E(X_{nk}-EX_{nk})^{2+2} = 0$.

Also "gilt der ZGWS für S_n": $\frac{S_n - ES_n}{\sqrt{var(S_n)}} \xrightarrow{D} Z,\ Z \sim N(0,1)$.

Wie verhalten sich ES_n und $var(S_n)$ mit $n\to\infty$?

$\frac{1}{n}ES_n = \sum_{k=1}^{\lceil\rho n\rceil}\frac{1}{1-\frac{k-1}{n}}\frac{1}{n}$ Interpretation als Riemannsumme führt auf $\frac{1}{n}E(S_n) \to \int_0^\rho\frac{1}{1-x}dx + O(\frac{1}{n}) = -\log(1-\rho) + O(\frac{1}{n})$

und ähnlich erhält man $\lim_{n\to\infty}\frac{1}{n}var(S_n) = \int_o^\rho\frac{x}{(1-x)^2}dx = \frac{\rho}{1-\rho} + \log(1-\rho)$.

Mit $a(\rho) := -\log(1-\rho),\ b(\rho) := \sqrt{\frac{\rho}{1-\rho} + \log(1-\rho)}$ folgt also $\frac{S_n - a(\rho)n}{b(\rho)\sqrt{n}} \xrightarrow{D} Z,\ Z\sim N(0,1)$.

Numerisches Beispiel: Wie groß muß Ihr Bekanntenkreis sein, damit mit W. ≥ 0.95 an 180 Tagen im Jahr Geburtstag gefeiert werden kann?

$n = 365,\ \rho = \frac{180}{365}$, bekannt: $P(Z \ge 1.645) \approx 0.05$, wenn $Z\sim N(0,1)$ und $S_n \ge k \iff \frac{S_n - a(\rho)n}{b(\rho)\sqrt{n}} \ge \underbrace{\frac{k-a(\rho)n}{b(\rho)\sqrt{n}}}_{\approx 1.645}$.
Wir erhalten den Mittelwert $k = 266$.

Das Verhalten von $a(\rho)$ und $b(\rho)$ mit $\rho\uparrow 1$ läßt vermuten, daß bei $\rho = 1$ (jede Kugel muß mindestens einmal erscheinen)

ein anderes Verhalten vorliegt.

$var(S_n) = n\sum_{k=1}^{n}\frac{k-1}{(n-k+1)^2} = n\sum_{k=1}^{n}\frac{n-k}{k^2} = n^2\sum_{k=1}^{n}\frac{1}{k^2} - n\sum_{k=1}^{n}\frac{1}{k} = n^2\frac{\pi^2}{6} + o(n^2)$.

$var(X_{nn}) = \frac{1-\frac{1}{n}}{(n-k+1)^2} = n^2 + o(n^2)$, d.h. die einzelnen Summanden sind nicht mehr vernachlässigbar ($\max_{1\le k\le r_n}\sigma_{nk}^2 \neq o(s_n^2)$), die Lindeberg-Bedingung ist nicht erfüllt.

$(*)$ $\quad\frac{S_n - n\log n}{n} \xrightarrow{D} Z,\ P(Z\le y) = \exp(-e^{-y})$ (Gumbel, Aufgabe 29)

Numerisches Beispiel: Soll der Bekanntenkreis so groß sein, daß mit W. 0.95 an jedem der 365 Tage im Jahr Geburtstag gefeiert werde kann, so erhält man die Bedingung: $k \ge 365\log 365 + 365\cdot 2.97 \approx 3237.51$

Heuristische Argumentation für die Gültigkeit von $(*)$:
Es sei A_{ik} das Ereignis, das die i-te Kugel nach k Zügen noch nicht aufgetaucht ist. Dann gilt:
$P(A_{ik}) = (1-\frac{1}{n})^k$ für $i = 1,\dots,n$.
Wir unterstellen nun, daß die Ereignisse $A_{1k},\dots,A_{nk}$ zumindest näherungsweise unabhängig sind.
Dann $P(S_n \le k_n) = P(\bigcap_{i=1}^{n} A_{ik_n}^c) \approx (1-(1-\frac{1}{n})^{k_n})^n$.
Für $k_n = n\log n + x\cdot n$ erhält man $P(\frac{S_n - n\log n}{n} \le x) = P(S_n \le k_n) \approx (1 - \exp(\overbrace{k_n}^{-\log n - x}\underbrace{\log(1-\frac{1}{n})}_{\approx -\frac{1}{n}}))^n \approx (1-\frac{1}{n}e^{-x})^n \to \exp(-e^{-x})$.

13.3.2 Rekorde

Es sei $X_1, X_2, \dots$ eine iid-Folge auf $(\Omega,\mathfrak{A},P)$ mit stetiger Verteilungsfunktion F.
Wir setzen $R_n = \begin{cases} 1 & , \quad X_n > X_1, X_n > X_2, \dots, X_n > X_{n-1} \\ 0 & , \quad \text{sonst}\end{cases}$. R_n zeigt an, ob die n-te Versuchswiederholung einen Rekord liefert oder nicht. Mit Fubini folgt für $i \neq j$ $P(X_i = X_j) = P^{X_i}\otimes P^{X_j}(\{(x,x) : x\in\mathbb{R}\}) = \int\underbrace{\int 1_{\{x\}}(y)P^{X_i}(dy)}_{F(x)-F(x-)=0}P^{X_j}dx = 0$, also ist $A := \{\omega\in\Omega : \exists i,j,\ i\neq j,\ X_i(\omega) = X_j(\omega)\} = \bigcup_{i=2}^{\infty}\bigcup_{j=1}^{i-1}\{X_i = X_j\}$ eine P-Nullmenge. Die Wahrscheinlichkeit für das Auftreten "mehrfacher Werte" (engl.: ties) ist 0, kann also vernachlässigt werden.

Sei $\mathbb{P}_n$ die Menge der Permutationen von $\{1,\dots,n\}$. Sei $\Psi_n : \Omega\to\mathbb{P}_n$ die zufällige Permutation, die $X_1,\dots,X_n$ in aufsteigende Reihenfolge bringt:
$\Psi_n = \Pi \iff X_{\Pi(1)} < X_{\Pi(2)} < \dots < X_{\Pi(n)}$
$\Psi_n^{-1}(\{\Pi\})$ ist der Durchschnitt von Mengen $\{X_i < X_j\}$, also in $\mathfrak{A}$, d.h. Ψ_n ist eine Zufallsgröße (meßbare Abbildung), wenn man $\mathbb{P}_n$ mit der Potenzmenge als σ-Algebra ausstattet.
Für jedes feste $\Pi\in\mathbb{P}_n$ haben die Vektoren $(X_{\Pi(1)},\dots,X_{\Pi(n)})$ und $(X_1,\dots,X_n)$ dieselbe Verteilung, nämlich $(P^{X_1})^{\otimes n}$. Sei $B := \{(x_1,\dots,x_n) : x_1 < x_2 < \dots < x_n\}$. Dann gilt $P(\Psi_n = \Pi) = P((X_{\Pi(1)},\dots,X_{\Pi(n)})\in B) = P((X_1,\dots,X_n)\in B) = P(\Psi_n = id)$, also:
(1) $\quad P(\Psi_n = \Pi) = \frac{1}{n!}$ für alle $\Pi\in\mathbb{P}_n$.
Damit (2) $\quad P(R_n = 1) = P(\Psi_n \in \{\Pi\in\mathbb{P}_n : \Pi(n) = n\}) = \frac{(n-1)!}{n!} = \frac{1}{n}$.

$\{R_{n+1} = 1\}\cap\{\Psi_n = \Pi\} = \{\Psi_{n+1} = \tilde\Psi\}$, $\quad\tilde\Psi(i) = \begin{cases}\Pi(i) & , \quad 1\le i\le n\\ n+1 & , \quad i = n+1\end{cases}$

Es folgt mit (1) und (2) $P(\Psi_n = \Pi, R_{n+1} = 1) = P(\Psi_n = \Pi)P(R_{n+1} = 1)\quad\forall\Pi\in\mathbb{P}_n$, d.h. $\sigma(\Psi_n)$ und $\sigma(R_{n+1})$, also die erzeugten σ-Algebren, sind unabhängig. Wegen $\sigma(R_1,\dots,R_n)\subset\sigma(\Psi_n)$ sind dann auch $\sigma(R_1,\dots,R_n)$ und $\sigma(R_{n+1})$ unabhängig. Für beliebige $i_1 < i_2 < \dots < i_n$ bedeutet dies:
$P(R_{i_1} = j_1,\dots,R_{i_n} = j_n) = P(R_{i_1} = j_1,\dots,R_{i_{n-1}} = j_{n-1})P(R_{i_n} = j_n) = \dots = P(R_{i_1} = j_1)\cdot\dots\cdot P(R_{i_n} = j_n)$, d.h. die ZV $R_1, R_2, \dots$ sind unabhängig.

Wieviele Rekorde gibt es unter den ersten n Versuchen? Sei $S_n := \sum_{i=1}^{n} R_i$.

$ES_n = \sum_{i=1}^{n} ER_i = \sum_{i=1}^{n} P(R_i = 1) \overset{(2)}{=} \sum_{i=1}^{n}\frac{1}{i} = \log n + \gamma + o(1)$ (γ ist Eulersche Konstante)

$var(S_n) \overset{\text{unab-}}{\underset{\text{hängig}}{=}} \sum_{i=1}^{n} var(R_i) = \sum_{i=1}^{n}\frac{1}{i}(1-\frac{1}{i}) = \sum_{i=1}^{n}\frac{1}{i} - \sum_{i=1}^{n}\frac{1}{i^2} = \log n + \gamma + o(1) - \frac{\pi^2}{6} + o(1)$

Chebychev liefert $P(|\frac{S_n}{ES_n} - 1|\ge\epsilon) \le \frac{1}{\epsilon^2}\frac{var(S_n)}{(ES_n)^2} \overset{n\to\infty}{\longrightarrow} 0$, d.h. $\frac{S_n}{ES_n}\xrightarrow{P} 1$.
Der ZGWS erlaubt eine genauere Aussage:

Mit $s_n := (var(S_n))^{\frac{1}{2}} \sim (\log n)^{\frac{1}{2}}$ hat man $\frac{1}{s_n^3}\sum_{i=1}^{n} E|R_i|^3 = \frac{1}{s_n^3}\sum_{i=1}^{n} ER_i \overset{n\to\infty}{\longrightarrow} 0$, d.h. Lyapunov mit $\delta = 1$ ist erfüllt, also $\frac{S_n - ES_n}{\sqrt{var(S_n)}}\xrightarrow{D} Z,\ Z\sim N(0,1)$.

Verwendet man die Eigenschaften der Verteilungskonvergenz ($\rightsquigarrow$ Aufgabe 34), so folgt hieraus $\frac{S_n - \log n}{\sqrt{\log n}}\xrightarrow{D} Z$, $Z\sim N(0,1)$.

Für große n ist also beispielsweise die W. dafür, daß die Anzahl der in den ersten n Durchgängen erzielten Rekorde zwischen $\log n - 1.96\sqrt{\log n}$ und $\log n + 1.96\sqrt{\log n}$ liegt, ungefähr 0.95.

13.3.3 Primteiler

$\omega(n)$: Anzahl der Primfaktoren von n (ohne Vielfachheiten), $\omega : \mathbb{N} \to \mathbb{N}$.

Klar: $\omega(p) = 1 \quad \forall p \in \mathbb{P}$ =Menge der Primzahlen= $\{p_1, p_2, \dots\},\ 2 =: p_1 < p_2 < \dots :\ \omega(\prod_{i=1}^{k} p_i) = k$

Läßt sich etwas über das "mittlere Verhalten" von ω aussagen?
Für jedes $n \in \mathbb{N}$ ist $(\Omega_n, \mathfrak{A}_n, \mathfrak{P}_n) := (\{1, \dots, n\}, \mathfrak{P}(\{1, \dots, n\}), unif\{1, \dots, n\})$ ein WRaum.

Sei $r_n := \max\{k : p_k < n\}$ die Anzahl der Primzahlen $\le n$, $X_{nk}(m) : \Omega_n \to \mathbb{R},\ X_{nk} = \begin{cases} 1 & , \quad p_k | m \\ 0 & , \quad \text{sonst} \end{cases} \qquad (1 \le k \le r)$.

Mit diesen ZV gilt $\omega(m) = \sum_{k=1}^{r_n} X_{nk}(m),\ m = 1, \dots, n.$

Fundamentaler Sachverhalt: $p_{i_1}, \dots, p_{i_j}$ teilen $m \iff (\prod_{l=1}^{j} p_{i_l}) | m$.

$P(X_{i_1} = 1, \dots, X_{i_j} = 1) = \frac{1}{n} \left\lfloor \frac{n}{\prod_{l=1}^{j} p_{i_l}} \right\rfloor,\ P_n(X_i = 1) = \frac{1}{n} \lfloor \frac{n}{p_i} \rfloor$.

Für große n gilt also näherungsweise $P_n(X_{i_1} = 1, \dots, X_{i_j} = 1) \approx P(X_{i_1} = 1) \cdot \dots \cdot P(X_{i_j} = 1)$, d.h. Unabhängigkeit.

$E \sum_{k=1}^{r_n} X_{nk} \approx \sum_{p<n} \frac{1}{p} \sim \log\log n,\ var(\sum_{k=1}^{r_n} X_{nk}) \approx \sum_{p \le n} \frac{1}{p}(1 - \frac{1}{p}) \sim \log\log n$ (aus der Zahlentheorie), also:

Satz 13.8 (Erdös–Kac)

$\frac{1}{n} \#\{1 \le m \le n : \frac{\omega(m) - \log\log n}{\sqrt{\log\log n}} \le x\} \overset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{y^2}{2}) dy$

Weitere Details, Beweis, ... $\rightsquigarrow$ Billingsley, §30.

Numerisches Beispiel: $\prod_{i=1}^{1000} p_i \approx 0.6786 \cdot 10^{3393},\ \log\log(\cdot) \approx 8.96$

$n = 10^{1.000.000}$: In etwa 99.9% der Zahlen $\{1, \dots, n\}$ haben nicht mehr als 27 Primfaktoren.

13.3.4 Maximum-Likelihood-Schätzer

Statistisches Problem: Gegeben sind Daten $x_1, \dots, x_n$ als Realisierungen von unabhängigen ZV $X_1, \dots, X_n$ mit derselben Verteilung, die die Lebesgue-Dichte $f(\cdot, \theta)$ hat. Hierbei ist θ ein unbekanntes Element des Parameterraumes $\Theta \subset \mathbb{R}$ (Stichprobe vom Umfang n). Der unbekannte Parameter soll geschätzt werden.

In Analogie zu §7.1 nennen wir $l : \Theta \to \mathbb{R},\ l(\theta)\ (= l(\theta | x_1, \dots, x_n)) := \prod_{i=1}^{n} f(x_i | \theta)$ die Likeliehood-Funktion,

$\log l$ die log-Likelihood-Funktion,

und eine Abbildung $\hat{\theta}_n : \mathbb{R}^n \to \Theta$ mit der Eigenschaft $l(\hat{\theta}(x_1, \dots, x_n) | x_1, \dots, x_n) = \sup_{\Theta} l(\theta(x_1, \dots, x_n))$ einen Maximum-Likelihood-Schätzer (ML-Schätzer).

Beispiel 13.9

(i) $\Theta = (0, \infty),\ f(x|\theta) = \theta e^{-\theta x} 1_{(0,\infty)}(x)$ (Exponentialverteilungen)

Man erhält $\log l(\theta) = n \log\theta - \theta \sum_{i=1}^{n} x_i$ und durch Ableiten und Nullsetzen etc.

$\hat{\theta}_n\ (= \hat{\theta}_n(x_1, \dots, x_n) = \frac{n}{\sum_{i=1}^{n} x_i}) = \frac{1}{\bar{X}_n}$.

(ii) Die Verteilung mit Verteilungsfunktion $F(x|\theta) = \frac{1}{1+\exp(-(x-\theta))},\ -\infty < x < \infty,\ \theta \in \Theta = \mathbb{R}$ bezeichnet man als logistische Verteilung mit Lageparameter θ (und Skalenparameter 1).

Eine zugehörige Dichte ist $f(x|\theta) = \frac{\exp(-(x-\theta))}{(1+\exp(-(x-\theta)))^2}$.

Man erhält $\sum_{i=1}^{n} \log f(x_i|\theta) = -\sum_{i=1}^{n}(x_i - \theta) - 2\sum_{i=1}^{n} \log(1 + \exp(-(x_i - \theta)))$, also $\frac{\partial}{\partial\theta} \sum_{i=1}^{n} \log f(x_i|\theta) = n - \sum_{i=1}^{n} \frac{2}{1+\exp(x_i - \theta)}$.

Rechts steht eine in θ streng monoton fallende Funktion, die mit $\theta \to -\infty$ gegen n und mit $\theta \to \infty$ gegen $-n$ strebt, also genau eine Nullstelle hat: Diese ist der ML-Schätzer.

In manchen Fällen hat man einfache, explizite Formeln für ML-Schätzer und kann diese benutzen, um das asymptotische Verhalten von $\hat{\theta}_n$ herzuleiten (Aufgabe 40 für Beispiel 13.9(i)).

Skizze eines Argumentes für asymptotische Normalität in allgemeinen, aber regulären Fällen:

Sei $s_n(\theta) := \frac{\partial}{\partial\theta} \log l(\theta | X_1, \dots, X_n) = \sum_{i=1}^{n} Y_i,\ Y_i := \frac{\partial}{\partial\theta} \log f(X_i|\theta)$. Die Y_i sind iid.

Mittelwertsatz: $s_n(\theta) = \underbrace{s_n(\hat{\theta}_n)}_{=0,\ \text{da ML-Schätzer}} + (\theta - \hat{\theta}_n) s_n'(\tilde{\theta}_n),\ \tilde{\theta}_n$ zwischen θ und $\hat{\theta}_n$.

Also: $(*)\quad \sqrt{n}(\hat\theta_n-\theta)=\frac{-\frac{1}{\sqrt{n}}s_n(\theta)}{\frac{1}{n}s_n'(\tilde\theta_n)}$, $\quad E_\theta Y_i \overset{Eg(x)=\int g(x)f(x)dx}{=} \int(\frac{\partial}{\partial\theta}f(x|\theta))\frac{1}{f(x|\theta)}f(x|\theta)dx \underset{\text{Regularität}}{\overset{\text{unter}}{=}} \frac{\partial}{\partial\theta}\int f(x|\theta)dx=0.$

Mit dem ZGWS folgt nun $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_i-EY_i)\xrightarrow{D} Z,\ Z\sim N(0,I(\theta))$ mit $I(\theta):=var_\theta Y_1=E_\theta Y_1^2=E_\theta(\frac{\partial}{\partial\theta}\log f(X_1|\theta))^2$
($\rightsquigarrow$ Fisher-Information, vgl. §7.2)
Was passiert im Nenner von $(*)$?
Ist $\hat\theta_n$ konsistent (d.h. $\hat\theta_n\to\theta$ in W.), so muß auch $\tilde\theta_n\to\theta$ in W. gelten. Mit einer gleichmäßigen Version der Gesetzes der großen Zahlen erhält man $\frac{1}{n}s_n'(\tilde\theta_n)=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{(\partial\theta)^2}\log f(X_1|\theta)\big|_{\theta=\tilde\theta_n}\xrightarrow{P} E_\theta\frac{\partial^2}{(\partial\theta)^2}\log f(X_1|\theta)$.

$$E_\theta\tfrac{\partial^2}{(\partial\theta)^2}\log f(X_1|\theta)=\int\tfrac{\partial}{\partial\theta}\Big(\tfrac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)}\Big)f(x|\theta)dx=\int\Big(\tfrac{\frac{\partial^2}{(\partial\theta)^2}f(x|\theta)}{f(x|\theta)}-\tfrac{(\frac{\partial}{\partial\theta}f(x|\theta))^2}{f(x|\theta)^2}\Big)f(x|\theta)dx$$
$$=\underbrace{\int\frac{\partial^2}{(\partial\theta)^2}f(x|\theta)\,dx}_{=0}-\int\Big(\tfrac{\partial}{\partial\theta}\log f(x|\theta)\Big)^2f(x|\theta)dx=-E_\theta\Big(\tfrac{\partial}{\partial\theta}\log f(X_1|\theta)\Big)^2=-I(\theta).$$

Insgesamt folgt aus $(*)$ mit den Eigenschaften der Verteilungskonvergenz (Aufgaben 32 & 34)

$$\boxed{\sqrt{n}(\hat\theta_n-\theta)\xrightarrow{D} Z,\ Z\sim N(0,\tfrac{1}{I(\theta)})}$$

Unter bestimmten Bedingungen ist $\frac{1}{I(\theta)}$ die kleinst mögliche asymptotische Varianz (vergleiche die Informationsgleichung aus Satz 7.8). ML-Schätzer sind dann asymptotisch optimal.

Beispiel 13.10 (Fortsetzung von Beispiel 13.9(ii))
Im Falle $f(x|\theta)=\frac{\exp(-(x-\theta))}{(1+\exp(-(x-\theta)))^2}$ erhält man $\frac{\partial^2}{(\partial\theta)^2}\log f(x|\theta)=\frac{-2\exp(-(x-\theta))}{(1+\exp(-(x-\theta)))^2}$ und damit

$$I(\theta)=-\int\Big(\tfrac{\partial^2}{(\partial\theta)^2}\log f(x|\theta)\Big)f(x|\theta)dx=\int_{-\infty}^{\infty}\tfrac{2\exp(-(x-\theta))^2}{(1+\exp(-(x-\theta)))^4}dx=\tfrac{1}{3}.$$

Obige Argumentation führt also auf

(1) $\quad\sqrt{n}(\hat\theta_n-\theta)\xrightarrow{D} Z,\ Z\sim N(0,3).$

$E_\theta(X_k-\theta)=\int(x-\theta)\frac{\exp(-(x-\theta))}{(1+\exp(-(x-\theta)))^2}dx=0$ (Symmetrie)
Daher ist der Mittelwert $\overline{X_n}$ ein erwartungstreuer Schätzer für θ, dieser Schätzer ist auch konsistent (nach dem schwachen Gesetz der großen Zahlen).
$\int_{-\infty}^{\infty}x^2\frac{\exp(-x)}{(1+\exp(-x))^2}dx=\frac{\pi^2}{3}$, also $var_\theta(X_1)=\frac{\pi^2}{3}$: Der ZGWS (in der Form von Satz 13.2) liefert:

(2) $\quad\sqrt{n}(\overline{X_n}-\theta)\xrightarrow{D} Z,\ Z\sim N(0,\frac{\pi^2}{3}).$

Vergleicht man (1) und (2), so sieht man, daß der ML-Schätzer eine etwas kleinere asymptotische Varianz hat als $\overline{X_n}$.

Aussagen zur asymptotischen Normalität eines Schätzers lassen sich zur Konstruktion von asymptotischen Konfidenzintervallen verwenden (vergleiche §7.5). Hat man $\sqrt{n}(\hat\theta_n-\theta)\xrightarrow{D} Z,\ Z\sim N(0,\sigma^2)$, so ist $[\hat\theta_n-\frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}},\hat\theta_n+\frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}]$ ein asymptotisch korrektes $100(1-\alpha)\%$ Konfidenzintervall für θ (hierbei: $P(Z\le u_\alpha)=\alpha$ bei $Z\sim N(0,1)$).
Kleinere asymptotische Varianz bedeutet: kürzere Intervalle.
Will man in der Situation von Beispiel 13.9 den schlechteren Wert von $\overline{X_n}$ durch größeres n ausgleichen, so wird man auf $\frac{n(\overline{X_n})}{n(\hat\theta_n)}=\frac{\frac{\pi^2}{3}}{3}\approx 1.0966$ geführt, d.h. man braucht etwas 10% mehr Beobachtungen.

Kapitel 14

Zufallsvektoren

14.1 Allgemeines

Hier: Zufallsgrößen X mit Werten in $(\mathbb{R}^d, \mathfrak{B}^d)$.
Elemente von $\mathbb{R}^d$ sind Spaltenvektoren.

$< \cdot, \cdot >$ ist das Skalarprodukt: $< x, y >= \sum_{k=1}^{d} x_k y_k,\ x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix},\ y = \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix}$.

Ist A^t die Transponierte einer $n \times m$-Matrix $A \in \mathbb{R}^{n \times m}$, und faßt man Elemente des $\mathbb{R}^d$ als $d \times 1$-Matrizen auf, so gilt $< x, y >= x^t y$.
Viele der Resultate der früheren Abschnitte lassen sich von $\mathbb{R}$ auf $\mathbb{R}^d,\ d > 1$, verallgemeinern (i.a. hier ohne Beweis.)
Ist X ein d-dimensinaler Zufallsvektor auf $(\Omega, \mathfrak{A}, P)$, so sind die Komponentenabbildungen $X_1, \ldots, X_d$ "gewöhnliche" Zufallsvariablen (auch die Umkehrung hiervon gilt, siehe Aufgabe 15.)
Existieren zu diesen die Erwartungswerte, so nennen wir $EX := \begin{pmatrix} EX_1 \\ \vdots \\ EX_d \end{pmatrix}$ der Erwartungswertvektor zu X (die Linearität bleibt erhalten.)
Hat man $EX_i^2 < \infty$ für $i = 1, \ldots, d$ (und damit auch $E|X_i X_j| < \infty$ für alle $i \neq j$ ($\rightsquigarrow$ Cauchy-Schwarz-Ungleichung), so nennt man

$$\boxed{cov(X) := \big(cov(X_i, X_j)\big)_{i,j=1}^{d}}$$

die Kovarianzmatrix zu X (Varianzmatrix wäre eigentlich besser.)
Auf der Diagonalen stehen die Varianzen der Komponenten.
Viele Rechenregeln lassen sich leicht übertragen: Ist z.B. $A : \mathbb{R}^d \to \mathbb{R}^k$ eine lineare Abbildung (also $k \times d$-Matrix), so gilt $E(AX) = AEX$ (Reihenfolge beachten, nicht $(EX)A$!!)
Dehnt man der Erwartungswertbegriff auf zufällige Matrizen aus (wieder komponentenweise), so gilt:

$$cov(X) = E(X - EX)(X - EX)^t.$$

Damit $cov(AX + b) = E((AX + b) - (AEX + b))(AX + b - (AEX + b))^t) =$
$= E((A(X - EX))(A(X - EX))^t) = AE((X - EX)(X - EX)^t)A^t = A \cdot cov(X) \cdot A^t.$

Ist X ein d-dimensionaler Zufallsvektor auf $(\Omega, \mathfrak{A}, P)$, so wird durch $\varphi_X : \mathbb{R}^d \to \mathbb{C},\ \varphi_X(\theta) = E \exp(i < \theta, X >) \quad \forall \theta \in \mathbb{R}^d$ die charakteristische Funktion zu X definiert. Auch hier gilt $\varphi_X = \varphi_Y \implies \mathfrak{L}(X) = \mathfrak{L}(Y)$ (Eindeutigkeitssatz).

Gibt es "interessante" mehrdimensionale Verteilungen?

14.2 Mehrdimensionale Normalverteilungen

Ein d-dimensionaler Zufallsvektor X heißt (d-dimensional) normalverteilt, wenn alle Linearkombinationen von Komponenten von X, also alle Zufallsvariablen der Form $< a, X >,\ a \in \mathbb{R}^d$, im 1-dimensionalen Sinn normalverteilt sind. (Im Fall $Z \equiv c \in \mathbb{R}$ betrachten wir Z als $N(c, 0)$-verteilt.)
Man hat $X_i = (0, \ldots, 0, 1, 0, \ldots, 0)X$, also existiert der Erwartungswert μ und die Kovarianzmatrix Σ zu X nach den Rechenregeln aus §14.1.:

$Ea^t X = a^t \mu \qquad \forall a \in \mathbb{R}^d,\ var(a^t X) = a^t \Sigma a.$

Damit $\varphi_{a^t X}(\theta) = \exp(i\theta(a^t \mu) - \frac{1}{2}\theta^2 (a^t \Sigma a)) \quad \forall \theta \in \mathbb{R} \qquad (\rightsquigarrow$ Beispiel 11.8)
Dies liefert $\varphi_X(\theta) = E \exp(i < \theta, X >) = E \exp(i \cdot 1 \cdot \theta^t X) = \varphi_{\theta^t X}(1) = \exp(i < \theta, \mu > -\frac{1}{2}\theta^t \Sigma \theta)$.

Erste Konsequenz: Auch im mehrdimensionalen Fall ist eine Normalverteilung durch die beiden Parameter μ und Σ eindeutig festgelegt. Weiter folgt hieraus (oder direkt aus der Definition), wie sich Normalverteilungen unter affinen

Transformationen verhalten:
$\varphi_{AX+b}(\theta) = E\exp(i<\theta, AX+b>)$
$= \exp(i<\theta,b>)E\exp(i<A^t\theta, X>)$
$= \exp(i<\theta,b>)\varphi_X(A^t\theta) = \exp(i<\theta, b+A\mu> -\frac{1}{2}\theta^t(A\Sigma A^t)\theta)$
Also: $X \sim N_d(\mu,\Sigma),\ A\in\mathbb{R}^{k\times d},\ b\in\mathbb{R}^k \implies Y := AX+b \sim N_k(A\mu+b, A\Sigma A^t)$.

Zu welchen $\mu\in\mathbb{R}^d,\ \Sigma\in\mathbb{R}^{d\times d}$ existiert eine solche Verteilung?
Wegen $a^t\Sigma a = var(a^tX) \geq 0 \quad \forall a\in\mathbb{R}^d$ muß Σ positiv semidefinit sein. Aus der Definition folgt, daß Σ symmetrisch sein muß. Zu jeder symmetrischen, positiv semidefiniten Matrix existiert eine "Wurzel", d.h. eine Matrix $A\in\mathbb{R}^{d\times d}$ mit $AA^t=\Sigma$.
Es seien nun $X_1,\dots,X_d$ unabhängige, $N(0,1)$-verteilte Zufallsvariablen. Für $X=(X_1,\dots,X_d)^t$ gilt

$$a^tX = \sum_{i=1}^d a_iX_i \sim N(0,\sum_{i=1}^d a_i^2) = N(0,a^tI_da).$$

Die Definition normalverteilter Zufallsvektoren zeigt also, daß X d-dimensional normalverteilt ist.
Klar: $X\sim N_d(0,I_d)$.
Für $Y := AX+\mu$ erhält man $Y\sim N_d(A\cdot 0+\mu, AI_dA^t) = N_d(\mu,\Sigma)$.
Bei nicht-singulärem Σ ist auch A nicht-singulär und mit Hilfe einer geeigneten Formel für das Verhalten von Dichten unter affinen Transformationen folgt, daß dann $N_d(\mu,\Sigma)$ die (l^d-) Dichte

$$(*)\qquad f(x|\mu,\Sigma) = (2\pi)^{-\frac{d}{2}}(det\Sigma)^{-\frac{1}{2}}\exp(-\tfrac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu))$$

hat.

Bei $det\Sigma=0$ gibt es ein $a\in\mathbb{R}^d$ mit $a^t\Sigma a=0$, also $var(a^tX)=0$, d.h. X ist auf $H:=\{x\in\mathbb{R}^d : a^tx=a^t\mu\}$ konzentriert. $l^d(H)=0$, d.h. es existiert keine Dichte.

Konstruktion zu $N(\mu,\Sigma)$ begann mit $X_1,\dots,X_d,\ X_i\sim N(0,1)$. Dann gilt $X=(X_1,\dots,X_d)^t\sim N_d(0,I_d)$.
Hat man umgekehrt $X\sim N_d(0,I_d)$, dann gilt $f_X(x_1,\dots,x_d) = (2\pi)^{-\frac{d}{2}}\exp(-\underbrace{\frac{1}{2}\sum_{i=1}^d x_i^2}_{x^tx})$, d.h. die gemeinsame Dichte der Komponenten $X_1,\dots,X_d$ ist das Produkt der einzelnen Dichten $f_{X_i}(x_i) = (2\pi)^{-\frac{1}{2}}\exp(-\frac{1}{2}x_i^2)$, d.h. $X_1,\dots,X_d$ sind unabhängig und $N(0,1)$-verteilt.

Definition 14.1

(i) Das WMaß auf $(\mathbb{R}_+,\mathfrak{B}_+)$ mit Dichte $f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}},\ x>0$ heißt Chiquadrat-Verteilung mit n Freiheitsgraden (kurz: χ^2_n.)

(ii) Die Studentsche t-Verteilung mit n Freiheitsgraden, kurz: t_n, ist die Verteilung auf $(\mathbb{R},\mathfrak{B})$ mit Dichte
$f(x) = \frac{1}{\sqrt{\pi n}}\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}(1+\frac{x^2}{n})^{-\frac{n+1}{2}}$.

Klar: $\chi^2_n = \Gamma(\frac{n}{2},\frac{1}{2})$ (siehe Aufgabe 26)

Lemma 14.2 (i) Sind X und Y unabhängige ZV mit Dichte f_X und f_Y, wobei $Y>0$, so hat $Z:=\frac{X}{Y}$ die Dichte

$$f_Z(x) = \int_0^\infty y f_Y(y)f_X(yz)dy.$$

(ii) Sind X und Y unabhängig mit $X\sim N(0,1),\ Y\sim\chi^2_n$, so gilt: $\frac{\sqrt{n}X}{\sqrt{Y}}\sim t_n$.
Beweis: Übungen.

Satz 14.3 Es sei $X_1,\dots,X_n$ eine Stichprobe aus $N(\mu,\sigma^2)$.
$\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$ (Stichprobenmittelwert), $S_n^2 := \frac{1}{n-1}\sum_{i=1}^n(X_i-\overline{X_n})^2$ (Stichprobenvarianz), $T_n = \frac{\sqrt{n}(\overline{X_n}-\mu)}{S_n}$. Dann gilt:
(i) $\overline{X_n}$ und S_n^2 sind unabhängig.
(ii) $\frac{\sqrt{n}(\overline{X_n}-\mu)}{\sigma}\sim N(0,1)$.
(iii) $\frac{(n-1)S_n^2}{\sigma^2}\sim\chi^2_{n-1}$.
(iv) $T_n\sim t_{n-1}$.

Beweis: Sei $Y_i := \frac{X_i-\mu}{\sigma},\ 1\leq i\leq n$. Dann ist $Y_1,\dots,Y_n$ eine Stichprobe aus $N(0,1)$, und es gilt:
$\frac{\sqrt{n}(\overline{X_n}-\mu)}{\sigma} = \sqrt{n}\overline{Y_n},\ (n-1)\frac{S_n^2}{\sigma^2} = \sum_{i=1}^n(Y_i-\overline{Y_n})^2$.
Sei nun A eine orthogonale $n\times n$-Matrix mit erster Zeile $(\frac{1}{\sqrt{n}},\dots,\frac{1}{\sqrt{n}})$ (kann mit Gram-Schmidt konstruiert werden, für uns reicht hier die Existenz). Sei $Y:=(Y_1,\dots,Y_n)^t,\ Z:=AY$. Dann gilt: $Y\sim N_n(0,I_n),\ Z\sim N_n(0,AI_nA^t) = N_n(0,I_n)$.
Also bilden die $Z_1,\dots,Z_n$ eine Stichprobe aus $N(0,1)$. Nach Konstruktion gilt $Z_1:=\sqrt{n}\overline{Y_n}$, also folgt (ii).
Die Orthogonalität von A liefert außerdem $\sum_{i=1}^n Z_i^2 = Z^tZ = Y^tY = \sum_{i=1}^n Y_i^2$.

$$\sum_{i=1}^{n}(Y_i-\overline{Y_n})^2=\sum_{i=1}^{n}Y_i^2-2\overline{Y_n}\underbrace{\sum_{i=1}^{n}Y_i}_{n\overline{Y_n}}+n\overline{Y_n}^2=\sum_{i=1}^{n}Y_i{}^2-n(\overline{Y_n})^2=\sum_{i=1}^{n}Z_i^2-Z_1^2=\sum_{i=2}^{n}Z_i^2.$$

Folglich ist S_n^2 eine Funktion von $Z_2,\dots,Z_n$ und damit unabhängig von $\overline{X_n}$, das ja eine Funktion von Z_1 ist. Hiermit folgt (i).

Die Bezeichnung $\frac{(n-1)S_n^2}{\sigma^2}=\sum_{i=2}^{n}Z_i^2$ liefert auch (iii), den nach Aufgabe 33(a) zur Stochastik I ist Z_i^2 χ_1^2-verteilt, und mit der Faltungsgleichung der Gammafunktion (Aufgabe 26) folgt hieraus $\sum_{i=2}^{n}Z_i^2\sim\chi_{n-1}^2\quad(=\Gamma(\frac{n}{2},\frac{1}{2}))$.

Teil (iv) folgt nun mit Lemma 14.2(ii) aus den bereits bewiesenen Teilen (i)-(iii). □

Bemerkung 14.4 (Tests und Konfidenzintervalle für Erwartungswert bei Stichproben aus der Normalverteilung)
Es sei $t_{n;\alpha}$ das α-Quantil zur Verteilung t_n (siehe Tafelwerk, mit Computerprogramm).
Ist $X_1,\dots,X_n$ eine Stichprobe aus $N(\mu,\sigma^2)$ mit unbekanntem $\mu\in\mathbb{R}$ und $\sigma^2>0$, so folgt mit Satz 14.3(iv)

$$P_{\mu,\sigma^2}(t_{n-1,\frac{\alpha}{2}}\le\frac{\overline{X_n}-\mu}{\frac{S_n}{\sqrt{n}}}\le t_{n-1;1-\frac{\alpha}{2}})=1-\alpha$$

Dies gilt für alle $\mu\in\mathbb{R},\ \sigma^2>0$. Löst man die Ungleichung nach μ auf, so sieht man, daß

$$[\overline{X_n}-\tfrac{1}{\sqrt{n}}t_{n-1;1-\frac{\alpha}{2}}S_n,\overline{X_n}-\tfrac{1}{\sqrt{n}}t_{n-1;\frac{\alpha}{2}}S_n]$$

ein $100(1-\alpha)\%$-Konfidenzintervall für μ ist.
Die Symmetrie der t-Verteilung impliziert $t_{n-1;\frac{\alpha}{2}}=-t_{n-1;1-\frac{\alpha}{2}}$ (damit: $\overline{X_n}\pm\frac{1}{\sqrt{n}}t_{n-1;1-\frac{\alpha}{2}}S_n$). Für große n kann t_n durch $N(0,1)$ ersetzt werden (siehe Aufgabe 48).

Der in §7.4 besprochene Zusammenhang zwischen Tests und Konfidenzintervallen führt auf den von Anwendern sehr (zu) häufig benutzten t-Test zur Hypothese $H_0:\mu=\mu_0$ (μ_0 ist ein vorgegebener Wert):
Verwerfe, wenn $|\overline{X_n}-\mu_0|>\frac{1}{\sqrt{n}}t_{n-1;1-\frac{\alpha}{2}}S_n$.
Die obige Rechnung zeigt, daß dies ein Test zum Niveau α ist, der bemerkenswerter Weise für alle $\sigma^2>0$ das Niveau α, die zugelassene maximale Wahrscheinlichkeit für einen Fehler 1. Art, "ausschöpft".
Es gibt hiervon auch 'einseitige' Versionen für Hypothesen der Form $H_0:\mu\le\mu_0,\ H_0:\mu\ge\mu_0$.

14.3 Grenzwertsätze

Gesetze der Großen Zahlen (stark und schwach)

Erweiterung von $\mathbb{R}$ auf $\mathbb{R}^d$ ist einigermaßen trivial (komponentenweise). Beim ZGWS muß zunächst "$\xrightarrow{D}$" definiert werden: $X_n\xrightarrow{D}X$ bedeutet $\lim_{n\to\infty}F_n(x)=F(x)\quad\forall$ Stetigkeitspunkte von F, wobei $F_n(n\in\mathbb{N}),\ F$ die Verteilungsfunktionen zu $X_n(n\in\mathbb{N}),\ X$.

Es gilt der Stetigkeitssatz für charakteristische Funktionen $X_n\xrightarrow{D}X\iff\varphi_{X_n}(\theta)\to\varphi_X(\theta)\quad\forall\theta\in\mathbb{R}^d$.
Damit:

Satz 14.5 (ein mehrdimensionaler ZGWS)
Es sei $(X_n)_{n\in\mathbb{N}}$ eine iid-Folge von d-dimensionalen Zufallsvariablen mit Erwartungswertvektor μ und Kovarianzmatrix Σ. Dann gilt:

$$\boxed{\sqrt{n}(\overline{X_n}-\mu)\to Z,\ Z\sim N_d(0,\Sigma).}$$

Beweis: Sei $\sigma\in\mathbb{R}^d$. Setze $Y_i:=\theta^tX_i,\ u\in\mathbb{N}$. Auf diese läßt sich Satz 13.2 anwenden:
$\sqrt{n}(\overline{Y_n}-\theta^t\mu)\longrightarrow Z_\theta,\ Z_\theta\sim N(0,\theta^t\Sigma\theta)$.
Der Stetigkeitssatz (für dim = 1) liefert nun $E\exp(i(\sqrt{n}(\overline{Y_n}-\theta^t\mu)))\longrightarrow\exp(-\frac{1}{2}\theta^t\Sigma\theta)$.
Da θ beliebig war, und $\overline{Y_n}=\theta^t\overline{X_n}$ gilt, folgt hieraus $E\exp(i\theta^t(\sqrt{n}(\overline{X_n}-\mu)))\longrightarrow E\exp(i\theta^tZ)\quad\forall\theta$, also mit dem "mehrdimensionalen" Stetigkeitssatz die Behauptung. □

Beispiel 14.6
Gegeben sei ein stochastisches Experiment mit d möglichen Resultaten, die mit den Wahrscheinlichkeiten $p_1,\dots,p_d$ eintreten. X_n sei der d-dimensionale Zufallsvektor, dessen k-te Komponente 1 oder 0 ist, je nachdem, ob Resultat k in der n-ten Wiederholung eingetreten ist oder nicht. Dann beschreibt $S_n:=X_1+\dots+X_n\ (=\overline{X_n})$ die Verteilung der Resultate nach n Durchgängen. $S_n\sim Multinom(n,p),\ p=(p_1,\dots,p_d)^d$ ($\rightsquigarrow$ Stochastik I, §6.2.5)
Aufgabe 45: $\mu=EX=p,\ \Sigma=diag(p)-pp^t$, in Komponenten: $(\Sigma)_{ij}=\begin{cases}p_i-p_i^2 & ,\ i=j\\ -p_ip_j & ,\ \text{sonst}\end{cases}$.

Satz 14.5 liefert $\quad(*)\quad\frac{1}{\sqrt{n}}(S_n-n\mu)\xrightarrow{D}Z,\ Z\sim N_d(0,\Sigma)$.

Die Testgröße des populären χ^2-Anpassungstests auf p lautet: $T(X_1,\dots,X_n):=\frac{1}{n}\sum_{i=1}^{n}\frac{(S_{ni}-np_i)^2}{p_i}$.

Konkretes Beispiel mit echten Daten:

189 Würfelwürfe haben das folgende Resultat ergeben:

i	1	2	3	4	5	6
Werte von S_{ni}	30	37	26	29	29	38

Ist der Würfel fair? Man erhält $T(X) = 3.73\ldots$

$Y_n := (\sqrt{n}\frac{\overline{X_{n1}}-p_1}{\sqrt{p_1}}, \ldots, \sqrt{n}\frac{\overline{X_{nd}}-p_d}{\sqrt{p_d}})^t = diag(\frac{1}{\sqrt{p_1}}, \ldots, \frac{1}{\sqrt{p_d}})\sqrt{n}(\overline{X_n} - p)$

Aus $(*)$ folgt: $Y_n \xrightarrow{D} diag(\frac{1}{\sqrt{p_1}}, \ldots, \frac{1}{\sqrt{p_d}})Z = \tilde{Z}$ (continuous mapping theorem, Eigenschaften von $\xrightarrow{D}$).

$\tilde{Z}$ ist wieder normalverteilt, mit Mittelwertvektor 0 und Kovarianzmatrix

$$\tilde{\Sigma} = diag(\tfrac{1}{\sqrt{p_1}}, \ldots, \tfrac{1}{\sqrt{p_d}})(diag(p) - pp^t)diag(\tfrac{1}{\sqrt{p_1}}, \ldots, \tfrac{1}{\sqrt{p_d}}) = I_d - \begin{pmatrix} \sqrt{p_1} \\ \vdots \\ \sqrt{p_d} \end{pmatrix} (\sqrt{p_1}, \ldots, \sqrt{p_d}).$$

Der Vektor $\begin{pmatrix} \sqrt{p_1} \\ \vdots \\ \sqrt{p_d} \end{pmatrix}$ läßt sich zu einer orthogonalen Matrix A ergänzen (vergleiche Beweis zu Satz 14.3)

Es sei $\overline{Z} := A^t\tilde{Z}$. Dann ist $\overline{Z}$ wieder normalverteilt mit $E\overline{Z} = 0$ und Kovarianzmatrix $\overline{\Sigma} = A^t\tilde{\Sigma}A = \underbrace{A^tI_dA}_{I_d} - \underbrace{A^t \begin{pmatrix} \sqrt{p_1} \\ \vdots \\ \sqrt{p_d} \end{pmatrix}}_{(1,0,\ldots,0)^t} \underbrace{(\sqrt{p_1}, \ldots, \sqrt{p_d})A}_{(1,0,\ldots,0)} = diag(0, 1, \ldots, 1)$

Also: $\overline{Z_1} = 0$ f.s., $\overline{Z_2}, \ldots, \overline{Z_d}$ unabhängig und $N(0,1)$-verteilt.

Damit $\overline{Z}^t\overline{Z} = \sum_{i=2}^{d} \overline{Z_i}^2 \sim \chi^2_{d-1}$ (siehe Ende von Beweis zu Satz 14.3)

Insgesamt $T_{(n)} = Y_n^tY_n \xrightarrow{D} \tilde{Z}^t\tilde{Z} = \overline{Z}^t\overline{Z} \sim \chi^2_{n-1}$ (continuous mapping theorem)
Wir haben damit die asymptotische Verteilung der Testgröße; diese hängt nicht von p ab.
Also: Wähle bei Signifikanzniveau α als kritischen Wert c_α (d.h. Ablehnung bei $T > c_\alpha$) das $1-\alpha$-Quantil zu χ^2_{d-1} ($\rightsquigarrow$ Tafel, Computer)
Dieser Test hält bei großem n das Niveau α ein. Faustregel: $np_i \geq 5$ für $i = 1, \ldots, d$.
Im numerischen Beispiel: $\chi^2_{5;1-0.05} = 11.1\ldots$, somit keine Ablehnung.

Kapitel 15

Martingale

Wichtiges Hilfsmittel der modernen Stochastik und hat zahlreiche Anwendunngen (u.a. bei Glücksspielstrategien)

15.1 Bedingte Erwartungswerte

$(\Omega, \mathfrak{A}, P)$ WRaum, $\mathfrak{F}$ Unter-σ-Algebra von $\mathfrak{A}$.

Definition 15.1 Es sei X eine Zufallsvariable mit existierendem Erwartungswert (also $E|X| \leq \infty$). Dann heißt $Y : \Omega \to \mathbb{R}$ eine Version des bedingten Erwartungswertes von X unter F, wenn gilt:

$$\boxed{Y \text{ ist } \mathfrak{F}\text{-meßbar}, \int_F Y dP = \int_F X dP}$$

Extremfälle:

- $\mathfrak{F} = \{\emptyset, \Omega\}$. Alle $\mathfrak{F}$-meßbaren $Y : \Omega \to \mathbb{R}$ sind konstant. Wähle $Y \equiv EX$: $\int_F Y dP = EX \cdot P(F) = \int_F X dP$, da nur $F = \emptyset,\ F = \Omega$ in Frage kommen.
- $\mathfrak{F} = \mathfrak{A}$. Wähle $Y = X$.

Satz 15.2 Es sei X eine Zufallsvariable mit $E|X| < \infty$. Dann gilt:

(i) Es existiert eine Version der bedingten Erwartung von X unter $\mathfrak{F}$.

(ii) Sind Y_1 und Y_2 Versionen der bedingten Erwartung von X unter $\mathfrak{F}$, so gilt $P(Y_1 = Y_2) = 1$ (In Worten: bedingte Erwartungswerte sind fast sicher eindeutig).

(iii) Ist Y eine Version der bedingten Erwartung von X unter $\mathfrak{F}$, so exitiert zu Y der Erwartungswert, und es gilt $EY = EX$.

Beweis:

(i) Es sei P_0 die Restriktion von P auf $\mathfrak{F}$. Durch $\nu_+(F) = \int_F X^+ dP$ für alle $F \in \mathfrak{F}$ (wobei $X^+ := \max\{X, 0\}$) wird ein P_0-stetiges Maß auf $(\Omega, \mathfrak{F})$ definiert ($P_0(N) = 0 \implies \nu_+(N) = 0$), also existiert nach dem Satz von Radon-Nikodym eine Dichte, d.h. eine $\mathfrak{F}$-meßbare Abbildung $Y^+ : \Omega \to \mathbb{R}^+$ mit

$$\int_F X^+ dP = \nu_+(F) = \int_F Y^+ dP_0 = \int_F Y^+ dP \quad \forall F \in \mathfrak{F}.$$

Verfahre analog bei X^-, und setze $Y := Y^+ - Y^-$.

$$\int_F Y dP = \int_F Y^+ dP - \int_F Y^- dP = \int_F X^+ dP - \int_F X^- dP = \int_F X dP.$$

(ii) $F_1 := \{Y_1 < Y_2\},\ F_2 := \{Y_1 > Y_2\}$, beide in $\mathfrak{F}$.

$$\int |Y_1 - Y_2| dP = \int_{F_1} (Y_2 - Y_1) dP + \int_{F_2} (Y_1 - Y_2) dP = \int_{F_1} Y_2 dP - \int_{F_1} Y_1 dP + \int_{F_2} Y_1 dP - \int_{F_2} Y_2 dP =$$

$$\int_{F_1} X dP - \int_{F_1} X dP + \int_{F_2} X dP - \int_{F_2} X dP = 0, \text{ also } |Y_1 - Y_2| = 0 \ P\text{-f.s. und damit } Y_1 = Y_2 \ P\text{-f.s.}$$

(iii) $\int Y^+ dP = \int_{\{Y>0\}} Y dP = \int_{\{Y>0\}} X dP \leq \int |X| dP < \infty$, analog: $\int Y^- dP < \infty$, also ist Y P-integrierbar.

Mit $\Omega \in \mathfrak{F}$ folgt schließlich $EY = \int_\Omega Y dP = \int_\Omega X dP = EX$. □

Aus Abschnitt 8.7: $\mathfrak{L}^p(\Omega, \mathfrak{A}, \mu) = \{f : \Omega \to \mathbb{R} : \int |f|^p d\mu < \infty\},\ ||f||_p := (\int |f|^p d\mu)^{\frac{1}{p}}$.
Auf diesen Räumen wird durch $f \sim g \ :\iff \mu(f \neq g) = 0$ eine Äquivalenzrelation definiert. Der Raum $L^p(\Omega, \mathfrak{A}, \mu)$ der Äquivalenzklassen $\tilde{f} := \{g \in \mathfrak{L}^p(\Omega, \mathfrak{A}, \mu) : g \sim f\}$ wird mit $||\tilde{f}||_p := ||f||_p$ (hängt nicht von der Wahl von $f \in \tilde{f}$ ab) zu einem Banach-Raum.
Im Falle $p = 2$ hat man mit $< \tilde{f}, \tilde{g} > := \int f g d\mu$ sogar einen Hilbert-Raum[1] (auf $\mathfrak{L}^p$ ist $|| \cdot ||_p$ nur eine Seminorm, da

[1] Wo ist alles erlaubt? — Im Krieg, in der Liebe und im Hilbert-Raum!

$||f||_p = 0$ auch bei $f \neq 0$ gelten kann.)

Nach Satz 15.2 (ii) bilden die verschiedenen Versionen der bedingten Erwartungswerte von X bzgl. $\mathfrak{F}$ eine Äquivalenzklasse (d.h. $L^1(\Omega, \mathfrak{F}, P) \subset L^1(\Omega, \mathfrak{A}, P)$), wir schreiben hierfür $E[X|\mathfrak{F}]$ bzw. $E[X|Z]$ oder $E[X|Z_i, i \in I]$ im Falle $\mathfrak{F} = \sigma(Z)$ bzw. $\mathfrak{F} = \sigma(\{Z_i : i \in I\})$. Man kann den Übergang von X zu $E[X|\mathfrak{F}]$ als Projektion von $L^1(\Omega, \mathfrak{A}, P)$ auf $L^1(\Omega, \mathfrak{F}, P)$ auffassen. Im folgenden werden wir den Unterschied Funktion/Äquivalenzklasse weitgehend ignorieren.
Konkret: Aussagen zu bedingten Erwartungswerten gelten i.a. nur P-f.s., z.B. $E[E[X|\mathfrak{F}]|\mathfrak{F}] = E[X|\mathfrak{F}]$ P-f.s.

Satz 15.3

(i) $E[\alpha X + \beta Y|\mathfrak{F}] = \alpha E[X|\mathfrak{F}] + \beta E[Y|\mathfrak{F}]$ P-f.s. (Linearität)

(ii) $X \leq Y \implies E[X|\mathfrak{F}] \leq E[Y|\mathfrak{F}]$ P-f.s. (Monotonie)

(iii) $\varphi : \mathbb{R} \to \mathbb{R}$ konvex und gilt $E|\varphi(X)| < \infty$, so folgt $\varphi(E[X|\mathfrak{F}]) \leq E[\varphi(X)|\mathfrak{F}]$ P-f.s. (bedingte Jensensche Ungleichtung)

Beweis:

(i) Sei $F \in \mathfrak{F}$. Dann gilt:
$$\int_F (\alpha E[X|\mathfrak{F}] + \beta E[Y|\mathfrak{F}])dP = \alpha \int_F E[X|\mathfrak{F}]dP + \beta \int_F E[Y|\mathfrak{F}]dP = \alpha \int_F XdP + \beta \int_F YdP = \int_F (\alpha X + \beta Y)dP$$
Da mit $E[X|\mathfrak{F}]$, $E[Y|\mathfrak{F}]$ auch $\alpha E[X|\mathfrak{F}] + \beta E[Y|\mathfrak{F}]$ $\mathfrak{F}$-meßbar ist, leistet die Linearkombination das Verlangte.

(ii) Wegen der bereits bewiesenen Linearität, reicht es, den Spezialfall $X \geq 0 \implies E[X|\mathfrak{F}] \geq 0$ P-f.s. zu beweisen.
Sei $F := \{E[X|\mathfrak{F}] < 0\}$ $(\in \mathfrak{F})$.
$$0 \geq \int_F E[X|\mathfrak{F}]dP = \int_F XdP \geq 0.$$
Also: $\int_F E[X|\mathfrak{F}]dP = 0$, d.h. $P(E[X|\mathfrak{F}]1_F = 0) = 1$ und damit $P(E[X|\mathfrak{F}] < 0) = 0$.

(iii) Übungsaufgabe. □

Satz 15.4

(i) (bedingte Version des Satzes von der monotonen Konvergenz)
Sind $X, X_1, X_2, \ldots$ Zufallsvariablen mit $0 \leq X_n \uparrow X$, so gilt $E[X_n|\mathfrak{F}] \uparrow E[X|\mathfrak{F}]$ P-f.s..

(ii) (bedingte Version des Lemmas von Fatou)
Sind $X_1, X_2, \ldots$ nicht negative Zufallsvariablen, so gilt $E[\liminf_{n\to\infty} X_n|\mathfrak{F}] \leq \liminf_{n\to\infty} E[X_n|\mathfrak{F}]$ P-f.s..

(iii) (bedingte Version des Satzes von der majorisierten Konvergenz)
Sind $X, Y, X_1, X_2, \ldots$ Zufallsvariablen mit $X_n \to X$ P-f.s., $|X_n| \leq Y$ und $EY < \infty$, so gilt $E[X_n|\mathfrak{F}] \to E[X|\mathfrak{F}]$ P-f.s..

Beweis:

(i) Die Monotonie von $Y_n := E[X_n|\mathfrak{F}]$ folgt aus der Monotonie-Eigenschaft des bedingten Erwartungswerts (Satz 15.3 (ii)): Sei $Y := \sup_{n\in\mathbb{N}} Y_n$, dann: $Y_n \uparrow Y$ $(P\text{-f.s.})$

Nach dem "unbedingten" Satz von der monotonen Konvergenz, angewendet auf die Folgen $(1_F Y_n)_{n\in\mathbb{N}}$ und $(1_F X_n)_{n\in\mathbb{N}}$ gilt für jedes $F \in \mathfrak{F}$: $\int_F YdP = \sup_n \int_F Y_n dP = \sup_n \int_F X_n dP = \int_F XdP$.

Da Y auch $\mathfrak{F}$-meßbar ist, leistet es das Verlangte.

(ii),(iii) Übungsaufgaben □

Einige Eigenschaften, die kein Analogon bei gewöhnlichen Erwartungswerten haben (können):

Satz 15.5

(i) Ist $\mathfrak{G}$ eine Unter-σ-Algebra von $\mathfrak{F}$, so gilt $E[E[X|\mathfrak{F}]|\mathfrak{G}] = E[X|\mathfrak{G}]$

(ii) Ist Y $\mathfrak{F}$-meßbar, so gilt: $E[Y \cdot X|\mathfrak{F}] = Y \cdot E[X|\mathfrak{F}]$ P-fast sicher, vorausgesetzt Y ist beschränkt oder $E|Y|^p < \infty$, $E|X|^q < \infty$ für ein Paar (p, q) mit $\frac{1}{p} + \frac{1}{q} = 1$ (wird gebraucht für $E|X \cdot Y| < \infty$).

(iii) Ist $\mathfrak{G}$ eine von $\sigma(\{\sigma(X), \mathfrak{F}\})$ unabhängige σ-Algebra, so gilt: $E[X|\sigma(\mathfrak{F}, \mathfrak{G})] = E[X|\mathfrak{F}]$ P-fast sicher.
Insbesondere: Sind $\sigma(X)$ und $\mathfrak{G}$ unabhängig, so gilt $E[X|\mathfrak{G}] = EX$ P-f.s..

Beweis:

(i) $E[X|\mathfrak{G}]$ ist $\mathfrak{G}$-meßbar. Für alle $G \in \mathfrak{G}$ gilt $\int_G E[X|\mathfrak{G}]dP = \int_G XdP = \int_G E[X|\mathfrak{F}]dP$, da $G \in \mathfrak{F}$, also leistet $E[X|\mathfrak{G}]$ das von $E[E[X|\mathfrak{F}]|\mathfrak{G}]$ verlangte.

(ii) Ein Fall für die "übliche Maschinerie":
Sei $Y = 1_A$ mit $A \in \mathfrak{F}$. Dann gilt für alle $F \in \mathfrak{F}$:
$$\int_F Y \cdot E[X|\mathfrak{F}]dP = \int_{\underbrace{F \cap A}_{\in \mathfrak{F}}} E[X|\mathfrak{F}]dP = \int_{F\cap A} XdP = \int_F 1_A X dP = \int_F X \cdot Y dP.$$
$Y \cdot E[X|\mathfrak{F}]$ ist $\mathfrak{F}$-meßbar, d.h. die Behauptung gilt für Indikatorfunktionen.

Linearität (Satz 15.3.(i)) liefert die Behauptung für primitive Y. Die bedingte Version des Satzes von der monotonen Konvergenz (Satz 15.4.(i)) liefert die Behauptung für $Y \geq 0$. Im letzten Schritt zerlege in Positiv- und Negativteil.

(iii) Wir können $X \geq 0$ annehmen (sonst X^+, X^- getrennt betrachten)
Betrachte $A \to \int\limits_A X dP$, $A \to \int\limits_A E[X|\mathfrak{F}]dP$. (Nach Definition von $E[X|\mathfrak{F}]$ stimmen diese Abbildungen auf $\mathfrak{F}$ überein.) Sei $F \in \mathfrak{F}$, $G \in \mathfrak{G}$. Da 1_G und $X1_F$ unabhängig sind, ebenso wie 1_G und $E[X1_F|\mathfrak{F}] = 1_F E[X|\mathfrak{G}]$ folgt:
$\int\limits_{F\cap G} X dP = \int 1_G(X \cdot 1_F)dP \overset{\text{Multiplikations-}}{\underset{\text{satz}}{=}} \int 1_G dP \cdot \int X1_F dP = \int 1_G dP \cdot \int\limits_F E[X|\mathfrak{F}]dP \overset{\text{Mult.-}}{\underset{\text{satz}}{=}} \int 1_G 1_F E[X|\mathfrak{F}]dP = \int\limits_{F\cap G} E[X|\mathfrak{F}]dP$, d.h. die beiden Maße stimmen auf einem $\cap$-stabilen Erzeuger von $\sigma(\mathfrak{F},\mathfrak{G})$ überein.
Da sie außerdem dieselben Gesamtmassen haben, sind sie gleich, d.h. $\int\limits_A E[X|\mathfrak{F}]dP = \int\limits_A X dP$ für alle $A \in \sigma(\mathfrak{F},\mathfrak{G})$.
Wegen der $\sigma(\mathfrak{F},\mathfrak{G})$-meßbarkeit von $E[X|\mathfrak{F}]$ (folgt aus $\mathfrak{F} \subset \sigma(\mathfrak{F},\mathfrak{G})$) hat man daher $E[X|\mathfrak{F}] = E[X|\sigma(\mathfrak{F},\mathfrak{G})]$ P-f.s..
Wähle $\mathfrak{F} = \{\emptyset,\Omega\}$. Dann $\sigma(\mathfrak{F},\mathfrak{G}) = \mathfrak{G}$, also $E[X|\mathfrak{G}] = E[X|\mathfrak{F}] = EX$ P-f.s.. □

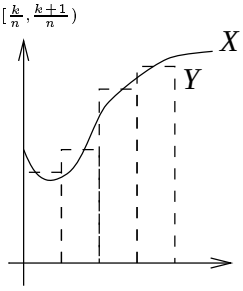
Satz 15.6 Gilt $EX^2 < \infty$, so minimiert $E[X|\mathfrak{F}]$ die Funktion $Y \to E(X-Y)^2$ auf der Menge der $\mathfrak{F}$-meßbaren Zufallsvariablen mit endlichem zweiten Moment (also: $E[\cdot|\mathfrak{F}]$ ist Orthogonalprojektion in $L^2(\Omega,\mathfrak{A},P)$ auf $L^2(\Omega,\mathfrak{F},P)$.)
Beweis: Für alle $\mathfrak{F}$-meßbaren Y mit $EY^2 < \infty$ gilt:
$E(X-Y)^2 = E((X-E[X|\mathfrak{F}]) + (E[X|\mathfrak{F}]-Y))^2 = E(X-E[X|\mathfrak{F}])^2 + E(E[X|\mathfrak{F}]-Y)^2$, denn
$E(X-E[X|\mathfrak{F}])(E[X|\mathfrak{F}]-Y) = E(E[(X-E[X|\mathfrak{F}])(E[X|\mathfrak{F}]-Y)|\mathfrak{F}]) = E((E[X|\mathfrak{F}]-Y)\underbrace{E[X-E[X|\mathfrak{F}]|\mathfrak{F}]}_{0}) = 0$

Der erste Term der Zerlegung hängt nicht von Y ab, der zweite ist ≥ 0 und $= 0$ für $Y = E[X|\mathfrak{F}]$. □

Stochastische Interpretation: $\mathfrak{F}$ steht für die "zugängliche Information", X soll 'vorhergesagt' werden. Wählt man die mittlere quadratische Abweichung als "Verlustfunktion", so besagt Satz 15.6, daß $E[X|\mathfrak{F}]$ optimal ist.

Beispiel 15.7

(i) $(\Omega,\mathfrak{A},P) = ([0,1),\mathfrak{B}_{[0,1)},l|_{[0,1)})$, $\mathfrak{F} = \sigma(\{[\frac{k}{n},\frac{k+1}{n}) : k = 0,1,\ldots,n-1\})$ (n fest).
Jede $\mathfrak{F}$-meßbare Funktion $\varphi : \Omega \to \mathbb{R}$ muß auf den erzeugenden Intervallen $[\frac{k}{n},\frac{k+1}{n})$, $k = 0,1,\ldots,n-1$, konstant sein. Ist y_{nk} der Wert von $Y = E[X|\mathfrak{F}]$ auf $[\frac{k}{n},\frac{k+1}{n})$, so muß $\int\limits_{[\frac{k}{n},\frac{k+1}{n})} Y dP = \frac{1}{n}y_{nk} = \int\limits_{[\frac{k}{n},\frac{k+1}{n})} X dP$ gelten.



Also: $Y = n\sum\limits_{k=0}^{n-1} (\int\limits_{[\frac{k}{n},\frac{k+1}{n})} X dP) 1_{[\frac{k}{n},\frac{k+1}{n})}$

Es gilt $\mathfrak{F} = \sigma(Z)$, mit $Z(\omega) = \frac{\lfloor n\omega \rfloor}{n}$.
Z zeigt an, in welches Intervall das Ergebnis des Zufallsexperiments fällt.
$E[X|\mathfrak{F}] = \varphi(Z)$ mit $\varphi(\frac{k}{n}) = n\int\limits_{[\frac{k}{n},\frac{k+1}{n})} X dP$.

(ii) Es sei $X_1, X_2, \ldots$ eine iid-Folge von Zufallsvariablen mit $E|X_n| < \infty$ und $S_n := \sum\limits_{k=1}^n X_k$. $E[X_1|S_n] = ?$
(Vermutung: $\frac{1}{n}S_n$)
Für alle $B \in \mathfrak{B}$ gilt: $\int X_1 1_{\{S_n \in B\}} dP = \int\limits_{\{x\in\mathbb{R}^n : \Sigma x_i \in B\}} \ldots \int x_1 P^{X_1}(dx_1)\ldots P^{X_n}(dx_n) = \int\limits_{\{x\in\mathbb{R}^n : \Sigma x_i \in B\}} \ldots \int x_j P^{X_1}(dx_1)\ldots P^{X_n}(dx_n) = \int X_j 1_{\{S_n \in B\}} dP$, denn $\bigotimes\limits_{i=1}^n P^{X_i}$ ist invariant unter Koordinatenpermutation.
Hat also Y die Eigenschaft $\int\limits_{\{S_n\in B\}} Y dP = \int\limits_{\{S_n\in B\}} X_1 dP$, so gilt sogar $\int\limits_{\{S_n\in B\}} Y dP = \int\limits_{\{S_n\in B\}} X_j dP$ für $j = 1,\ldots,n$, also:
$E[X_1|S_n] = E[X_2|S_n] = \ldots = E[X_n|S_n]$.
Bilde Summen: $(nE[X_1|S_n] =)$ $\sum\limits_{i=1}^n E[X_i|S_n] = E[\sum\limits_{i=1}^n X_i|S_n] = E[S_n|S_n] = S_n$.
Damit folgt: $\boxed{E[X_1|S_n] = \frac{1}{n}S_n}$
Betrachtet man $\mathfrak{F} = \sigma\{S_m : m \geq n\}$, so folgt mit Satz 15.5(iii): $E[X_1|S_n, S_{n+1}, \ldots] = \frac{1}{n}S_n$.

15.2 Martingale

Es seien $(\Omega,\mathfrak{A},P)$ ein WRaum und $(T,\leq)$ eine gerichtete Menge, d.h.
(i) "$\leq$" ist Halbordnung auf T.
(ii) $\forall s,t \in T\ \exists u \in T : s \leq u, t \leq u$.
Weiter sei $(\mathfrak{F}_t)_{t\in T}$ eine bzgl. "$\leq$" aufsteigende Familie von Unter-σ-Algebren von $\mathfrak{A}$, d.h. $s \leq t \implies \mathfrak{F}_s \leq \mathfrak{F}_t$ (Filtration)

Definition 15.8

Der stochastische Prozeß $(X_t)_{t\in T}$ (d.h. $\forall t \in T$ ist $X_t : \Omega \to \mathbb{R}$ eine ZV) heißt zu $(\mathfrak{F}_t)_{t\in T}$ adaptiert, wenn X_t $\mathfrak{F}_t$-meßbar ist für alle t.
$(X_t,\mathfrak{F}_t)_{t\in T}$ heißt Martingal, wenn $E|X_t| < \infty$ $\forall t \in T$ und $s,t \in T, s \leq t \implies E[X_t|\mathfrak{F}_s] = X_s$ P-f.s..
Im Falle $X_s \leq E[X_t|\mathfrak{F}_s]$ für alle $s \leq t$ nennt man $(X_t,\mathfrak{F}_t)_{t\in T}$ ein Submartingal, bei $\geq$ Supermartingal.

Interpretation: T Zeit, $(X_t)_{t\in T}$ beschreibt die zufällige Entwicklung des "stochastischen Systems": X_t ist Position eines Partikels zur Zeit t, Aktienkurs o.ä.

Wenn $(X_t)_{t\in T}$ gegeben ist, so läßt sich eine zugehörige Filtration $(\mathfrak{F}_t)_{t\in T}$ konstruieren durch $\mathfrak{F}_t := \sigma(\{X_s : s \le t\})$. Diese Familie nennt man die natürliche Filtration zu $(X_t)_{t\in T}$.
Klar: $(X_t)_{t\in T}$ ist zur zugehörigen Filtration adaptiert (X_t ist $\mathfrak{F}_t$-meßbar.)

Grob: Bei einem Martingal ist der gegenwärtige Wert der beste Vorhersagewert für zukünftige Werte (Satz 15.6). Bei einem Martingal bleiben wegen $EX_t = E(E[X_t|\mathfrak{F}_s]) = EX_s$ die Erwartungswerte konstant, bei einem Supermartingal nehmen sie im Laufe der Zeit ab, bei Submartingalen zu.

Ist $(X_t, \mathfrak{F}_t)_{t\in T}$ ein Supermartingal, so gilt:
$(*)$ $\int\limits_A X_s dP \ge \int\limits_A X_t dP$ für alle $s,t \in T$ mit $s \le t$ und alle $A \in \mathfrak{F}_s$.
Ist umgekehrt $(X_t)_{t\in T}$ zu $(\mathfrak{F}_t)_{t\in T}$ adaptiert, und gilt $E|X_t| < \infty \quad \forall t \in T$, so impliziert $(*)$, daß $(X_t, \mathfrak{F}_t)_{t\in T}$ ein Supermatingal ist ($\rightsquigarrow$ Aufgabe 54.)

Die analoge Aussage für Submartingale und Martingale erhält man mit:
$(X_t, \mathfrak{F}_t)_{t\in T}$ Submartingal $\iff$ $(-X_t, \mathfrak{F}_t)_{t\in T}$ ist Supermatingal.
$(X_t, \mathfrak{F}_t)_{t\in T}$ Martingal $\iff$ $(X_t, \mathfrak{F}_t)_{t\in T}$ ist Super- und Submartingal.

Man sagt (abkürzend): "$(X_t)_{t\in T}$ Martingal", wenn $(X_t, \mathfrak{F}_t)_{t\in T}$ mit der natürlichen Filtration $(\mathfrak{F}_t)_{t\in T}$ ein Martingal ist.

Beispiel 15.9

(i) Es sei $T = \mathbb{N}$, $(X_n)_{n\in\mathbb{N}}$ eine iid-Folge von Zufallsvariablen mit Erwartungswert μ. $(S_n)_{n\in\mathbb{N}}$ sei die Folge der Partialsummen $S_n := \sum\limits_{k=1}^{n} X_k$. Weiter sei $\mathfrak{F}_n := \sigma(\{X_1,\dots,X_n\})$ $(= \sigma(\{S_1,\dots,S_n\}))$ (also: $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ ist die natürliche Filtration zu $(S_n)_{n\in\mathbb{N}}$).
$E[S_{n+1}|\mathfrak{F}_n] = E[S_n + X_{n+1}|\mathfrak{F}_n] = E[S_n|\mathfrak{F}_n] + E[X_{n+1}|\mathfrak{F}_n] = S_n + \overbrace{\mu}^{E(X_n)}$ (nach Satz 15.3(i), Satz 15.5(iii)) für alle $n \in \mathbb{N}$, also $E[S_{n+1}|\mathfrak{F}_n] \overset{\ge}{\underset{\le}{=}} S_n$, wenn $\mu \overset{\ge}{\underset{\le}{=}} 0$.
$E[S_{n+2}|\mathfrak{F}_n] = E[E[S_{n+2}|\mathfrak{F}_{n+1}]|\mathfrak{F}_n]$ ("Turm")
Hieraus folgt, daß $(S_n)_{n\in\mathbb{N}}$ ein Submartingal / Martingal / Supermartingal ist, wenn $\mu \ge 0$ / $\mu = 0$ / $\mu \le 0$ gilt ($\rightsquigarrow$ Satz 15.5(i))

(ii) (In diesem Beispiel ist T keine Teilmenge von $\mathbb{R}$ ...)
Es sei Q ein weiteres WMaß auf $(\Omega, \mathfrak{A})$, daß von P dominiert wird (d.h. $P(N) = 0 \implies Q(N) = 0$)
Sei T die Menge der endlichen meßbaren Zerlegungen (Partitionen) $\mathfrak{Z}$ von Ω, d.h. $\mathfrak{Z} = (A_1,\dots,A_n)$ mit $n \in \mathbb{N}$, $A_1,\dots,A_n \in \mathfrak{A}$, $A_i \cap A_j = \emptyset$ für $i \ne j$, $\bigcup\limits_{i=1}^{n} A_i = \Omega$. Sind $\mathfrak{Z}_1 = (A_1,\dots,A_n)$ und $\mathfrak{Z}_2 = (B_1,\dots,B_m)$ zwei solche Zerlegungen, so nennen wir $\mathfrak{Z}_2$ feiner als $\mathfrak{Z}_1$ und schreiben $\mathfrak{Z}_1 \le \mathfrak{Z}_2$, wenn gilt: $\forall k \in \{1,\dots,n\} \exists I \subseteq \{1,\dots,m\}$: $A_k = \bigcup\limits_{i\in I} B_i$.
Zu zwei Partitionen gibt es stets eine gemeinsame feinere (beispielsweise das System aller Durchschnitte), also $(T, \le)$ ist eine gerichtete Menge.
Für jedes $t = \mathfrak{Z} = (A_1,\dots,A_n) \in T$ setzen wir nun $\mathfrak{F}_t := \sigma(\{A_1,\dots,A_n\})$, $X_t = \sum\limits_{i=1}^{n} \frac{Q(A_i)}{P(A_i)} 1_{A_i}$
(=0, wenn $P(A_i) = 0$).
Offensichtlich ist dann $(X_t)_{t\in T}$ zur Filtration $(\mathfrak{F}_t)_{t\in T}$ adaptiert, und es gilt
$$E|X_t| = EX_t = \sum_{i=1}^{n} \frac{Q(A_i)}{P(A_i)} E1_{A_i} = \sum_{i=1}^{n} Q(A_i) = 1.$$
Es seien nun $s = (A_1,\dots,A_n)$ und $t = (B_1,\dots,B_m)$ zwei Zerlegungen mit $s \le t$. Für $k = 1,\dots,n$ gilt:
$$\int\limits_{A_k} X_t dP = \sum_{i=1}^{m} \frac{Q(B_i)}{P(B_i)} \underbrace{\int\limits_{A_k} 1_{B_i} dP}_{P(A_k\cap B_i)} = \sum_{\{i:B_i\subseteq A_k\}} \frac{Q(B_i)}{P(B_i)} P(B_i) = Q(A_k).$$
Außerdem: $\int\limits_{A_k} X_s dP = \sum\limits_{i=1}^{n} \frac{Q(A_i)}{P(A_i)} \underbrace{\int\limits_{A_k} 1_{A_i} dP}_{P(A_i\cap A_k)} = Q(A_k).$
Mit Aussage $(*)$ folgt, daß $(X_t, \mathfrak{F}_t)_{t\in T}$ ein Martingal ist.

(iii) Es sei $(\mathfrak{F}_t)_{t\in T}$ eine Filtration, X eine Zufallsvariable mit endlichem Erwartungswert und $X_t := E[X|\mathfrak{F}_t]$.
Nach Konstruktion ist dann $(X_t)_{t\in T}$ zu $(\mathfrak{F}_t)_{t\in T}$ adaptiert und mit Satz 15.5(i) folgt für $s \le t$ $E[X_t|\mathfrak{F}_s] = E[E[X|\mathfrak{F}_t]|\mathfrak{F}_s] = E[X|\mathfrak{F}_s] = X_s$ (wieder P-f.s.), also ist $(X_t, \mathfrak{F}_t)_{t\in T}$ eine Martingal.

Satz 15.10 (Submartingalungleichung von Doob)
Ist $(X_t, \mathfrak{F}_t)_{t=1,\dots N}$ ein Submartingal, so gilt für alle $c > 0$:
$$c \cdot P(\max_{1\le k\le n} X_k \ge c) \le \int\limits_{\{\max\limits_{1\le k\le n} X_k \ge c\}} X_n dP \le EX_n^+$$

Beweis: $A := \{\max_{1\le k\le n} X_k \ge c\},\ A_i := \{X_j < c \text{ für } j = 1,\dots,i-1,\ X_i \ge c\}$.
Dann gilt: $A = A_1 + A_2 + \dots + A_n;\ A_i \in \mathfrak{F}_i,\ X_i \ge c$ auf A_i.
Dann $\int_{A_i} X_n dP \ge \int_{A_i} X_i dP \ge c \cdot P(A_i)$.
Summiere nun über i: $\int_A X_n dP \ge c \cdot P(A)$, der zweite Teil der Ungleichung folgt mit $X_n 1_A \le X_n^+$. □

Mit den Resultaten des ersten Paragraphen erhält man leicht eine Reihe von Eigenschaften der Klasse der Martingale (Supermartingale bzw. Submartingale). Aus der Linearität des bedingten Erwartungswertes folgt beispielsweise unmittelbar, daß mit $(X_t,\mathfrak{F}_t)_{t\in T}$ und $(Y_t,\mathfrak{F}_t)_{t\in T}$ auch $(\alpha X_t + \beta Y_t,\mathfrak{F}_t)_{t\in T}$ ein Supermartingal ist, wenn $\alpha \ge 0,\ \beta \ge 0$; die analoge Aussage für Martingale gilt für alle $\alpha,\beta \in \mathbb{R}$. Ist T_0 eine gerichtete Teilmenge von T, so ist mit $(X_t,\mathfrak{F}_t)_{t\in T}$ auch $(X_t,\mathfrak{F}_t)_{t\in T_0}$ ein Submartingal. Die bedingte Version der Jensenschen Ungleichung liefert unmittelbar:

Satz 15.11
Ist $(X_t,\mathfrak{F}_t)_{t\in T}$ ein Martingal und $\varphi : \mathbb{R} \to \mathbb{R}$ eine konvexe Funktion mit $E|\varphi(X_t)| < \infty$ für alle $t \in T$,
so ist $(\varphi(X_t),\mathfrak{F}_t)_{t\in T}$ ein Submartingal.

Insbesondere ist bei einem quadratisch integrierbaren Martingal $(X_t,\mathfrak{F}_t)_{t\in T}$, d.h. $E({X_t}^2) < \infty \quad \forall t \in T$, der Prozeß $({X_t}^2,\mathfrak{F}_t)_{t\in T}$ ein Submartingal. Weitere Eigenschaften dieser Art werden in den Übungsaufgaben behandelt.

15.3 Stoppzeiten und Transformationen

Wir betrachten in diesem Absatz Prozesse mit Zeitmenge $T = \mathbb{N}_0 = \{0,1,2,\dots\}$. Es ist $(\Omega,\mathfrak{A},P)$ ein WRaum und $(\mathfrak{F}_t)_{t\in\mathbb{N}_0}$ eine Filtration. Martingale etc. sollen sich stets auf $(\mathfrak{F}_n)_{n\in\mathbb{N}_0}$ beziehen. Man sieht leicht, daß bei diesem T die (Super-, Sub-) Martingaleigenschaft bereits aus $E[X_{n+1}|\mathfrak{F}_n] \begin{matrix}\le\\ \ge\end{matrix} X_n$ für alle $n \in \mathbb{N}_0$ folgt (Induktion, Satz 15.5(i)).

Im nachfolgenden ist gelegentlich eine Interpretation über Glücksspiele hilfreich. Hierbei ist X_n das Kapital eines Spielers nach n Durchgängen, wenn jede Runde eine Geldeinheit gesetzt wird (X_0 ist das Anfangskapital), $\mathfrak{F}_n := \sigma(\{X_0,\dots,X_n\})$ repräsentiert die nach n Durchgängen vorhandene Information. Weiter ist $X_{n+1} - X_n$ der Gewinn pro gesetzter Geldeinheit in der nächsten Runde, $E[X_{n+1} - X_n|\mathfrak{F}_n]$ ist der zu erwartende Gewinn in der nächsten Runde bei Kenntnis des gesamten bisherigen Spielverlaufs. Also wegen $E[X_{n+1} - X_n|\mathfrak{F}_n] = E[X_{n+1}|\mathfrak{F}_n] - X_n$ gilt:

$(X_n)_{n\in\mathbb{N}_0}$ ist $\left\{\begin{array}{l}\text{Supermartingal}\\ \text{Martingal}\\ \text{Submartingal}\end{array}\right\}$ bedeutet, das Spiel ist $\left\{\begin{array}{l}\text{nachteilig}\\ \text{fair}\\ \text{vorteilhaft}\end{array}\right\}$ (aus Sicht des Spielers).

Betrachten wir nun die strategischen Möglichkeiten: In der n-ten Runde werden c_n Geldeinheiten eingesetzt, wobei natürlich c_n $\mathfrak{F}_{n-1}$-meßbar sein muß. Das Kapital des Spielers nach n Durchgängen ist dann $X_0 + \sum_{k=1}^{n} c_k(X_k - X_{k-1})$.

Definition 15.12 Ein reellwertiger stochastischer Prozeß $(c_n)_{n\in\mathbb{N}}$ heißt vorhersagbar bzgl. $(\mathfrak{F}_n)_{n\in\mathbb{N}_0}$, wenn c_n $\mathfrak{F}_{n-1}$-meßbar ist für alle $n \in \mathbb{N}$.

Satz 15.13 Es sei $(c_n)_{n\in\mathbb{N}_0}$ ein vorhersagbarer Prozeß, $(X_n)_{n\in\mathbb{N}_0}$ ein Prozeß mit $E|c_{n+1}(X_{n+1} - X_n)| < \infty$ für alle $n \in \mathbb{N}_0$, $Y_n := X_0 + \sum_{k=1}^{n} c_k(X_k - X_{k-1})$. Dann gilt:

(i) Ist $(X_n)_{n\in\mathbb{N}_0}$ ein Martingal, so ist $(Y_n)_{n\in\mathbb{N}_0}$ ein Martingal.
(ii) Ist $(X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal und ist $(c_n)_{n\in\mathbb{N}_0}$ nicht-negativ, so ist auch $(Y_n)_{n\in\mathbb{N}_0}$ ein Supermatingal.

Beweis: Da c_{n+1} $\mathfrak{F}_n$-meßbar ist, folgt mit Satz 15.5(ii) $E[Y_{n+1} - Y_n|\mathfrak{F}_n] = c_{n+1}E[X_{n+1} - X_n|\mathfrak{F}_n]$, woraus sich beide Behauptungen ergeben. □

Aus einem (un-)fairen Spiel wird also durch variieren der Einsätze kein vorteilhaftes Spiel (wenn man nicht über prophetische Gaben verfügt ...)
Kann man sein Glück dadurch verbessern, daß man im geeigneten Moment aufhört?

Definition 15.14
Eine Abbildung $\tau : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ heißt Stoppzeit (auch Optionszeit) bzgl. der Filtration $(\mathfrak{F}_n)_{n\in\mathbb{N}_0}$, wenn $\{\tau \le n\} \in \mathfrak{F}_n$ für alle $n \in \mathbb{N}_0$ gilt. Für eine Stoppzeit τ ist $\mathfrak{F}_\tau := \{A \subseteq \Omega : A \cap \{\tau \le n\} \in \mathfrak{F}_n \quad \forall n \in \mathbb{N}_0\}$ eine σ-Algebra (Übungsaufgabe), die σ-Algebra der τ-Vergangenheit.
(Interpretation: Die bis einschließlich zur (zufälligen) Zeit τ vorhandene Information.)

Als "optional stopping" bezeichnet man den Übergang von $(X_n)_{n\in\mathbb{N}_0}$ zu $(X_{\tau\wedge n})_{n\in\mathbb{N}_0}$[2], wobei $X_{\tau\wedge n} : \Omega \to \mathbb{R}$, $(X_{\tau\wedge n})(\omega) := X_{\tau(\omega)\wedge n}(\omega)$.
Ist $(\tau_n)_{n\in T_0}$, $T_0 \subseteq \mathbb{N}_0$, eine wachsende Folge von Stoppzeiten, so bezeichnet man den Übergang von $(X_n)_{n\in\mathbb{N}_0}$ zu $(Y_n)_{n\in T_0}$, $Y_n := X_{\tau_n}$ als "optional sampling". Mit τ ist auch $\tau_n := \tau \wedge n$ eine Stoppzeit ($\rightsquigarrow$ Übungsaufgabe), also ist optional stopping ein Spezialfall des optional sampling.

[2] $\wedge$ ist hierbei das Minimum ...

Bei der Glücksspiel-Situation könnte eine mögliche Strategie darin bestehen, bis zum Erreichen eines Mindestgewinnes c zu spielen und dann aufzuhören; dies entspricht dem optional stopping mit $\tau = \inf\{n \in \mathbb{N}_0 : X_n \geq X_0 + c\}$ (wir vereinbaren: $\inf \emptyset = \infty$).

Beispiel 15.15

(i) Konstante τ's sind Stoppzeiten, denn im Falle $\tau(\omega) = n_0 \quad \forall \omega \in \Omega$ gilt
$\{\tau \leq n\} = \Omega$ für $n \geq n_0$
$\{\tau \leq n\} = \emptyset$ für $n < n_0$, man erhält also jedesmal ein Element von $\mathfrak{F}_n$.

(ii) Ist $(Y_n)_{n\in\mathbb{N}_0}$ ein zu $(\mathfrak{F}_n)_{n\in\mathbb{N}_0}$ adaptierter (reellwertiger) Prozeß und A eine Borel-Menge, so ist die Eintrittszeit dieses Prozesses in A $\tau_A(\omega) = \inf\{n \in \mathbb{N}_0 : Y_n(\omega) \in A\}$ (wieder mit $\inf \emptyset = \infty$) eine Stoppzeit, denn $\{\tau_A \leq n\} = \bigcup_{i=0}^{n} \{Y_i \in A\} \in \mathfrak{F}_n$, wegen $\{Y_i \in A\} \in \mathfrak{F}_i$, $(\mathfrak{F}_n)_{n\in\mathbb{N}_0}$ isoton.
Die oben genannte 'Aufhörstrategie' basiert also tatsächlich auf einer Stoppzeit.

Wir wollen nun zeigen, daß unter bestimmten Voraussetzungen auch unter einer optional-sampling-Transformation die (Super-) Martingaleigenschaft erhalten bleibt. Hierzu benötigen wir einige einfache Hilfsaussagen:

Lemma 15.16

(i) Ist τ eine Stoppzeit, so ist $X_\tau^* : \Omega \to \mathbb{R}$, $X_\tau^*(\omega) := \begin{cases} X_{\tau(\omega)}(\omega) & , \text{ wenn } \tau(\omega) \leq \infty \\ 0 & , \text{ sonst} \end{cases}$ $\mathfrak{F}_\tau$-meßbar.

(ii) Sind τ_1, τ_2 Stoppzeiten mit $\tau_1 \leq \tau_2$, so gilt $\mathfrak{F}_{\tau_1} \subseteq \mathfrak{F}_{\tau_2}$.

Beweis: (i) Für jede Borelmenge $A \in \mathfrak{B}$ und jedes $n \in \mathbb{N}_0$ gilt: $\{X_\tau^* \in A\} \cap \{\tau \leq n\} = \bigcup_{k=0}^{n} \{X_k \in A\} \cap \{\tau = k\}$.

Mit $\{\tau = k\} = \underbrace{\{\tau \leq k\}}_{\in \mathfrak{F}_k} \cap \underbrace{\{\tau \leq k-1\}^c}_{\in \mathfrak{F}_{k-1} \subseteq \mathfrak{F}_k}$ folgt $\{\tau = k\} \in \mathfrak{F}_k$, also insgesamt $\{X_\tau^* \in A\} \cap \{\tau \leq n\} \in \mathfrak{F}_n$.

(ii) Sei $A \in \mathfrak{F}_{\tau_1}$. Für alle $n \in \mathbb{N}_0$ gilt: $\{\tau_2 \leq n\} \subseteq \{\tau_1 \leq n\}$, also folgt
$A \cap \{\tau_2 \leq n\} = \underbrace{A \cap \{\tau_1 \leq n\}}_{\in \mathfrak{F}_n, \text{ da } A \in \mathfrak{F}_{\tau_1}} \cap \underbrace{\{\tau_2 \leq n\}}_{\in \mathfrak{F}_n, \text{ da } \tau_2 \text{ Stoppzeit}} \in \mathfrak{F}_n$, d.h. $A \in \mathfrak{F}_{\tau_2}$. □

Ohne weitere Voraussetzungen kann beim Übergang von (X_n) zu (Y_n) die Martingaleigenschaft verloren gehen.

Beispiel 15.17

Es sei $(Y_n)_{n\in\mathbb{N}}$ eine iid-Folge von Zufallsvariablen mit $P(Y_n = -1) = P(Y_n = 1) = \frac{1}{2}$, $X_n := \sum_{i=1}^{n} 2^{i-1} Y_i$,
$\mathfrak{F}_n := \sigma(\{X_1, \ldots, X_n\})$ $(= \sigma(\{Y_1, \ldots, Y_n\}))$, $\tau := \inf\{n \in \mathbb{N} : Y_n = 1\}$.
Dann ist $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}}$ ein Martingal und τ eine Stoppzeit ($\rightsquigarrow$ Beispiel 15.15(ii))

Interpretation: Eine faire Münze wird wiederholt geworfen. Der Spieler setzt 2^{n-1}DM in der n-ten Runde und verliert seinen Einsatz, wenn 'Zahl' erscheint ($Y_n = -1$), bei 'Kopf' ($Y_n = 1$) gewinnt er 2^nDM. Beim ersten Eintreten von 'Kopf' wird das Spiel beendet; X_τ ist der Gesamtgewinn bei Spielende.
Es gilt: $P(\tau > k) = 2^{-k}$, insbesondere $P(\tau < \infty) = 1$, sowie

$$X_\tau = \sum_{k=1}^{\infty} X_k 1_{\{\tau=k\}} = \sum_{k=1}^{\infty} \underbrace{\left(-\sum_{i=1}^{k-1} 2^{i-1} + 2^{k-1}\right)}_{\equiv 1} 1_{\{\tau=k\}} \equiv 1 \quad (P\text{-f.s.})$$

Sei nun $T_0 = \{1, 2\}$, $\tilde{Y}_1 = X_1$, $\tilde{Y}_2 = X_\tau$.
Dann gilt: $E\tilde{Y}_1 = 0$, $E\tilde{Y}_2 = EX_\tau = 1$.
Also ist $(\tilde{Y}_n, \mathfrak{G}_n)_{n\in T_0}$ kein Martingal ($E|X_{\tau-1}| = \infty$.)

Lemma 15.18 Es seien $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}_0}$ ein Supermartingal und σ, τ Stoppzeiten mit $\sigma \leq \tau \leq \sigma + 1 \leq C$ für ein konstantes $C < \infty$. Dann gilt $E|X_\tau| < \infty$, sowie $E[X_\tau | X_\sigma] \leq X_\sigma$.

Beweis: Es gilt $E|X_\tau| = \sum_{k=0}^{C} \int_{\{\tau=k\}} |X_k| dP \leq \sum_{k=0}^{C} E|X_k| < \infty$.

Sei nun $A \in \mathfrak{F}_\sigma$, $A_k := A \cap \{\sigma = k\} \cap \{\tau > \sigma\} = A \cap \{\sigma = k\} \cap \{\tau > k\} = A \cap \{\sigma = k\} \cap \{\tau = k+1\}$.

Dann gilt: $\int_A (X_\sigma - X_\tau) dP = \sum_{k=0}^{C-1} \int_{A_k} (X_\sigma - X_\tau) dP = \sum_{k=0}^{C-1} \int_{A_k} (X_k - X_{k+1}) dP \geq 0$, wenn $A_k \in \mathfrak{F}_k$.

$A_k \in \mathfrak{F}_k$ folgt mit $A \cap \{\sigma = k\} = \underbrace{(A \cap \{\sigma \leq k\})}_{\in \mathfrak{F}_k, \text{ da } \sigma \text{ Stoppzeit}} \cap (\bigcup_{i=1}^{k-1} \underbrace{(A \cap \{\sigma \leq i\})}_{\in \mathfrak{F}_i \subseteq \mathfrak{F}_k})^c$. □

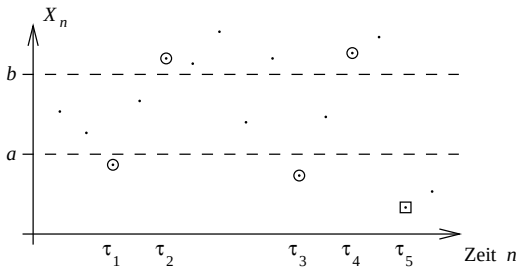
Satz 15.19 (Optional Sampling Theorem)
Es seien $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}_0}$ ein Supermartingal (Submartingal, Martingal) und $(\tau_n)_{n\in\mathbb{N}_0}$ eine isotone Folge von beschränkten Stoppzeiten (d.h. $\forall n \in \mathbb{N}\ \exists C_n < \infty : P(\tau_n \leq C_n) = 1$); $Y_n := X_{\tau_n}$; $\mathfrak{G}_n := \mathfrak{F}_{\tau_n}$. Dann ist $(Y_n, \mathfrak{G}_n)_{n\in\mathbb{N}_0}$ wieder ein Supermartingal (Submartingal, Martingal).
Beweis: Es reicht, den Satz für Supermatingale zu beweisen.
Nach Lemma 15.16(ii) liefert $\mathfrak{G}_n \subseteq \mathfrak{G}_{n+1}\ \forall n \in \mathbb{N}_0$; $(\mathfrak{G}_n)_{n\in\mathbb{N}_0}$ ist also eine Filtration.

Nach Lemma 15.16(i) ist $(Y_n)_{n\in\mathbb{N}_0}$ hierzu adaptiert. Es bleibt zu zeigen: $E[X_{\tau_{n+1}}|\mathfrak{F}_{\tau_n}] \le X_{\tau_n}$ P-f.s. für alle $n \in \mathbb{N}$.
Sei $n \in \mathbb{N}$. Setze $\sigma_k := \tau_{n+1} \wedge (\tau_n + k)$ für $k = 0, \dots, \lceil C_{n+1} \rceil =: m$.
Dann ist $\sigma_0, \dots, \sigma_m$ eine isotone Folge von Stoppzeiten mit $\sigma_0 = \tau_n$, $\sigma_m = \tau_{n+1}$ und $\sigma_{k+1} - \sigma_k \le 1$ für $k = 0, \dots, m-1$.
Lemma 15.18 liefert $E[X_{\sigma_{k+1}}|\mathfrak{F}_{\sigma_k}] \le X_{\sigma_k}$, also:
$E[X_{\tau_{n+1}}|\mathfrak{F}_{\tau_n}] = E[X_{\sigma_m}|\mathfrak{F}_{\sigma_0}] = E[E[X_{\sigma_m}|\mathfrak{F}_{\sigma_{m-1}}]|\mathfrak{F}_{\sigma_0}] \le E[X_{\sigma_{m-1}}|\mathfrak{F}_{\sigma_0}] \le \dots \le E[X_{\sigma_0}|\mathfrak{F}_{\sigma_0}] = X_{\sigma_0} = X_{\tau_n}$ P-f.s. □

15.4 Konvergenzsätze

15.4.1 Fast sichere Konvergenz

Zu einer Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen definieren wie die Größen $U_n[a,b]$, $-\infty < a < b < \infty$, $n \in \mathbb{N}$, wie folgt: $U_n[a,b]$ ist die Anzahl der aufsteigenden Überquerungen des Intervalls $[a,b]$ durch $X_1, X_2, \dots, X_n$, also das größte $k \in \mathbb{N}_0$ mit $\exists 1 \le i_1 < i_2 < i_3 < \dots < i_{2k} \le n$ mit $X_{i_{2j-1}} \le a$, $X_{i_{2j}} \ge b$ für $j = 1, \dots k$. ($U_n[a,b]$ ist eine Zufallsvariable.)



Lemma 15.20 Ist $(X_n)_{n\in\mathbb{N}}$ ein Supermartingal, so gilt: $EU_n[a,b] \le \frac{1}{b-a} E(X_n - a)^-$.

Beweis: Sei $p = \lfloor \frac{n}{2} \rfloor + 1$, $\tau_0 \equiv 0$, und für $k = 1, \dots, p$,

$$\tau_{2k-1} := \begin{cases} \min\{j \ge \tau_{2k-2}, j \le n : X_j \le a\} & , \text{ wenn } \{\dots\} \ne \emptyset \\ n & , \text{ sonst} \end{cases}$$

$$\tau_{2k} := \begin{cases} \min\{j \ge \tau_{2k-1}, j \le n : X_j \ge b\} & , \text{ wenn } \{\dots\} \ne \emptyset \\ n & , \text{ sonst} \end{cases}$$

Dies sind Stoppzeiten, klar: $1 \le \tau_1 \le \tau_2 \le \dots \le \tau_p = n$.
Die aufsteigenden Überquerungen gehören zu den Paaren $(\tau_1, \tau_2), \dots, (\tau_{2k_0-1}, \tau_{2k_0})$ mit $k_0 = U_n[a,b]$, d.h. $X_{\tau_{2k}} - X_{\tau_{2k-1}} \ge b - a$ für $k = 1, \dots, k_0$.
Die hierauf folgende Differenz $X_{\tau_{2k_0+2}} - X_{\tau_{2k_0+1}}$ ist nur dann nicht 0, wenn noch eine a-Unterschreitung (aber keine b-Überschreitung) mehr stattfindet. Sie ist dann von der Form $X_n - X_j$ mit einem $X_j \le a$, ist also $\ge X_n - a \ge \min\{0, X_n - a\} = -(X_n - a)^-$.
Es gilt $(*)$ $\sum_{k=1}^{p} (X_{\tau_{2k}} - X_{\tau_{2k-1}}) \ge (b-a)U_n[a-b] - (X_n - a)^-$.
Das Optional Sampling Theorem (Satz 15.19) liefert $EX_{\tau_{2k}} \le EX_{\tau_{2k-1}}$, also ist der Erwartungswert auf der linken Seite von $(*)$ ≤ 0, d.h. $(b-a)U_n[a,b] - E(X_n - a)^- \le 0$.
(siehe Williams für einen auf Satz 15.13 beruhenden Beweis.) □

Ist $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ eine Filtration, so sei $\mathfrak{F}_\infty := \sigma(\bigcup_{n\in\mathbb{N}} \mathfrak{F}_n)$.

Satz 15.21 (Vorwärts-Konvergenzsatz von Doob)
Es sei $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}}$ ein Supermartingal mit der Eigenschaft $\sup_{n\in\mathbb{N}} E|X_n| < \infty$
(Man nennt das Supermartingal dann $\mathfrak{L}^1$-beschränkt).
Dann existiert eine $\mathfrak{F}_\infty$-meßbare Zufallsvariable X_∞ mit $\lim_{n\to\infty} X_n = X_\infty$ P-fast sicher.

Beweis: Wie zeigen zunächst $P(N) = 0$ für $N := \{\omega \in \Omega : \liminf_{n\to\infty} X_n(\omega) < \limsup_{n\to\infty} X_n(\omega)\}$.
Mit $U_\infty[a,b] := \lim_{n\to\infty} U_n[a,b]$ gilt $N \subset \bigcup_{\substack{a,b\in\mathbb{Q} \\ a<b}} \{\omega \in \Omega : U_\infty[a,b](\omega) = \infty\}$.
Lemma 15.20 liefert $(b-a)EU_n[a,b] \le E(X_n - a)^- \le |a| + E|X_n|$.
Nach Voraussetzung gibt es hierfür eine von n unabhängige Oberschranke, also folgt $EU_\infty[a,b] < \infty$ mit dem Satz von der der majorisierten Konvergenz – insbesondere $P(U_\infty[a,b] = \infty) = 0$.
Damit: N ist in einer abzählbaren Vereiniging von Nullmengen enthalten, also $P(N) = 0$.
Sei $\tilde{X}_\infty(\omega) := \begin{cases} \lim_{n\to\infty} X_n(\omega) & , \quad \omega \in N^c \\ 0 & , \text{ sonst} \end{cases}$ (bestimmte Divergenz ist zugelassen.)
Das Lemma von Fatou liefert $E|\tilde{X}_\infty| = E \liminf |X_n| \le \liminf E|X_n| \le \sup_n E|X_n| < \infty$, d.h. $P(\tilde{N}) = 0$ mit $\tilde{N} = \{\omega \in \Omega : |\tilde{X}_\infty| = \infty\}$.
$X_\infty := 1_{\tilde{N}^c} \cdot \tilde{X}_\infty$ leistet das Verlangte. □

Bemerkung 15.22 $\mathfrak{L}^1$-Beschränktheit folgt bereits aus $\sup_n EX_n^- < \infty$ (Übungsaufgabe 64)

15.4.2 Gleichgradige Integrierbarkeit

Hintergrund: Wann hat man (bei Martingalen) Konvergenz im Mittel (in $\mathfrak{L}^1$).

Definition 15.23 Eine Familie $\{X_i : i \in I\}$ von Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$ heißt gleichgradig integrierbar, wenn gilt:

$\forall \epsilon > 0\ \exists C < \infty\ \forall i \in I : \int\limits_{|X_i|>C} |X_i| dP < \epsilon.$

Beispiel 15.24

(i) Ist X integrierbar (d.h. $E|X| < \infty$), so folgt mit dem Satz von der majorisierten Konvergenz $\lim\limits_{C\to\infty} \int\limits_{|X|>C} |X| dP = 0$, also ist die "Familie" $\{X\}$ gleichgradig integrierbar.
Offensichtlich ist die endliche Vereinigung gleichgradig integrierbarer Familien wieder gleichgradig integrierbar, insbesondere sind alle endlichen Familien von Zufallsvariablen mit endlichen Erwartungswerten gleichgradig integrierbar.

(ii) Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen, Y eine weitere Zufallsvariable mit $|X_n| < Y \quad \forall n$, $EY < \infty$ (Y ist integrierbare Majorante zur Folge $(X_n)_{n\in\mathbb{N}}$), so ist $\{X_n : n \in \mathbb{N}\}$ gleichgradig integrierbar:
$\int\limits_{|X_n|>C} |X_n| dP \le \int\limits_{Y>C} Y dP$ für alle $n \in \mathbb{N}$, die rechte Seite hängt nicht von n ab und geht mit $C \to \infty$ gegen 0 (vergleicht Teil (i)) (geht auch mit beliebigen Familien $\{X_i : i \in I\}$).

(iii) Ist $\{X_i : i \in I\}$ $\mathfrak{L}^p$-beschränkt für ein $p > 1$ (d.h. $\sup\limits_{i\in I} E|X_i|^p < \infty$), so folgt mit
$\int\limits_{|X_i|>K} |X_i| dP \le \int\limits_{|X_i|>K} K^{1-p}|X_i|^p dP \le K^{1-p} \sup\limits_{i\in I} E|X_i|^p$ die gleichgradig Integrierbarkeit der Familie (wegen $p > 1$ geht die rechte Seite, die nicht von i abhängt mit $K \to \infty$ gegen 0). Die analoge Aussage mit $p = 1$ gilt NICHT (Aufgabe 65).

Lemma 15.25 Es sei X ein Zufallsvariable mit $E|X| < \infty$. Dann gilt:

$\forall \epsilon > 0\ \exists \delta > 0\ \forall A \in \mathfrak{A} : \ P(A) < \delta \implies \int\limits_A |X| dP < \epsilon.$

Beweis: Gilt die Aussage nicht, so existiert ein $\epsilon_0 > 0$ und eine Folge $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{A}$ mit
$P(A_n) \le 2^{-n}$, $\int\limits_{A_n} |X| dP \ge \epsilon_0 \quad \forall n \in \mathbb{N}$.

Sei $A = \limsup\limits_{n\to\infty} A_n$. Das Borel-Cantelli-Lemma liefert $P(A) = 0$.

Andererseits folgt mit Fatou $\quad (0 =) \int\limits_A |X| dP = E|X| - \int\limits_{A^c} |X| dP = E|X| - \int \liminf (1_{A_n{}^c} |X|) dP$

$\ge E|X| - \liminf \int\limits_{A_n{}^c} |X| dP = \limsup \int\limits_{A_n} |X| dP \ge \epsilon, \quad \#.$ □

Satz 15.26

Ist X eine Zufallsvariable mit $E|X| < \infty$, so ist $\{E[X|\mathfrak{F}] : \mathfrak{F}$ Unter-σ-Algebra von $\mathfrak{A}\}$ gleichmäßig integrierbar.

Beweis: Sei $\epsilon > 0$. Nach Lemma 15.25 existiert ein $\delta > 0$ mit

$(*) \qquad P(A) < \delta \implies \int\limits_A |X| dP < \epsilon \quad \forall A \in \mathfrak{A}.$

Wähle nun, zu diesem δ ein C mit $\frac{1}{C} E|X| < \delta$. Sei nun $\mathfrak{F}$ irgendeine Unter-σ-Algebra von $\mathfrak{A}$ und Y eine Version von $E[X|\mathfrak{F}]$. Die Jensensche Ungleichung für bedingte Erwartungswerte liefert

$(**) \qquad |Y| \le E\big[|X|\big|\mathfrak{F}\big]$ P-fast sicher und damit $C \cdot P(|Y| \ge C) \le E|Y| \le E|X|$, also $P(|Y| \ge C) \le \frac{1}{C} E|Y| \overset{\text{s.o.}}{\le} \delta.$

Wegen $\{|Y| \ge C\} \in \mathfrak{F}$ folgt insgesamt $\int\limits_{|Y|\ge C} |Y| dP \overset{(**)}{\le} \int\limits_{|Y|\ge C} E\big[|X|\big|\mathfrak{F}\big] dP = \int\limits_{|Y|\ge C} |X| dP \overset{(*)}{\le} \epsilon.$ □

Das nun folgende Hauptresultat dieses Abschnitts verallgemeinert den Satz von der majorisierten Konvergenz (wegen Beispiel 15.24(ii))

Satz 15.27 Es seien $X, X_1, X_2, \ldots$ Zufallsvariablen mit endlichem Erwartungswert. Dann sind äquivalent:

(i) $X_n \to X$ in $\mathfrak{L}^1$ (d.h. $E|X_n - X| \to 0$)

(ii) $X_n \to X$ in Wahrscheinlichkeit und $\{X_n : n \in \mathbb{N}\}$ gleichgradig integrierbar.

Beweis:

(i)$\Longrightarrow$(ii): Wegen $(0 \le) \ \epsilon \cdot P(|X_n - X| > \epsilon) \le E|X_n - X|$ folgt aus der $\mathfrak{L}^1$-Konvergenz die Konvergenz in Wahrscheinlichkeit.
Zum Nachweise der gleichgradigen Integrierbarkeit sei $\epsilon > 0$.
Zu diesem ϵ existiert ein n_0 mit $E|X_n - X| \le \frac{\epsilon}{2} \quad \forall n \ge n_0$.
Lemma 15.25 (n_0-mal angewendet) liefert ein $\delta > 0$ derart, daß für alle $A \in \mathfrak{A}$ mit $P(A) < \delta$ gilt:
$(*) \qquad \int\limits_A |X_n| dP < \epsilon$ für $n = 1, \ldots, n_0 - 1$, $\int\limits_A |X| dP < \frac{\epsilon}{2}$.
Wegen $E|X_n| \le E|X_n - X| + E|X|$ ist die Folge $(X_n)_{n\in\mathbb{N}}$ in $\mathfrak{L}^1$ beschränkt, es existiert also ein $C > 0$ mit $\frac{1}{C} \sup\limits_{n\in\mathbb{N}} E|X_n| < \delta$.
Insbesondere: $P(|X_n| \ge C) < \delta \quad$ für alle $n \in \mathbb{N} \quad (\rightsquigarrow$ Markov-Ungleichung).
Mit $(*)$: $\int\limits_{|X_n|\ge C} |X_n| dP < \epsilon$ für $n = 1, \ldots, n_0 - 1$,

$$\int\limits_{|X_n|\geq C} |X_n|dP \leq \underbrace{\int\limits_{|X_n|\geq C} |X|dP}_{<\frac{\epsilon}{2}} + \underbrace{\int\limits_{|X_n|\geq C} |X_n - X|dP}_{\leq E|X_n - X| \leq \frac{\epsilon}{2}} < \epsilon \quad \text{für } n \geq n_0.$$

Da $\epsilon > 0$ beliebig war, folgt hieraus die behauptete gleichgradige Integrierbarkeit.

(ii)$\Longrightarrow$(i): Für alle $C > 0$ definiere $\varphi_C : \mathbb{R} \to \mathbb{R}$ durch $\varphi_C(x) := \begin{cases} C & , \quad x > C \\ x & , \quad |x| \leq C \\ -C & , \quad x < -C \end{cases}$

Zwischenbehauptung: $(*)$ $\lim\limits_{n\to\infty} |\varphi_C(X_n) - \varphi_C(X)| = 0$ für alle $C > 0$.

Sei also $\epsilon > 0$. Man hat $|\varphi_C(x) - \varphi_C(y)| \leq |x - y|$, also

$E|\varphi_C(X_n) - \varphi_C(X)| = \int\limits_{|X_n - X|>\epsilon} |\varphi_C(X_n) - \varphi_C(X)|dP + \int\limits_{|X_n - X|\leq\epsilon} |\varphi_C(X_n) - \varphi_C(X)|dP \leq$

$2CP(|X_n - X| > \epsilon) + \epsilon \overset{n\to\infty}{\longrightarrow} \epsilon$, also folgt $(*)$.

Sei wieder $\epsilon > 0$. Der Satz von der majorisierten Konvergenz liefert $E|\varphi_C(X) - X| \to 0$ mit $C \to \infty$. Hiermit und mit der gleichgradigen Integrierbarkeit von $\{X_n : n \in \mathbb{N}\}$ erhält man ein $C > 0$ $(C < \infty)$ mit: $E|\varphi_C(X) - X| \leq \frac{\epsilon}{3}$.

$E|\varphi_C(X_n) - X_n|$ $(\leq \int\limits_{|X_n|\geq C} |X_n|dP)$ $< \frac{\epsilon}{3}$ $\forall n \in \mathbb{N}$

und damit $E|X_n - X| \leq E|X_n - \varphi_C(X_n)| + E|\varphi_C(X_n) - \varphi_C(X)| + E|\varphi_C(X) - X| \leq \frac{2}{3}\epsilon + E|\varphi_C(X_n) - \varphi_C(X)|$.

Nach $(*)$ existiert ein n_0 mit: $E|\varphi_C(X_n) - \varphi_C(X)| < \frac{\epsilon}{3}$ $\forall n \geq n_0$. □

15.4.3 Gleichgradig integrierbare Martingale

Es sei wieder $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ eine Filtration und $\mathfrak{F}_\infty := \sigma(\bigcup\limits_{n=1}^{\infty} \mathfrak{F}_n)$.

Klar: $(\mathfrak{F}_n)_{n\in\mathbb{N}\cup\{\infty\}}$ ist wieder eine Filtration.

Satz 15.28 Es sei $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}}$ ein gleichgradig integrierbares Supermartingal. Dann existiert eine $\mathfrak{F}_\infty$-meßbare Zufallsvariable X_∞ mit $X_n \to X_\infty$ P-f.s. und in $\mathfrak{L}^1$ mit $n \to \infty$, sowie $X_n \geq E[X_\infty|\mathfrak{F}_n]$ $\forall n \in \mathbb{N}$. (auch $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}\cup\{\infty\}}$ ist also ein Supermartingal.)

Beweis:

Gleichgradig integrierbare Funktionen sind $\mathfrak{L}^1$-beschränkt ($\rightsquigarrow$ Aufgabe 65), also läßt sich Satz 15.21 anwenden und liefert ein X_∞ mit $X_n \to X_\infty$ P-f.s.

Hieraus folgt $X_n \to X_\infty$ in Wahrscheinlichkeit (Satz 10.2), also mit Satz 15.27 $E|X_n - X_\infty| \to 0$, d.h. die $\mathfrak{L}^1$-Konvergenz.

Sei nun $A \in \mathfrak{F}_n$. Die Supermaringaleigenschaft liefert $\int\limits_A X_n dP \geq \int\limits_A X_m dP$ $\forall m \geq n$.

Wegen $|\int\limits_A X_m dP - \int\limits_A X_\infty dP| \leq E|X_m - X_\infty| \overset{m\to\infty}{\longrightarrow} 0$ folgt $\int\limits_A X_n dP \geq \int\limits_A X_\infty dP$ $\forall A \in \mathfrak{F}_n$, d.h. $X_n \geq E[X_\infty|\mathfrak{F}_n]$ ($\rightsquigarrow$ Aufgabe 54) □

Sei nun T eine beliebige gerichtete Indexmenge.

Ist $(\mathfrak{F}_t)_{t\in T}$ eine Filtration und X eine Zufallsvariable mit $E|X| < \infty$, so ist $(X_t, \mathfrak{F}_t)_{t\in T}$ mit $X_t := E[X|\mathfrak{F}_t]$ ein gleichgradig integrierbares Martingal ($\rightsquigarrow$ Beispiel 15.9(iii), Satz 15.26)

Satz 15.29 Es sei $(X_t, \mathfrak{F}_t)_{t\in T}$ ein gleichgradig integrierbares Martingal, $\mathfrak{F}_\infty := \sigma(\bigcup\limits_{t\in T} \mathfrak{F}_t)$.

Dann existiert eine $\mathfrak{F}_\infty$-meßbare Zufallsvariable X_∞ mit $X_t = E[X_\infty|\mathfrak{F}_t]$ (P-f.s.) für alle $t \in T$.

Beweis: Wir zeigen zunächst:

(1) $\forall \epsilon > 0\ \exists t(\epsilon) \in T\ \forall s, t \in T : s \geq t(\epsilon), t \geq t(\epsilon) \implies E|X_s - X_t| < \epsilon$ ($(X_t)_{t\in T}$ ist eine Cauchy-Folge "entlang T")

Angenommen, (1) ist falsch. Dann existiert $\epsilon_0 > 0$ mit

(2) $\forall t \in T\ \exists s, u \in T, s \geq t, u \geq t$ mit $E|X_s - X_u| \geq 2\epsilon_0$.

Wir können dazu rekursiv eine isotone Folge $(t_n)_{n\in\mathbb{N}} \subset T$ konstruieren mit:

(3) $E|X_{t_{n+1}} - X_{t_n}| \geq \epsilon_0$ $\forall n \in \mathbb{N}$.

Sei hierzu $t_1 \in T$ beliebig. Sind $t_1, \ldots, t_n$ bereits konstruiert, so existieren nach (2) $s, u \geq t_n$ mit $E|X_s - X_u| \geq 2\epsilon_0$. Nach der Dreiecksungleichung muß dann $E|X_s - X_{t_n}| \geq \epsilon_0$ oder $E|X_u - X_{t_n}| \geq \epsilon_0$ gelten. Wähle entsprechend $t_{n+1} = s$ oder $t_{n+1} = u$.

Nun ist $(X_{t_n}, \mathfrak{F}_{t_n})_{n\in\mathbb{N}}$ ein gleichgradig integrierbares Martingal, konvergiert nach Satz 15.28 im Widerspruch zu (3) im Mittel. Damit ist (1) bewiesen.

Wähle $\epsilon = \frac{1}{n}$ in (1), die Folge $t(\frac{1}{n})$, $n \in \mathbb{N}$, kann als isoton vorausgesetzt werden (hierbei wird "T gerichtet" gebrauchtet.) Dann liefert (1)

(4) $E|X_t - X_{t(\frac{1}{n})}| \leq \frac{1}{m}$ für alle $m \in \mathbb{N}$, $n \geq m$, $t \geq t(\frac{1}{m})$.

Abermalige Anwendung von Satz 15.28 liefert eine Zufallsvariable $X \in \mathfrak{L}^1$, gegen die $X_{t(\frac{1}{n})}$ mit $n \to \infty$ im Mittel konvergiert. Läßt man in (4) $n \to \infty$ gehen, so folgt:

(5) $E|X_t - X| \leq \frac{1}{m}$ für alle $m \in \mathbb{N},\ t \geq t(\frac{1}{m})$ (X_t konvergiert gegen X "entlang T")
Da $\mathfrak{L}^1$-Konvergenz Konvergenz in Wahrscheinlichkeit impliziert, und da zu in W. konvergenten Folgen f.s. konvergente Teilfolgen existieren ($\rightsquigarrow$ Aufgabe 31), gibt es eine Folge $(m_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$ mit $X_{t(\frac{1}{m_n})} \to X$ P-f.s. Setze nun $X_\infty(\omega) = \begin{cases} X(\omega) & , \text{ wenn } \lim\limits_{n\to\infty} X_{t(\frac{1}{m_n})} \text{ existiert} \\ 0 & , \text{ sonst} \end{cases}$
Dann ist X_∞ $\mathfrak{F}_\infty$-meßbar, da alle $X_{t(\frac{1}{m_n})}$ $\mathfrak{F}_\infty$-meßbar sind. Es bleibt zu zeigen, daß alle $t \in T$ X_t eine Version von $E[X_\infty|\mathfrak{F}_t]$ ist. Sei hierzu $t \in T$, $A \in \mathfrak{F}_t$. Die Martingaleigenschaft liefert
(6) $\int\limits_A X_t dP = \int\limits_A X_s dP$ für alle $s \geq t$.
Zu jedem $m \in \mathbb{N}$ existiert ein $s \geq t(\frac{1}{m})$, $s \geq t$, also folgt mit (6) und (5)
$|\int\limits_A X_t dP - \int\limits_A X_\infty dP| \overset{P(X=X_\infty)=1}{=} |\int\limits_A X_t dP - \int\limits_A X dP| \leq E|X_s - X| \leq \frac{1}{m}$ für alle $m \in \mathbb{N}$.
Mit $m \to \infty$ folgt $\int\limits_A X_t dP = \int\limits_A X_\infty dP \quad \forall A \in \mathfrak{F}_t$. □

15.4.4 Ein Rückwärtskonvergenzsatz

Jetzt: Indexmenge $T = -\mathbb{N}$, Grenzübergang $n \to -\infty$.

Satz 15.30 (Rückwärtskonvergenzsatz von Doob)
Es sei $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}}$ ein Martingal, $\mathfrak{F}_{-\infty} := \bigcap\limits_{n=1}^{\infty} \mathfrak{F}_{-n}$. Dann gilt mit $n \to -\infty$:
$X_n \to X_{-\infty} := E[X_{-1}|\mathfrak{F}_{-\infty}]$ P-f.s und im Mittel.
Beweis: Es sei $U_{-n}[a,b]$ die Anzahl der aufsteigenden Überquerungen von $[a,b]$ durch $X_{-n}, \ldots, X_{-1}$, $U_{-\infty}[a,b] = \lim\limits_{n\to\infty} U_{-n}[a,b]$.
Lemma 15.20 liefert $EU_{-n}[a,b] \leq \frac{1}{b-a} E(X_{-1} - a)^-$ (unabhängig von n)
Also: monotone Konvergenz ergibt $EU_{-\infty}[a,b] < \infty$, und damit $P(\exists a,b \in \mathbb{Q} : U_{-\infty}[a,b] = \infty) = 0$.
Dies liefert die Existenz eines (möglicherweise $\pm\infty$-wertigen) f.s. Grenzwertes $\tilde{X}_{-\infty}$.
Mit dem Lemma von Fatou folgt $E|\tilde{X}_{-\infty}| = E \liminf\limits_{n\to\infty} |X_{-n}| \leq \liminf\limits_{n\to\infty} E|X_{-n}| \leq E|X_{-1}| < \infty$, da $(X_{-n}, \mathfrak{F}_{-n})_{n\in\mathbb{N}}$ ein Submartingal ist (Satz 15.11.)
Hieraus folgt $P(|X_{-\infty}| = \infty) = 0$; Abänderung auf einer Nullmenge liefert eine $\mathbb{R}$-wertige $\mathfrak{F}_{-\infty}$-meßbare Zufallsvariable $X_{-\infty}$ mit $X_{-n} \to X_{-\infty}$ f.s.; $\mathfrak{L}^1$-Konvergenz folgt mit Satz 15.26 und Satz 15.27.
Dies wiederum kann wie im Beweis zu Satz 15.28 verwendet werden, um zu zeigen, daß $X_{-\infty}$ eine Version von $E[X_{-1}|\mathfrak{F}_{-\infty}]$ ist. □

15.5 Anwendungen

15.5.1 Ein 0-1-Gesetz von Kolmogorov

Als 0-1-Gesetze bezeichnet man Aussagen, die für die Wahrscheinlichkeit bestimmter Ereignisse nur die "Extremfälle" 0 und 1 zulassen.

Satz 15.31 Es seien $X_1, X_2, \ldots$ unabhängige Zufallsvariablen. $\mathfrak{T}_n := \sigma\{X_m : m \geq n\}$, $\mathfrak{T} := \bigcap\limits_{n=1}^{\infty} \mathfrak{T}_n$
($\mathfrak{T}$ heißt auch "σ-Algebra der terminalen Ereignisse".) Dann gilt $P(A) = 0$ oder $P(A) = 1$ für alle $A \in \mathfrak{T}$.
Beweis: Sei $A \in \mathfrak{T}$. Setze $\mathfrak{F}_n := \sigma\{X_1, \ldots, X_n\}$, $Y := 1_A$, $Y_n := E[Y|\mathfrak{F}_n]$.
Wegen $|Y_n| < 1$ (Jensen) läßt sich Satz 15.28 anwenden: Es gibt eine $\mathfrak{F}_\infty$-meßbare Zufallsvariable Y_∞ mit $Y_n \to Y_\infty$ P-f.s. und im Mittel, und es gilt $Y_n = E[Y_\infty|\mathfrak{F}_n]$ für alle $n \in \mathbb{N}$.
Hiermit $\int\limits_B Y_\infty dP = \int\limits_B Y_n dP = \int\limits_B Y dP$ für alle $B \in \mathfrak{F}$.
Man kann $Y_\infty \geq 0$ annehmen (Monotonie des bedingten Erwartungswertes)
Die Maße $B \to \int\limits_B Y_\infty dP$, $B \to \int\limits_B Y dP$ stimmen also auf dem $\cap$-stabilen Erzeuger $\bigcup\limits_{n=1}^{\infty} \mathfrak{F}_n$ von $\mathfrak{F}_\infty$ und damit auf $\mathfrak{F}_\infty$ überein. Wegen $\mathfrak{T}_n \subseteq \mathfrak{F}_\infty$ $\forall n$ gilt $\mathfrak{T} \subseteq \mathfrak{F}_\infty$, also ist Y $\mathfrak{F}_\infty$-meßbar.
Der Eindeutigkeitssatz für Dichten führt also auf $P(Y = Y_\infty) = 1$.
Andererseits ist $\mathfrak{T}$ als Teilmenge von $\mathfrak{T}_{n+1}$ von $\mathfrak{F}_n$ unabhängig, und damit $Y_n = E[1_A|\mathfrak{F}_n] = E1_A = P(A)$ P-f.s.
Insgesamt: $P(\{\omega \in \Omega : \lim\limits_{n\to\infty} P(A) = 1_A(\omega)\}) = 1$, d.h. $P(A) = 0$ oder $P(A) = 1$. □

Beispiel 15.32 Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen, $\overline{X_n} := \frac{1}{n} \sum\limits_{i=1}^{n} X_i$. Offensichtlich ändern sich die Häufungspunkte der Folge $(\overline{X_n})_{n\in\mathbb{N}}$ nicht, wenn man beliebig viele Anfangsglieder $X_1, \ldots, X_n$ verändert.
Dies bedeutet, daß $Y := \limsup\limits_{n\to\infty} \overline{X_n}$ $\mathfrak{T}_n$-meßbar ist für alle $n \in \mathbb{N}$, also sogar $\mathfrak{T}$-meßbar.
Nach Satz 15.31 gilt $P(Y \in A) \in \{0,1\}$ $\forall A$, d.h. es existiert eine Konstante $C \in [-\infty, \infty]$ mit $P(\limsup \overline{X_n} = C) = 1$.

Analog: Auch der lim inf ist fast sicher konvergent.
Hieraus folgt insgesamt: Konvergiert $\overline{X_n}$ f.s. gegen eine Zufallsvariable $\overline{X_\infty}$, so ist $\overline{X_\infty}$ P-f.s. konstant.

15.5.2 Das starke Gesetz der großen Zahlen

Ein Martingalbeweis für Satz 10.4:

Satz 15.33 Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit (existierendem) Erwartungswert μ, so gilt $\overline{X_n} \to \mu$ P-f.s. $(\frac{1}{n}S_n \to \mu$ f.s.)

Beweis:

Es sei $\mathfrak{F}_{-n} := \sigma\{S_m : m \geq n\}$ $\forall n$, wobei $S_m := X_1 + \ldots + X_m$, $\mathfrak{F}_{-\infty} := \bigcap_{n=1}^{\infty} \mathfrak{F}_{-n}$. Aus Beispiel 15.7(ii) ist $E[X_1|\mathfrak{F}_{-n}] = \overline{X_n}$, d.h. $(Y_n, \mathfrak{F}_n)_{n\in-\mathbb{N}}$ mit $Y_n := \overline{X_{-n}}$ ist ein Martingal (siehe auch Beispiel 15.9(iii)).
Satz 15.30 liefert die fast sichere Konvergenz gegen eine (f.s. endliche) $\mathfrak{F}_{-\infty}$-meßbare Zufallsvariable $X_{-\infty}$.
In der Notation von §15.5.1 gilt $\mathfrak{F}_{-\infty} = \mathfrak{T}$, also folgt mit Satz 15.31 und Beispiel 15.32 $Y_{-\infty} \equiv c$ P-f.s. mit einem $c \in \mathbb{R}$. Da außerdem Konvergenz in $\mathfrak{L}^1$ gilt, hat man $E\overline{X_n} \to EY_{-\infty}$, also $c = \mu$. □

15.5.3 Der Satz von Radon-Nikodym

Wir beschränken uns auf "normierte" Maße (WMaße). Der allgemeine Fall kann hieraus leicht hergeleitet werden.

Satz 15.34 Es sei $(\Omega, \mathfrak{A}, P)$ ein WRaum und Q ein weiteres WMaß auf $(\Omega, \mathfrak{A})$ mit $Q \ll P$ (Q wird von P dominiert, d.h. $P(A) = 0 \implies Q(A) = 0$ $\forall A \in \mathfrak{A}$). Dann existiert eine $\mathfrak{A}$-meßbare Funktion $f : \Omega \to \mathbb{R}_+$ mit $Q(A) := \int_A f dP$ $\forall A \in \mathfrak{A}$ (In Worten: Q hat eine Dichte bzgl. P.)

Beweis: Wir benötigen eine Hilfsaussage, die ganz ähnlich wie Lemma 15.25 bewiesen wird:

$(*)$ $\forall \epsilon > 0\ \exists \delta > 0\ \forall A \in \mathfrak{A} : P(A) < \delta \implies Q(A) < \epsilon$.

Wir verwenden die Konstruktion aus Beispiel 15.9(ii):
T ist die Menge der (endlichen, meßbaren) Zerlegungen $t = \{A_1, \ldots, A_n\}$, $A_1, \ldots, A_n \in \mathfrak{A}$, $A_i \cap A_j = \emptyset$ für $i \neq j$, $\bigcup_{i=1}^n A_i = \Omega$; $\mathfrak{F}_t := \sigma(\{A_1, \ldots, A_n\})$, $X_t = \sum_{i=1}^n \frac{Q(A_i)}{P(A_i)} 1_{A_i}$ (Dichte von $Q|_{\mathfrak{F}_t}$ bzgl. $P|_{\mathfrak{F}_t}$.)
Bereits bekannt: $(X_t, \mathfrak{F}_t)_{t\in T}$ ist ein nicht-negatives Martingal. Jedes $B \in \mathfrak{F}_t$ ist endliche disjunkte Vereinigung von A_i-Mengen, insbesondere

$(**)$ $\int_B X_i dP = \sum_{i:A_i\subseteq B} \frac{Q(A_i)}{P(A_i)} P(A_i \cap B) = \sum_{A_i\subseteq B} Q(A_i) = Q(B)$.

Speziell: $E_P X_t = 1$, $\int_{X_t > C} X_t dP = Q(X_t > C)$ $\forall C > 0$.

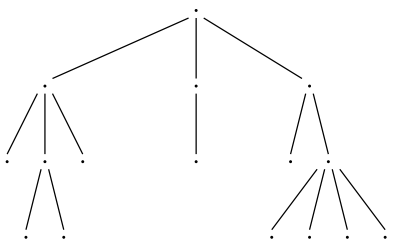
Wir wollen nun zeigen, daß $\{X_t : t \in T\}$ gleichgradig integrierbar ist. Sei hierzu $\epsilon > 0$. Wähle δ gemäß $(*)$ und $C > \delta^{-1}$. Dann gilt $P(X_t > C) \leq \frac{1}{C} E_P X_t < \delta$, also $\int_{X_t > C} X_t dP = Q(X_t > C) < \epsilon$ wegen $(*)$.

Nach Satz 15.29 existiert nun eine $\mathfrak{F}_\infty$-meßbare Zufallsvariable X_∞ mit

$(***)$ $X_t = E[X_\infty|\mathfrak{F}_t]$ $\forall t \in T$.

Klar: $\mathfrak{F}_\infty \subseteq \mathfrak{A}$. Betrachte, zu gegebenem $A \in \mathfrak{A}$, die Zerlegung $t = \{A, A^c\}$. Dies zeigt $\mathfrak{A} \subset \mathfrak{F}_\infty$, also $\mathfrak{A} = \mathfrak{F}_\infty$; außerdem folgt mit $(**)$ und $(***)$ $Q(A) = \int_A X_t dP = \int_A X_\infty dP$, also leistet $f := X_\infty$ das Verlangte. □

15.5.4 Verzweigungsprozesse



$\{X_{nj} : n \in \mathbb{N}, j \in \mathbb{N}\}$ unabhängige $\mathbb{N}_0$-wertige Zufallsvariablen, $p_k := P(X_{nj} = k)$ für $k \in \mathbb{N}_0$, $n, j \in \mathbb{N}$.
Interpretation: X_{nj} ist die Anzahl der Nachkommen vom Mitglied j der Generation n. Der Verzweigungsprozeß $(Z_n)_{n\in\mathbb{N}_0}$ wird rekursiv definiert durch

$Z_0 := 1$, $Z_{n+1} := \sum_{k=1}^{Z_n} X_{n+1,k}$ für $n = 1, 2, \ldots$

Anwendungen: überleben von Familiennahmen (19. Jh.)
Kernspaltung (20. Jh.)

Mittlere Familiengröße $\mu := EX_{nj} = \sum_{k=0}^{\infty} kp_k < \infty$.

Mit welcher Wahrscheinleichkeit stirbt ein solcher Prozeß aus?
Klar: $Z_n = 0 \implies Z_k = 0$ $\forall k \geq n$
$A_n := \{Z_n = 0\}$, $q_n := P(A_n)$
$A_n \uparrow A := \{\exists n : Z_n = 0\}$ ("Prozeß stirbt aus")
$q_n \uparrow q := P(A)$ Frage: $q =$?
Der Fall $p_0 = 0$ liefert $q = 0$, und wird nun ausgeschlossen.

Wahrscheinlichkeitserzeugende Funktion: Y sei eine $\mathbb{N}_0$-wertige Zufallsvariable.

Dann heißt $g(t) := Et^Y = \sum_{k=0}^{\infty} t^k P(Y=k)$ die wahrscheinlichkeitserzeugende Funktion zu Y

(bzw. zur Verteilung von Y.)

Man hat $g_{X+Y}(t) = g_X(t) g_Y(t)$, wenn X, Y unabhängig

$g'(1) = EX$ etc.

Ziel: "Darstellung" der Aussterbewahrscheinlichkeit q über die wahrscheinlichkeitserzeugende Funktion $g(t) = \sum t^k p_k$ der Nachkommenverteilung.

Lemma 15.35 Es sei g_n die wahrscheinlichkeitserzeugende Funktion zu Z_n.

Dann gilt $g_n = g \circ g \circ \ldots \circ g$ (n-fache Hintereinanderschaltung)

Beweis: $Z_1 = X_{11}$, die Behauptung gilt also für $n=1$. Gilt sie für ein $n \in \mathbb{N}$, so folgt

$$g_{n+1}(t) = Et^{Z_{n+1}} = \sum_{j=0}^{\infty} E \underbrace{t^{\sum_{k=1}^{j} X_{n+1,k}}}_{} \underbrace{1_{\{Z_n=j\}}}_{} = \sum_{j=0}^{\infty} g(t)^j P(Z_n = j) = g_n(g(t)). \qquad \square$$

Satz 15.36 Die Aussterbewahrscheinlickeit q ist die kleinste nicht-negative Lösung der Gleichung $g(t) = t$.

Beweis: Sei wieder $q_n := P(Z_n = 0)$. Mit Lemma 15.35 folgt $q_n = g_n(0) = g(g_{n-1}(0)) = g(q_{n-1})$.

Da g auf $[0,1]$ stetig ist, liefert der Grenzübergang $n \to \infty$ die Gleichung $q = g(q)$.

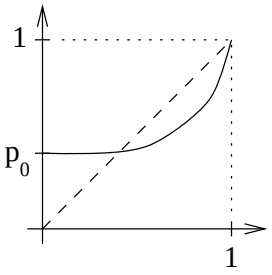
Ist andererseits r irgendeine nicht-negative Lösung von $g(t) = t$, so folgt, da g schwach monoton wachsend ist,

$q_1 = g(0) \le g(r) = r$

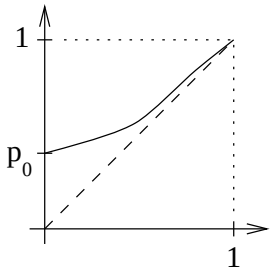
$q_2 = g(q_1) \le g(r) = r$, also $g_n \le r \quad \forall n \in \mathbb{N}$ und damit $g \le r$. $\square$

Korollar 15.37 Es sei p_0 $(= P(X_{nk} = 0)) > 0$. Dann gilt: $q = 1 \iff \mu \le 1$.

Beweis: $g''(t) = \sum_{k=2}^{\infty} p_k k(k-1) t^{k-2} \ge 0$ für $0 \le t \le 1$. Also ist g auf $[0,1]$ konvex.



Fall $\mu > 1$



Fall $\mu \le 1$

(Formaler Beweis: $\rightsquigarrow$ Krengel) $\square$

Bei Modellen dieser Art ist die Suche nach zugehörigen Martingalen oft lohnend:

Setzt man $\mathfrak{F}_n := \sigma\{Z_1, \ldots, Z_n\}$, so erhält man

$$E[Z_{n+1}|\mathfrak{F}_n] = E[\sum_{j=1}^{Z_n} X_{n+1,j}|\mathfrak{F}_n] = E[\sum_{k=1}^{\infty}(\sum_{j=1}^{k} X_{n+1,j}) 1_{\{Z_n=k\}}|\mathfrak{F}_n] \overset{\text{monotone Konvergenz}}{=} \sum_{k=1}^{\infty} E[(\sum_{j=1}^{k} X_{n+1,j}) 1_{\{Z_n=k\}}|\mathfrak{F}_n] \overset{15.5(ii)\ 15.5(iii)}{=}$$

$$\sum_{k=0}^{\infty} E(\underbrace{\sum_{j=1}^{k} X_{n+1,j}}_{k\mu}) 1_{\{Z_n=k\}} = \sum_{k=0}^{\infty} k\mu 1_{\{Z_k=k\}} = \mu Z_n, \text{ also ist } (Y_n, \mathfrak{F}_n)_{n\in\mathbb{N}} \text{ mit } \boxed{Y_n := \mu^{-n} Z_n} \text{ ein Martingal.}$$

Mit $Z_n \ge 0$ folgt $E|Y_n| = EY_n = EY_1 = EX_{11} = \mu$, also ist das Martingal $\mathfrak{L}^1$-beschränkt und mit Satz 15.21 folgt:

Satz 15.38 Es gibt eine Zufallsvariable Y_∞ mit $\frac{1}{\mu^n} Z_n \to Z_\infty$ P-f.s.

Bei $\mu \le 1$ hat man wegen $P(Z_n = 0) \to q = 1$ natürlich $Y_\infty \equiv 0$ P-f.s.

Dies liefert ein Beispiel für ein Martingal, das fast sicher, aber nicht im Mittel konvergiert: $EY_n \equiv \mu \not\to 0 = EY_\infty$.

Index

Symbols

σ-Algebra, 1, 2
    der teminalen Ereignisse, 82
    Erzeuger, 19
σ-endlich, 47

A

Alternative, 36
Axiome, 2

C

charakteristische Funktion, 69

D

Dichte, 24, 45
Diracmaß, 2
Dynkin-System, 20

E

Einpunktmaß, 2
Ereignis, 1
Ereignisraum, 1
Erwartungswert, 13, 45
    bedingter, 15

F

Faltung, 57
Filtration, 75
    natürlich, 76

G

Gesetz der großen Zahlen
    schwach, 17, 51
    stark, 52

H

Hypothese, 36

I

Indikatorfunktion, 14
integrierbar
    gleichgradig, 80

K

Konfidenz
    Bereich, 38
    Intervall, 68
    Niveau, 38
Konsistenz, 36
Konvergenz
    entlang T, 81
    fast sicher, 51
    in Verteilung, 29
    in Wahrscheinlichkeit, 51
    schwach, 29
Korrelation, 16
Kovarianz, 16, 49
    -matrix, 69

L

Laplace-Experiment, 2, 3, 7
Law, 11
Lebesgue-Integral
    Konstruktion, 41
    Lemma von Fatou, 44, 74
    Majorisierte Konvergenz, 44, 74
    Monotone Konvergenz, 44, 74
    Transformationsformel, 43
limes
    superior, 1

M

Martingal, 75
Massenfunktion, 11
Maximum-Likelihood, 33
Menge
    gerichtet, 75

O

optional sampling, 77
optional stopping, 77
Optionszeit, 77

P

p-Wert, 38
Partialsumme
    standardisiert, 31
Probleme
    Bridge, 9, 15
    Geburtstage, 9
    Paradox von de Méré, 8
    Postbote, 10
Produktmaß, 47
Prozeß
    stochastisch, 75
    vorhersagbar, 77

Q

Quantil, 23

S

Schätzer
    Likeliehood-Fonktion, 67
    log-Likelihood-Funktion, 67
    ML, 67
Standardabweichung, 14, 45

Stetigkeitskorrektur, 31
Stichproben
    -mittelwert, 34
    -raum, 33
    -varianz, 34
Stochastik, 1
Stoppzeit, 77
straff, 60

U

Unabhängigkeit, 6
Ungleichungen
    Cauchy-Schwarz, 16, 46
    Chebychev, 17, 46
    Hölder, 46
    Informations-, 35
    Jensen, 46, 74
    Markov, 17, 46
    Minkovski, 46

V

Varianz, 14, 45
Verteiltung
    absolutstetige, 48
Verteilung, 11, 21
    bedingte, 15
    diskret
        bernoulli, 12
        binomial, 11
        geometrisch, 12
        hypergeometrisch, 12
        multinomial, 13
        negativ-binomial, 12
        poisson, 12
    gemeinsam, 49
    kontinuierlich
        Chiquadrat, 70
        exponential, 25
        Gamma, 25
        Gleich, 24
        logistisch, 67
        normal, 25
        Rechteck, 24
        t, 70
    Rand-, 14
Verteilungsfunktion
    empirisch, 40
Verzweigungsprozeß, 83

W

Wahrscheinlichkeit
    bedingte, 5
Wharscheinlichkeitserzeugende Funktion, 84
WMaß, 2
    Dichte, 24
    dominiert, 83

Z

ZGWS
    Lindeberg-Bedingung, 64
    Lyapunov-Bedingung, 65
Zufallsexperiment, 1
Zufallsgröße, 11