Stochastik I
Gliederung zur Vorlesung im Sommersemester 2016

Markus Reiß
Humboldt-Universität zu Berlin

Vorläufige Version vom 14. September 2016

Inhaltsverzeichnis

Markus Reiß

Vorlesung
Stochastik I
Sommersemester 2016

Ein paar Literaturempfehlungen

- Hans-Otto Georgii, *Stochastik*, de Gruyter: exzellentes Lehrbuch inkl. Maßtheorie
- Ulrich Krengel, *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, Vieweg: Klassiker mit vielen Beispielen und Diskussionen, ohne Maßtheorie
- Herold Dehling, Beate Haupt, *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, Springer: Lehrbuch mit vielen erklärenden Skizzen und Diagrammen, ohne Maßtheorie
- William Feller, *An introduction to probability theory and its applications I*, Wiley: das alte Testament, eine Fundgrube, immer noch Standardreferenz
- Kai Lai Chung, *A Course in Probability Theory*, Academic Press: Englisch-sprachiges Standardwerk, besonders empfehlenswert für charakteristische Funktionen und Konvergenzresultate
- Richard Dudley, *Real Analysis and Probability*, Cambridge University Press: ausgezeichnetes und recht anspruchsvolles Lehrbuch zu Maßtheorie, Analysis und W-Theorie, insbesondere für Konvergenzarten und charakteristische Funktionen
- Achim Klenke, *Wahrscheinlichkeitstheorie*, Springer: Lehrbuch für Stochastik I und II, aus Vorlesungen entstanden
- Jürgen Elstrodt, *Maß- und Integrationstheorie*, Springer: mit viel Liebe und historischen Anmerkungen verfasstes, ausführliches Maßtheoriebuch
- Heinz Bauer, *Wahrscheinlichkeitstheorie*, de Gruyter: umfassendes deutsches Standardwerk, auf dem Maßtheoriebuch des Autors aufbauend
- Albert N. Shiryaev, *Probability*, Springer: umfassendes Lehrbuch, gut als Nachschlagewerk für Stochastik I und II
- Jean Jacod, Philip Protter, *Probability Essentials*, Springer: alle wichtigen Ergebnisse auf hohem Niveau, kurz und knapp
- John A. Rice, *Mathematical Statistics and Data Analysis*, Thomson: gutes einführendes Lehrbuch in die mathematische Statistik, viele Beispiele
- Jun Shao, *Mathematical Statistics*, Springer: deckt weite Themen der math. Statistik ab, gut für den Überblick und zum Nachschlagen

1 Wahrscheinlichkeitsräume

1.1 Ereignisse, Wahrscheinlichkeiten und Zufallsvariablen

1.1 Beispiele.

(a) Würfeln mit zwei unterscheidbaren Würfeln wird durch die Grundmenge $\Omega = \{1,2,3,4,5,6\}^2$ beschrieben. Interpretation ist, dass Versuchsausgang $(n_1, n_2) \in \Omega$ bedeutet Augenzahl des 1. Würfels $= n_1$, des 2. Würfels $= n_2$. Ereignisse sind z.B. $A_1 =$*es wird ein Pasch gewürfelt*, $A_2 =$*die Augensumme ist 7*, $A_3 =$*es tritt keine 1 auf*. Ereignisse werden als Teilmengen von Ω modelliert: $A_1 = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$, $A_2 = \{(n_1,n_2) \in \Omega \,|\, n_1 + n_2 = 7\}$, $A_3 = \{(n_1,n_2) \in \Omega \,|\, n_1 \neq 1 \text{ und } n_2 \neq 1\}$. Sind alle Versuchsausgänge gleichwahrscheinlich, so ist die Wahrscheinlichkeit eines Ereignisses $A \subseteq \Omega$ gleich der Anzahl aller für A günstiger Versuchsausgänge geteilt durch die Anzahl aller möglichen Versuchsausgänge, d.h. $P(A) = \frac{|A|}{|\Omega|}$. Wir erhalten $P(A_1) = \frac{1}{6}$, $P(A_2) = \frac{1}{6}$, $P(A_3) = \frac{25}{36}$.

(b) Beim n-fachen Wurf einer Münze setzen wir $\Omega = \{0,1\}^n$ mit Interpretation $(b_1, \dots, b_n) \in \Omega$ kodiert Ergebnis in den Würfen 1 bis n via $b_i = 1$ für Kopf im i-ten Wurf, $b_i = 0$ für Zahl im i-ten Wurf. Wir betrachten die Ereignisse $A_1 = \{(1,\dots,1)\}$ (n-mal Kopf), $A_2 = \{b \in \Omega \,|\, b_1 + \cdots + b_n = n-1\}$ ($(n-1)$-mal Kopf), $A_3 = \{b \in \Omega \,|\, b_1 + \cdots + b_n \leqslant n-1\}$ (mindestens einmal Zahl). Sind alle Versuchsausgänge gleichwahrscheinlich, so erhalten wir $P(A_1) = 2^{-n}$, $P(A_2) = n2^{-n}$, $P(A_3) = 1 - 2^{-n}$. Beachte, dass $A_3 = A_1^\complement$ und $P(A_1) + P(A_3) = 1$ gilt. A_1 und A_3 heißen auch komplementäre Ereignisse oder Gegenereignisse.

(c) Wir nehmen nun an, dass die Münze beliebig oft hintereinander geworfen wird. Die Versuchsausgänge modellieren wir dann durch 0-1-Folgen $B = (b_n)_{n\in\mathbb{N}} = (b_1, b_2, \dots)$, d.h. $\Omega = \{0,1\}^{\mathbb{N}}$. Ereignisse sind dann $A_1 = \{(1,1,\dots)\}$ (immer nur Kopf), $A_2 = \{b \in \Omega \,|\, \forall M \in \mathbb{N}\, \exists k \in \mathbb{N} : b_k = b_{k+1} = \cdots = b_{k+M-1} = 1\}$ (für jedes $M \in \mathbb{N}$ gibt es einen Kopf-run der Länge M). Wegen $|\Omega| = \infty$ gibt es keine Gleichverteilung auf den Versuchsausgängen. Intuitiv würden wir aus (b) schließen $P(A_1) = \lim_{n\to\infty} 2^{-n} = 0$, während $P(A_2)$ nicht klar ist. Es fehlt eine mathematisch formale Einführung des Wahrscheinlichkeitsmaßes P und seiner Eigenschaften.

1.2 Definition. Mit Ω werde die nichtleere Menge der möglichen Versuchsausgänge oder Ergebnismenge, Grundmenge bezeichnet. Ein Teilmengensystem $\mathscr{F} \subseteq \mathscr{P}(\Omega)$ heißt Menge der interessierenden Ereignisse oder mathematisch σ-Algebra, falls gilt:

(a) $\Omega \in \mathscr{F}$;

(b) $A \in \mathscr{F} \Rightarrow A^c \in \mathscr{F}$;

(c) $A_n \in \mathscr{F}, n \in \mathbb{N} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathscr{F}$.

Die Elemente von $\mathscr{F}$ heißen Ereignisse. Ein Wahrscheinlichkeitsmaß P (auch Wahrscheinlichkeitsverteilung genannt) auf $\mathscr{F}$ ist eine Abbildung $P : \mathscr{F} \to [0,1]$, die den *Kolmogorovschen Axiomen* (1933) genügt:

(a) $P(\Omega) = 1$ (Normierung);

(b) für $A_n \in \mathscr{F}$, $n \in \mathbb{N}$, paarweise disjunkt gilt

$$P\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} P(A_n) \text{ (σ-Additivität).}$$

Ein Wahrscheinlichkeitsraum ist ein Tripel $(\Omega, \mathscr{F}, P)$, bestehend aus einer Ergebnismenge Ω, einer σ-Algebra $\mathscr{F}$ über Ω sowie einem Wahrscheinlichkeitsmaß P auf $\mathscr{F}$.

1.3 Beispiele. Auf jeder nichtleeren Ergebnismenge Ω existieren die triviale σ-Algebra $\{\varnothing, \Omega\}$ sowie die Potenzmenge $\mathscr{P}(\Omega)$ als σ-Algebren. Für $\omega_0 \in \Omega$ ist das Einpunkt- oder Diracmaß $\delta_{\omega_0}(A) = \mathbf{1}(\omega_0 \in A)$, $A \in \mathscr{F}$, ein Wahrscheinlichkeitsmaß auf jeder σ-Algebra $\mathscr{F}$ über Ω. Sind $(P_n)_{n\geqslant 1}$ Wahrscheinlichkeitsmaße auf einer σ-Algebra $\mathscr{F}$, so auch jede Konvexkombination $\sum_{n\geqslant 1} w_n P_n$ mit $w_n \geqslant 0$ und $\sum_{n\geqslant 1} w_n = 1$. Die Einpunktmaße bilden Extremalpunkte der konvexen Menge aller Wahrscheinlichkeitsmaße auf $\mathscr{F}$.

1.4 Lemma. *Für jede σ-Algebra $\mathscr{F}$ gilt:*

(a) $\varnothing \in \mathscr{F}$;

(b) $A_1, A_2 \in \mathscr{F} \Rightarrow A_1 \cup A_2 \in \mathscr{F}$;

(c) $A_n \in \mathscr{F}, n \in \mathbb{N} \Rightarrow \bigcap_{n\in\mathbb{N}} A_n, A_1 \cap A_2 \in \mathscr{F}$.

1.5 Lemma. *Für jedes Wahrscheinlichkeitsmaß $P : \mathscr{F} \to [0,1]$ gilt:*

(a) $P(\varnothing) = 0$;

(b) $A, B \in \mathscr{F}$, $A \subseteq B \Rightarrow P(A) \leqslant P(B)$;

(c) $\forall\, A, B \in \mathscr{F} :\ P(A \cup B) = P(A) + P(B) - P(A \cap B)$;

(d) $\forall\, A_n \in \mathscr{F}$, $n \geqslant 1 :\ P(\bigcup_{n\geqslant 1} A_n) \leqslant \sum_{n\geqslant 1} P(A_n)$ *(Subadditivität)*;

(e) Für $A_n \in \mathscr{F}$, $n \geqslant 1$, *mit* $A_n \uparrow A$ *(d.h.* $A_n \subseteq A_{n+1}$, $\bigcup_n A_n = A$*) gilt* $P(A) = \lim_{n\to\infty} P(A_n)$ *(σ-Stetigkeit).*

Andererseits ist jede normierte, additive Mengenfunktion $Q : \mathscr{F} \to [0,1]$ (d.h. $Q(\Omega) = 1$, $Q(A \cup B) = Q(A) + Q(B)$ für alle disjunkten $A, B \in \mathscr{F}$), die σ-stetig ist, auch σ-additiv und damit ein Wahrscheinlichkeitsmaß.

1.6 Definition. Es sei $(\Omega, \mathscr{F}, P)$ ein Wahrscheinlichkeitsraum und $(S, \mathscr{S})$ ein Messraum. Dann heißt eine Funktion $g : \Omega \to S$ messbar (bzgl. $(\mathscr{F}, \mathscr{S})$), falls

$$\forall\, A \in \mathscr{S} :\ g^{-1}(A) \in \mathscr{F}$$

gilt. Jede solche messbare Funktion heißt $(S, \mathscr{S})$-wertige Zufallsvariable. Für $S = \mathbb{R}^d$ wird kanonisch $\mathscr{S} = \mathfrak{B}_{\mathbb{R}^d}$ (Borel-σ-Algebra, siehe unten) gewählt, und man spricht bloß von einer Zufallsvariablen ($d = 1$) bzw. einem Zufallsvektor ($d \geqslant 2$).
Die Verteilung einer $(S, \mathscr{S})$-wertigen Zufallsvariablen X ist das Wahrscheinlichkeitsmaß (!)

$$P^X(A) := P(X \in A) = P(X^{-1}(A)), \quad A \in \mathscr{S}.$$

Die Verteilung P^X von X ist also das Bildmaß von P unter X. Mit der Verteilungsfunktion (Dichte, Zähldichte) von X werden wir stets die zu P^X gehörige Größe meinen.

Wir schreiben kurz $\{X \in A\} := \{\omega \in \Omega \,|\, X(\omega) \in A\}$, $\{X = x\} := \{\omega \in \Omega \,|\, X(\omega) = x\}$, $P(X \in A) := P(\{X \in A\})$, $P(X = x) := P(\{X = x\})$ etc.

1.7 Beispiele.

(a) Beim Würfeln mit 2 Würfeln interessiert uns die Augensumme X als abgeleiteter Parameter. Formal betrachten wir $\Omega = \{1, 2, 3, 4, 5, 6\}^2$, $X : \Omega \to \mathbb{R}$ mit $X((n_1, n_2)) = n_1 + n_2$. Das Ereignis $\{X = 7\}$ ist dann kurz für $\{\omega \in \Omega : | X(\omega) = 7\} = X^{-1}(\{7\}) = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} \subseteq \Omega$ und $\{X \text{ ist gerade}\} = \{\omega \in \Omega \,|\, X(\omega) \in \{2, 4, 6, 8, 10, 12\}\}$.

X transportiert das Wahrscheinlichkeitsmaß P auf Ω nach $\mathbb{R}$: $P^X(A) := P(X^{-1}(A)) = P(\{\omega \in \Omega \,|\, X(\omega) \in A\}) = P(X \in A)$ für $A \subseteq \mathbb{R}$. Beachte, dass hier jede Funktion $X : \Omega \to \mathbb{R}$ messbar und damit Zufallsvariable ist, da Ω mit der Potenzmenge $\mathscr{P}(\Omega)$ als σ-Algebra versehen ist, so dass P^X sogar auf $\mathscr{P}(\mathbb{R})$ wohldefiniert ist.

(b) Auf einem Wahrscheinlichkeitsram $(\Omega, \mathscr{F}, P)$ ist die Indikatorfunktion $X(\omega) = \mathbf{1}_A(\omega) = \mathbf{1}(\omega \in A)$, $\omega \in \Omega$, genau dann eine reellwertige Zufallsvariable, wenn $A \in \mathscr{F}$ gilt. Für ihre Verteilung gilt $P^X = P(A^\complement)\delta_0 + P(A)\delta_1$.

1.2 Diskrete Verteilungen

1.8 Definition. Ist Ω eine abzählbare (d.h. endliche oder abzählbar unendliche) Menge und P ein Wahrscheinlichkeitsmaß auf $\mathscr{F} = \mathscr{P}(\Omega)$, so heißt $(\Omega, \mathscr{F}, P)$ diskreter Wahrscheinlichkeitsraum. Man nennt eine S-wertige Zufallsvariable X diskret verteilt, falls sie bezüglich $\mathscr{P}(S)$ messbar ist und einen diskreten Wahrscheinlichkeitsraum $(S, \mathscr{P}(S), P^X)$ generiert. Ist X eine diskrete S-wertige Zufallsvariable und $S \subseteq \mathbb{R}^d$ (S abzählbar und somit $\mathscr{P}(S) \subseteq \mathfrak{B}_{\mathbb{R}^d}$), so bezeichnet man X auch als diskrete $\mathbb{R}^d$-wertige Zufallsvariable.

1.9 Beispiel. Das Modell des Würfel- und des n-fachen Münzwurfs bildet einen diskreten Wahrscheinlichkeitsraum. Die Augensumme beim Münzwurf ist eine diskret verteilte Zufallsvariable.

1.10 Lemma.

(a) Ist $(\Omega, \mathscr{F}, P)$ ein diskreter Wahrscheinlichkeitsraum, so ist P eindeutig durch seine <u>*Zähldichte*</u> *$p : \Omega \to [0,1]$ mit $p(\omega) := P(\{\omega\})$ festgelegt.*

Ebenso legt bei einer diskret verteilten S-wertigen Zufallsvariablen X die zugehörige Zähldichte $p^X(s) = P(X = s)$, $s \in S$, die Verteilung P^X eindeutig fest.

(b) Ist andererseits Ω eine endliche oder abzählbar unendliche Menge und besitzt $p : \Omega \to [0,1]$ die Eigenschaft $\sum_{\omega \in \Omega} p(\omega) = 1$, so wird durch

$$P(A) := \sum_{\omega \in A} p(\omega), \quad A \subseteq \Omega,$$

ein Wahrscheinlichkeitsmaß P auf $\mathscr{F} = \mathscr{P}(\Omega)$ definiert, dessen Zähldichte p ist.

1.11 Lemma (Urnenmodelle)**.** *In einer Urne liegen N Kugeln mit den Aufschriften $1, 2, \dots, N$. Es werden n Kugeln gezogen. Dann gilt für die Anzahl verschiedener Versuchsausgänge:*

Mit Zurücklegen, mit Betrachtung der Reihenfolge:
$\Omega_1 = \{1, \dots, N\}^n$, $|\Omega_1| = N^n$.

Ohne Zurücklegen, mit Betrachtung der Reihenfolge:
$\Omega_2 = \{(k_1, \dots, k_n) \,|\, k_1, \dots, k_n \in \{1, \dots, N\}$ *paarweise verschieden*$\}$, $|\Omega_2| = \frac{N!}{(N-n)!}$ *für* $n \leqslant N$.

Ohne Zurücklegen, ohne Betrachtung der Reihenfolge:
$\Omega_3 = \{A \subseteq \{1, \dots, N\} \,|\, |A| = n\}$, $|\Omega_3| = \binom{N}{n}$ *für* $n \leqslant N$.

Mit Zurücklegen, ohne Betrachtung der Reihenfolge:
$\Omega_4 = \{(s_1, \dots, s_n) \,|\, 1 \leqslant s_1 \leqslant s_2 \leqslant \dots \leqslant s_n \leqslant N\}$, $|\Omega_4| = \binom{N+n-1}{n}$.

1.12 Beispiel. Wie groß ist die Wahrscheinlichkeit, dass in einem Raum mit n Personen keine zwei Personen am selben Tag Geburtstag haben? Geht man von 365 Tagen im Jahr aus, so ist die Menge aller Geburtstagskombinationen gerade Ω_1 mit $N = 365$. Das Ereignis, das keine zwei Personen am selben Tag Geburtstag haben, entspricht dann gerade Ω_2. Unter der Annahme einer Gleichverteilung ergibt sich daher für die gesuchte Wahrscheinlichkeit $\frac{N!}{N^n (N-n)!}$. Approximativ ergibt sich $\exp(-n(n-1)/(2N))$ und konkret $0,432$ für $n = 25$; $4,4 \bullet 10^{-4}$ für $n = 50$; $2,2 \bullet 10^{-9}$ für $n = 80$; $2,7 \bullet 10^{-14}$ für $n = 100$.

1.13 Definition. Folgende Zähldichten beschreiben wichtige Verteilungen:

Laplace-/Gleich-Verteilung: $p_{Lap(\Omega)}(\omega) = \frac{1}{|\Omega|}$, $\omega \in \Omega$, für $|\Omega| < \infty$;

hypergeometrische Verteilung: Parameter $0 \leqslant n \leqslant N$, $0 \leqslant W \leqslant N$

$$p_{Hyp(N,W,n)}(w) = \frac{\binom{N-W}{n-w}\binom{W}{w}}{\binom{N}{n}}, \quad w \in \{0, \dots, W\}.$$

Bernoulli-Schema: Länge $n \in \mathbb{N}$, Erfolgswahrscheinlichkeit $p \in [0,1]$

$$p_{Bern(n,p)}(\omega) = p^{\sum_{i=1}^n \omega_i}(1-p)^{n-\sum_{i=1}^n \omega_i}, \quad \omega = (\omega_1, \dots, \omega_n) \in \{0,1\}^n.$$

Binomialverteilung: Länge $n \in \mathbb{N}$, Erfolgswahrscheinlichkeit $p \in [0,1]$

$$p_{Bin(n,p)}(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0,1,\dots,n\}.$$

Geometrische Verteilung: Erfolgswahrscheinlichkeit $p \in (0,1]$

$$p_{Geo(p)}(k) = (1-p)^{k-1}p, \quad k \in \mathbb{N}\,.$$

Poissonverteilung: Parameter $\lambda > 0$

$$p_{Pois(\lambda)}(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0\,.$$

1.14 Satz (Poissonscher Grenzwertsatz)**.** *Es seien $p_n \in [0,1]$ gegeben mit $\lim_{n\to\infty} np_n = \lambda > 0$. Dann gilt für alle $k \in \mathbb{N}_0$*

$$\lim_{n\to\infty} p_{Bin(n,p_n)}(k) = p_{Pois(\lambda)}(k).$$

1.15 Satz (Vitali, 1903)**.** *Sei $\Omega = \{0,1\}^{\mathbb{N}}$ der Ergebnisraum des unendlich oft wiederholten Münzwurfs. Dann gibt es kein Wahrscheinlichkeitsmaß P auf der Potenzmenge $\mathscr{P}(\Omega)$, das folgender Invarianzeigenschaft genügt:*

$$\forall A \subseteq \Omega,\, n \in \mathbb{N}:\ P(T_n(A)) = P(A),$$

wobei $T_n(\omega) = T_n(\omega_1, \omega_2, \dots) = (\omega_1, \dots, \omega_{n-1}, 1-\omega_n, \omega_{n+1}, \dots)$ das Ergebnis des n-ten Wurfs umkehrt.

1.3 Maßtheorie und Wahrscheinlichkeitsmaße im $\mathbb{R}^d$

1.16 Lemma. *Es sei $\mathscr{E} \subseteq \mathscr{P}(\Omega)$ ein System von Teilmengen von Ω. Dann gibt es eine kleinste σ-Algebra $\mathscr{F}$, die $\mathscr{E}$ enthält.*

1.17 Definition. In der Situation des vorigen Lemmas sagt man, dass die σ-Algebra $\mathscr{F}$ von $\mathscr{E}$ erzeugt wird. $\mathscr{E}$ heißt Erzeuger von $\mathscr{F}$ und man schreibt $\mathscr{F} = \sigma(\mathscr{E})$.

1.18 Definition. Es sei (S,d) ein metrischer Raum. Dann heißt $\mathfrak{B}_S := \sigma(\{O \subseteq S \,|\, O \text{ offen}\})$ Borel-σ-Algebra über S.

1.19 Satz.

(a) Die Borel-σ-Algebra $\mathfrak{B}_{\mathbb{R}}$ über $\mathbb{R}$ wird auch erzeugt von folgenden Mengensystemen:

(i) $\mathscr{E}_1 := \{(a,b) \,|\, a,b \in \mathbb{R}\}$;

(ii) $\mathscr{E}_2 := \{[a,b] \,|\, a,b \in \mathbb{R}\}$;

(iii) $\mathscr{E}_3 := \{(a, b] \mid a, b \in \mathbb{R}\}$*;*

(iv) $\mathscr{E}_4 := \{(-\infty, b] \mid b \in \mathbb{R}\}$*;*

(v) $\mathscr{E}_5 := \{(-\infty, b) \mid b \in \mathbb{R}\}$*.*

(b) *Die Borel-σ-Algebra $\mathfrak{B}_{\mathbb{R}^d}$ über $\mathbb{R}^d$ wird auch erzeugt von folgenden Mengensystemen:*

(i) $\mathscr{E}_1^d := \{(a_1, b_1) \times \cdots \times (a_d, b_d) \mid a_k, b_k \in \mathbb{R}, k = 1, \dots, d\}$*;*

(ii) $\mathscr{E}_2^d := \{[a_1, b_1] \times \cdots \times [a_d, b_d] \mid a_k, b_k \in \mathbb{R}, k = 1, \dots, d\}$*;*

(iii) $\mathscr{E}_3^d := \{(a_1, b_1] \times \cdots \times (a_d, b_d] \mid a_k, b_k \in \mathbb{R}, k = 1, \dots, d\}$*;*

(iv) $\mathscr{E}_4^d := \{(-\infty, b_1] \times \cdots \times (-\infty, b_d] \mid b_k \in \mathbb{R}, k = 1, \dots, d\}$*;*

(v) $\mathscr{E}_5^d := \{(-\infty, b_1) \times \cdots \times (-\infty, b_d) \mid b_k \in \mathbb{R}, k = 1, \dots, d\}$*.*

1.20 Lemma. *Eine Funktion $g : \Omega \to S$ ist bereits $(\mathscr{F}, \mathscr{S})$-messbar, falls für einen Erzeuger $\mathscr{E}$ von $\mathscr{S}$ gilt*

$$\forall A \in \mathscr{E} : \ g^{-1}(A) \in \mathscr{F}.$$

1.21 Korollar.

(a) *Jede stetige Funktion $g : S \to T$ zwischen metrischen Räumen (S, d_S) und (T, d_T) ist <u>Borel-messbar</u>, d.h. $(\mathfrak{B}_S, \mathfrak{B}_T)$-messbar.*

(b) *Jede Funktion $g : \Omega \to \mathbb{R}$ mit $\{g \leqslant y\} \in \mathscr{F}$ für alle $y \in \mathbb{R}$ ist $(\mathscr{F}, \mathfrak{B}_\mathbb{R})$-messbar.*

(c) *Falls $g_n : \Omega \to \mathbb{R}$ $(\mathscr{F}, \mathfrak{B}_\mathbb{R})$-messbar sind für alle $n \geqslant 1$, so auch $\inf_n g_n$, $\sup_n g_n$, $\limsup_n g_n$, $\liminf_n g_n$, sofern diese Funktionen endlich sind. Falls der punktweise Grenzwert $\lim_n g_n$ überall existiert, so ist auch dieser $(\mathscr{F}, \mathfrak{B}_\mathbb{R})$-messbar.*

(d) *Sind $g_1, \dots, g_d : \Omega \to \mathbb{R}$ $(\mathscr{F}, \mathfrak{B}_\mathbb{R})$-messbar und ist $h : \mathbb{R}^d \to \mathbb{R}^k$ Borel-messbar, so ist $\omega \mapsto h(g_1(\omega), \dots, g_d(\omega))$ $(\mathscr{F}, \mathfrak{B}_{\mathbb{R}^k})$-messbar; insbesondere sind also messbar: $(g_1, \dots, g_d)$, $g_1 + g_2$, $g_1 - g_2$, $g_1 \bullet g_2$, g_1/g_2 (falls überall wohldefiniert), $\max(g_1, g_2)$, $\min(g_1, g_2)$.*

(e) *Ist $g : \Omega \to S$ $(\mathscr{F}, \mathscr{S})$-messbar und $h : S \to T$ $(\mathscr{S}, \mathscr{T})$-messbar, so ist die Komposition $h \circ g$ $(\mathscr{F}, \mathscr{T})$-messbar.*

1.22 Definition. Es sei Ω eine nichtleere Menge. Dann heißt $\mathscr{A} \subseteq \mathscr{P}(\Omega)$ <u>Algebra</u> über Ω, falls gilt:

(a) $\Omega \in \mathscr{A}$;

(b) $A \in \mathscr{A} \Rightarrow A^c \in \mathscr{A}$;

(c) $A, B \in \mathscr{A} \Rightarrow A \cup B \in \mathscr{A}$.

Eine Abbildung $\mu : \mathscr{A} \to [0, \infty]$ heißt <u>Prämaß</u> über $\mathscr{A}$, falls

(a) $\mu(\varnothing) = 0$;

(b) für $A_n \in \mathscr{A}$, $n \in \mathbb{N}$, paarweise disjunkt mit $\bigcup_n A_n \in \mathscr{A}$ gilt

$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n) \text{ (σ-Additivität).}$$

μ heißt Maß, falls $\mathscr{A}$ bereits eine σ-Algebra ist. Ein Maß μ heißt σ-endlich, falls es $A_n \in \mathscr{A}$, $n \in \mathbb{N}$, gibt mit $\mu(A_n) < \infty$ und $\Omega = \bigcup_n A_n$. Konsistent mit obiger Definition heißt ein Maß μ Wahrscheinlichkeitsmaß, falls $\mu(\Omega) = 1$ gilt.

1.23 Satz (Maßerweiterungssatz von Carathéodory, 1917)**.** *Jedes Prämaß μ auf einer Algebra $\mathscr{A}$ kann zu einem Maß $\tilde{\mu}$ auf der von $\mathscr{A}$ erzeugten σ-Algebra $\mathscr{F} = \sigma(\mathscr{A})$ fortgesetzt werden, d.h. $\tilde{\mu}$ ist ein Maß auf $\mathscr{F}$ mit $\tilde{\mu}(A) = \mu(A)$ für alle $A \in \mathscr{A}$.*

1.24 Satz (Eindeutigkeitssatz)**.** *Es seien μ und ν σ-endliche Maße auf $(\Omega, \mathscr{F})$ und es gebe $A_n \in \mathscr{F}$, $n \in \mathbb{N}$, mit $\mu(A_n) = \nu(A_n) < \infty$ und $\bigcup_n A_n = \Omega$. Stimmen μ und ν auf einem Erzeuger $\mathscr{E}$ von $\mathscr{F}$ überein, der in dem Sinne $\cap$-stabil ist, dass $A, B \in \mathscr{E} \Rightarrow A \cap B \in \mathscr{E}$ gilt, so stimmen μ und ν auf der ganzen σ-Algebra $\mathscr{F}$ überein. Insbesondere ist ein Wahrscheinlichkeitsmaß durch seine Werte auf einem $\cap$-stabilen Erzeuger eindeutig festgelegt.*

1.25 Definition. Für ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$ ist die zugehörige Verteilungsfunktion gegeben durch $F(x) := P((-\infty, x])$, $x \in \mathbb{R}$; für $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$-wertige Zufallsvariablen X wird durch $F^X(x) := P^X((-\infty, x]) = P(X \leqslant x)$, $x \in \mathbb{R}$, die zugehörige Verteilungsfunktion definiert.

1.26 Lemma. *Jede Verteilungsfunktion F ist monoton wachsend, rechtsstetig und erfüllt $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$.*

1.27 Satz. *Es sei $F : \mathbb{R} \to \mathbb{R}$ eine monoton wachsende, rechtsstetige Funktion. Dann existiert ein Maß μ auf $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$ mit*

$$\mu((a, b]) = F(b) - F(a), \quad a < b \in \mathbb{R}.$$

μ ist eindeutig durch F definiert und heißt Lebesgue-Stieltjes-Maß zu F.

1.28 Korollar. *Es gibt genau ein Maß λ auf $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$ mit $\lambda((a, b]) = b - a$, das Lebesguemaß.*

1.29 Korollar. *Ist $F : \mathbb{R} \to [0, 1]$ monoton wachsend und rechtsstetig mit $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$, so existiert genau ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$ mit $P((a, b]) = F(b) - F(a)$ für alle $a < b$. Insbesondere ist F die Verteilungsfunktion von P.*

1.30 Beispiel. Für eine reellwertige Zufallsvariable X mit Verteilungsfunktion F^X betrachte $U = F^X(X)$. F^X ist Borel-messbar und somit U eine Zufallvariable mit Werten in $[0, 1]$. Ist F^X stetig, so gilt mit der Rechtsinversen $(F^X)^{-1}(p) = \inf\{x \in \mathbb{R} \mid F(x) \geqslant p\}$, dass U die Verteilungsfunktion $F^U(u) = P(F^X(X) \leqslant u) = F^X((F^X)^{-1}(u)) = u$ für $u \in (0, 1)$ besitzt und somit $U((0, 1))$-verteilt ist.

Ist andererseits U eine $U((0,1))$-verteilte Zufallsvariable (oder eine Pseudozufallszahl in Anwendungen), so gilt für jede Verteilungsfunktion F, dass die Zufallsvariable $X = F^{-1}(U)$ die Verteilungsfunktion $F^X = F$ besitzt. F^{-1} heißt Quantilsfunktion und die Simulationsmethode Quantilstransformation. Eine einfache Anwendung ist die Erzeugung einer $\mathrm{Bin}(1,p)$-verteilten Zufallsvariablen X: es gilt $F^X(x) = (1-p)\mathbf{1}(x \geqslant 0) + p\mathbf{1}(x \geqslant 1)$ und $(F^X)^{-1}(y) = \mathbf{1}(y \geqslant 1-p)$, $y \in (0,1)$, so dass $\mathbf{1}(U \geqslant 1-p)$ $\mathrm{Bin}(1,p)$-verteilt ist für eine $U((0,1))$-verteilte Zufallsvariable U.

1.31 Definition. Ist $f : \mathbb{R}^d \to [0,\infty)$ eine Lebesgue-integrierbare Funktion mit $\int_{\mathbb{R}^d} f(x)\,dx = 1$, so heißt f Wahrscheinlichkeitsdichte oder kurz Dichte auf $\mathbb{R}^d$.

1.32 Korollar. *Jede Wahrscheinlichkeitsdichte f auf $\mathbb{R}$ erzeugt mittels*

$$P_f((a,b]) = \int_a^b f(x)\,dx, \quad a,b \in \mathbb{R},\, a < b,$$

ein eindeutiges Wahrscheinlichkeitsmaß P_f auf $\mathfrak{B}_{\mathbb{R}}$.

1.33 Lemma.

(a) Ist f die Dichte eines Wahrscheinlichkeitsmaßes P auf $\mathfrak{B}_{\mathbb{R}}$ mit Verteilungsfunktion F, so gilt $F(x) = \int_{-\infty}^x f(y)\,dy$ für alle $x \in \mathbb{R}$.

(b) Ist die Verteilungsfunktion F eines Wahrscheinlichkeitsmaßes P auf $\mathfrak{B}_{\mathbb{R}}$ (schwach) differenzierbar, so ist $f(x) := F'(x)$ die zugehörige Wahrscheinlichkeitsdichte.

Vollkommen Analoges gilt für die Dichte f^X, die Verteilungsfunktion F^X und die Verteilung P^X einer reellwertigen Zufallsvariablen X.

1.34 Satz. *Jede Wahrscheinlichkeitsdichte f auf $\mathbb{R}^d$ erzeugt mittels*

$$P_f((a_1,b_1] \times \cdots \times (a_d,b_d]) = \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} f(x_1,\ldots,x_d)\,dx_d \cdots dx_1$$

für $a_k, b_k \in \mathbb{R}$ mit $a_k < b_k$ ein eindeutiges Wahrscheinlichkeitsmaß P_f auf $\mathfrak{B}_{\mathbb{R}^d}$, und es gilt $P_f(B) = \int_B f(x)dx$.

1.35 Definition. Folgende Wahrscheinlichkeitsdichten beschreiben wichtige Verteilungen auf $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$:

Gleichverteilung: $f_{U(G)}(x) = \frac{1}{\lambda(G)}\mathbf{1}_G(x)$ für $G \in \mathfrak{B}_{\mathbb{R}}$ mit Lebesguemaß $\lambda(G) \in (0,\infty)$;

Exponentialverteilung: $f_{Exp(\lambda)}(x) = \lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}^+}(x)$ mit Parameter $\lambda > 0$;

Normalverteilung: $f_{N(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$ mit Parametern $\mu \in \mathbb{R}$, $\sigma > 0$.

1.36 Definition. Sind $f_1, \dots, f_d$ Wahrscheinlichkeitsdichten auf $\mathbb{R}$, so heißt

$$f(x_1, \dots, x_d) = \prod_{k=1}^{d} f_k(x_k), \quad x_1, \dots, x_d \in \mathbb{R},$$

Produktdichte der $(f_k)_{k=1,\dots,d}$ im $\mathbb{R}^d$. Insbesondere ist die d-dimensionale Standard-Normalverteilung $N(0, E_d)$ im $\mathbb{R}^d$ definiert über die Dichte

$$f(x) = (2\pi)^{-d/2} e^{-|x|^2/2},\ x \in \mathbb{R}^d, \quad \text{mit } |x|^2 = \sum_{i=1}^{d} x_i^2.$$

1.37 Satz (Dichtetransformationssatz)**.** *Es seien X ein d-dimensionaler Zufallsvektor mit Dichte f^X sowie $Y = \varphi(X)\mathbf{1}(X \in U)$ für einen C^1-Diffeomorphismus $\varphi : U \to V$ mit $U, V \subseteq \mathbb{R}^d$ offen und $P(X \in U) = 1$. Dann ist Y ein Zufallsvektor mit Dichte*

$$f^Y(y) = f^X(\varphi^{-1}(y))|\det(D(\varphi^{-1})(y))|\mathbf{1}(y \in V), \quad y \in \mathbb{R}^d.$$

Ist allgemeiner $\varphi : U \to V$ und $U = \bigcup_{i=1}^n U_i$ eine Zerlegung in offene Mengen U_i derart, dass $P(X \in U) = 1$ und $\varphi_i = \varphi|_{U_i} : U_i \to \varphi(U_i)$, $i = 1, \dots, n$, jeweils einen C^1-Diffeomorphismus definiert, so besitzt $Y = \varphi(X)$ die Dichte $f^Y = \sum_{i=1}^n g_i$ mit

$$g_i(y) = f^X(\varphi_i^{-1}(y))|\det(D(\varphi_i^{-1})(y))|\mathbf{1}(y \in \varphi(U_i)), \quad y \in \mathbb{R}^d.$$

Konvention: setze $A \bullet 0 := 0$ auch für Ausdrücke A, die nicht wohldefiniert sind.

1.38 Korollar.

(a) *Ist X eine reellwertige Zufallsvariable mit Dichte f^X, so besitzt $Y = aX + b$ für $a \in \mathbb{R} \setminus \{0\}$, $b \in \mathbb{R}$ die Dichte $f^Y(y) = |a|^{-1} f^X(a^{-1}(y - b))$.*

(b) *Ist X ein d-dimensionaler Zufallsvektor mit Dichte f^X, so besitzt $Y = AX + b$ für $A \in \mathbb{R}^{d \times d}$ invertierbar und $b \in \mathbb{R}^d$ die Dichte $f^Y(y) = |\det(A)|^{-1} f^X(A^{-1}(y - b))$.*

1.39 Beispiele.

(a) Ist X eine $N(0,1)$-verteilte Zufallsvariable, so ist $Y = \mu + \sigma X$ eine $N(\mu, \sigma^2)$-verteilte Zufallsvariable mit $\mu \in \mathbb{R}$, $\sigma > 0$. Weiterhin ist $Z = X^2$ eine $\chi^2(1)$-verteilte Zufallsvariable, wobei die χ^2-Verteilung mit einem Freiheitsgrad gegeben ist durch die Wahrscheinlichkeitsdichte $f_{\chi^2(1)}(x) = \frac{1}{\sqrt{2\pi x}} e^{-x/2} \mathbf{1}_{\mathbb{R}^+}(x)$, $x \in \mathbb{R}$.

(b) Sind X ein d-dimensionaler standard-normalverteilter Zufallsvektor sowie $\mu \in \mathbb{R}^d$, $A \in \mathbb{R}^{d \times d}$ mit $\det(A) \neq 0$ deterministisch, so ist $Y = \mu + AX$ ein d-dimensionaler Zufallsvektor mit Dichte

$$\varphi_{\mu,\Sigma}(x) = (2\pi)^{-d/2} \det(\Sigma)^{-1/2} \exp\Big(-\frac{1}{2} \langle \Sigma^{-1}(x - \mu), x - \mu \rangle \Big), \quad x \in \mathbb{R}^d,$$

wobei $\Sigma = A^\top A$ eine symmetrische positiv-definite Matrix ist. Y ist normalverteilt mit Mittelwertvektor μ und Kovarianzmatrix Σ, kurz $Y \sim N(\mu, \Sigma)$.

2 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

2.1 Bedingte Wahrscheinlichkeiten und Bayes-Formel

2.1 Definition. Es seien A und B Ereignisse mit $P(B) > 0$. Dann wird mit

$$P(A \,|\, B) := \frac{P(A \cap B)}{P(B)}$$

die bedingte Wahrscheinlichkeit von A gegeben (oder: unter) B bezeichnet.

2.2 Beispiel. Beim Wurf von zwei Würfeln mit den Ereignissen $A =$„Pasch“ und $B =$„beide Augenzahlen ungerade“ gilt $P(A \,|\, B) = \frac{3/36}{9/36} = \frac{1}{3} > \frac{1}{6} = P(A)$.

2.3 Satz. *Auf dem Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, P)$ sei B ein Ereignis mit $P(B) > 0$. Dann gilt:*

(a) *Durch $Q(A) := P(A \,|\, B)$ wird ein Wahrscheinlichkeitsmaß Q auf $\mathscr{F}$ definiert.*

(b) *(Formel von der totalen Wahrscheinlichkeit) Es sei $B = \bigcup_{i=1}^{N} B_i$ Vereinigung paarweise disjunkter Ereignisse B_i mit $P(B_i) > 0$. Dann folgt für jedes Ereignis A*

$$P(A \cap B) = \sum_{i=1}^{N} P(B_i) P(A \,|\, B_i).$$

(c) *(Bayesformel) Für jedes Ereignis A mit $P(A) > 0$ und jede Zerlegung $\Omega = \bigcup_{i=1}^{N} B_i$ von Ω in paarweise disjunkte Ereignisse B_i mit $P(B_i) > 0$ gilt*

$$P(B_i \,|\, A) = \frac{P(B_i) P(A \,|\, B_i)}{\sum_{j=1}^{N} P(B_j) P(A \,|\, B_j)}.$$

In (b) und (c) kann auch $N = \infty$ gesetzt werden.

2.4 Beispiel. Eine seltene Krankheit kommt bei $0,5\%$ der Bevölkerung vor. Ein Test führt bei 99% der Kranken zu einer positiven Reaktion, aber auch bei 2% der Gesunden. Die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person mit positivem Testergebnis krank ist, beträgt dann nach der Bayesformel $\frac{0{,}005 \bullet 0{,}99}{0{,}005 \bullet 0{,}99 + 0{,}995 \bullet 0{,}02} \approx 0,2$, d.h. nur ca. 20%.

2.5 Lemma (Multiplikationsformel/Pfadregel)**.** *Für Ereignisse $A_1, \ldots, A_n$ mit $P(A_1 \cap \cdots \cap A_{n-1}) > 0$ gilt*

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) P(A_2 \,|\, A_1) P(A_3 \,|\, A_1 \cap A_2) \cdots P(A_n \,|\, A_1 \cap \cdots \cap A_{n-1}).$$

2.6 Beispiel. Beim Ziehen aus einer Urne mit W weißen und S schwarzen Kugeln ohne Zurücklegen und mit Beachtung der Reihenfolge ergibt sich mit $N = S + W$ für das Ergebnis 'SSW' (d.h. 1. und 2. Kugel schwarz, 3. Kugel weiß) nach der Pfadregel die Wahrscheinlichkeit $\frac{S}{N} \bullet \frac{S-1}{N-1} \bullet \frac{W}{N-2}$. Dieselbe Wahrscheinlichkeit besitzen die Versuchsausgänge 'SWS' und 'WSS' (man nennt die Verteilung *austauschbar*).

2.7 Beispiel (Scheinkorrelationen)**.** An einer Universität werden von 825/560/325 männlichen Bewerbern für Fach 1/2/3 jeweils 62%/63%/34% zugelassen, von 108/25/593 weiblichen Bewerberinnen hingegen 82%/68%/37%. Obwohl die Zulassungsquote in jedem Fach für Frauen höher war, ergibt sich nach der Formel von der totalen Wahrscheinlichkeit insgesamt eine Zulassungsquote von ca. 57% für Männer und von ca. 45% für Frauen, weil Letztere sich stärker für Fach 3 mit schwierigerer Zulassung beworben haben.

2.2 Unabhängige Ereignisse und Lemma von Borel-Cantelli

2.8 Definition.

(a) Zwei Ereignisse A und B heißen (stochastisch) unabhängig (unter P), falls $P(A \cap B) = P(A)P(B)$ gilt.

(b) Eine Familie $(A_i)_{i \in I}$ von Ereignissen, $I \neq \varnothing$ beliebige Indexmenge, heißt (stochastisch) unabhängig, falls für jede endliche Teilmenge $J \subseteq I$ gilt

$$P\Big(\bigcap_{j \in J} A_j\Big) = \prod_{j \in J} P(A_j).$$

2.9 Beispiel. Beim Würfeln mit zwei Würfeln haben die Ereignisse „Augensumme ist 7“ und “erste Augenzahl ist 6“ jeweils Wahrscheinlichkeit 1/6 unter Gleichverteilung. Der Schnitt der beiden Ereignisse ist „erste Augenzahl ist 6, zweite Augenzahl ist 1“ und hat Wahrscheinlichkeit 1/36, so dass die beiden Ereignisse (unter Gleichverteilung) unabhängig sind.

2.10 Definition. Für eine Folge $(A_n)_{n \geqslant 1}$ von Ereignissen setze

$$\limsup_{n \to \infty} A_n := \bigcap_{m \geqslant 1} \bigcup_{n \geqslant m} A_n = \{\omega \in \Omega \,|\, \omega \in A_n \text{ für unendlich viele } n\}.$$

2.11 Satz (Lemma von Borel-Cantelli)**.** *Für eine Folge $(A_n)_{n \geqslant 1}$ von Ereignissen gilt:*

(a) Aus $\sum_{n \geqslant 1} P(A_n) < \infty$ folgt $P(\limsup_{n \to \infty} A_n) = 0$.

(b) Gilt $\sum_{n \geqslant 1} P(A_n) = \infty$ und ist die Folge $(A_n)_{n \geqslant 1}$ unabhängig, so folgt $P(\limsup_{n \to \infty} A_n) = 1$.

2.12 Beispiele.

(a) Ist A ein Ereignis mit $P(A) \in (0,1)$, so gilt $\sum_{n \geqslant 1} P(A_n) = \infty$ für $A_n := A$, jedoch $P(\limsup_{n \to \infty} A_n) = P(A) < 1$. Auf die Unabhängigkeit in Teil (b) des Lemmas von Borel-Cantelli kann also nicht verzichtet werden.

(b) Im unendlichen Münzwurfexperiment bezeichne $A_n^M := \{\omega \in \Omega \,|\, \omega_n = \omega_{n+1} = \cdots = \omega_{n+M-1} = 1\}$ das Ereignis eines M-runs von Einsen ('Kopf') ab Wurf n. Dann ist die Familie $(A_{kM}^M)_{k \geqslant 1}$ unabhängig (auf dem entsprechenden Münzwurfmodell in den Übungen) mit $P(A_{kM}^M) = 2^{-M}$. Aus Teil (b) des Lemmas von Borel-Cantelli folgt daher $P(\limsup_{k\to\infty} A^M_{kM}) = 1$ für jedes $M \in \mathbb{N}$. Dies impliziert $P(\limsup_{n\to\infty} A^M_n) = 1$. Es gilt sogar $P(\bigcap_{M\geqslant 1} \limsup_{n\to\infty} A^M_n) = 1$ („Abzählbarer Schnitt von Einsmengen ist wiederum Einsmenge").

2.13 Definition. Es seien $\mathscr{M}_i \subseteq \mathscr{F}$, $i \in I$, Mengen von Ereignissen. Dann heißt $(\mathscr{M}_i)_{i\in I}$ unabhängig, falls für jede beliebige Auswahl von Ereignissen $A_i \in \mathscr{M}_i$ die Familie $(A_i)_{i\in I}$ unabhängig ist.

2.3 Unabhängige Zufallsvariablen und σ-Algebren

2.14 Definition. Eine Familie $(X_i)_{i\in I}$ von $(S_i, \mathscr{S}_i)$-wertigen Zufallsvariablen heißt unabhängig, falls für jede beliebige Wahl von $A_i \in \mathscr{S}_i$ die Familie von Ereignissen $(\{X_i \in A_i\})_{i\in I}$ unabhängig ist. Äquivalent ist die Familie $(X_i)_{i\in I}$ unabhängig, falls die von X_i erzeugten σ-Algebren $\mathscr{F}^{X_i} = \{X_i^{-1}(A) \mid A \in \mathscr{S}_i\}$, $i \in I$, unabhängig sind.

2.15 Lemma. *Sind $(X_i)_{i\in I}$ eine Familie unabhängiger $(S_i, \mathscr{S}_i)$-wertiger Zufallsvariablen und $g_i : S_i \to T_i$ $(\mathscr{S}_i, \mathscr{T}_i)$-messbare Funktionen, so besteht auch die Familie $(g_i(X_i))_{i\in I}$ aus unabhängigen Zufallsvariablen.*

2.16 Satz. *Es seien $(X_i)_{i\in I}$ eine Familie von Zufallsvariablen auf $(\Omega, \mathscr{F}, P)$ mit Werten in $(S_i, \mathscr{S}_i)$ und $\mathscr{E}_i$ $\cap$-stabile Erzeuger von $\mathscr{S}_i$, $i \in I$. Dann ist $(X_i)_{i\in I}$ bereits unabhängig, falls $(\{X_i \in A_i\})_{i\in I}$ unabhängig ist für beliebige $A_i \in \mathscr{E}_i$.*

2.17 Beispiel. Ist $(A_i)_{i\in I}$ eine Familie unabhängiger Ereignisse, so sind $X_i = \mathbf{1}_{A_i}$, $i \in I$, unabhängige $\{0,1\}$-wertige Zufallsvariablen. Zum Nachweis reicht es, jeweils den $\cap$-stabilen Erzeuger $\mathscr{E} = \{1\}$ der Potenzmenge $\mathscr{P}(\{0,1\})$ zu betrachten.

2.18 Korollar. *Sind $(A_i)_{i\in I}$ unabhängige Ereignisse, so sind auch die erzeugten σ-Algebren $\mathscr{F}_i := \{\varnothing, \Omega, A_i, A_i^c\}$, $i \in I$, unabhängig.*

2.19 Korollar. *Es seien $X_1, \dots, X_n$ Zufallsvariablen auf $(\Omega, \mathscr{F}, P)$.*

(a) Sind X_k diskret-verteilte S_k-wertige Zufallsvariablen, so sind $X_1, \dots, X_n$ genau dann unabhängig, wenn gilt

$$p^{(X_1,\dots,X_n)}(s_1, \dots, s_n) = \prod_{k=1}^n p^{X_k}(s_k) \text{ für alle } s_k \in S_k.$$

(b) Hat jedes X_k Werte in $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$, so sind $X_1, \dots, X_n$ genau dann unabhängig, wenn gilt

$$P(X_1 \leqslant b_1, \dots, X_n \leqslant b_n) = \prod_{k=1}^n P(X_k \leqslant b_k) \text{ für alle } b_k \in \mathbb{R}.$$

2.20 Beispiel. Beim Würfelwurf mit zwei Würfeln sind $X_i : \Omega \to \{1, \dots, 6\}$ mit $X_i(\omega_1, \omega_2) = \omega_i$ für $i \in \{1,2\}$ unabhängige Zufallsvariablen (unter Gleichverteilung). Dazu reicht es, für $k_1, k_2 \in \{1, \dots, 6\}$ für die entsprechenden Zähldichten $p^{X_1}(k_1) = p^{X_2}(k_2) = 1/6$ sowie $p^{(X_1,X_2)}(k_1, k_2) = 1/36$ nachzuprüfen und Teil (a) des Korollars anzuwenden.

2.21 Satz. *Es sei $X = (X_1, \dots, X_n)$ ein Zufallsvektor auf $(\Omega, \mathscr{F}, P)$ mit Dichte $f^X : \mathbb{R}^n \to [0, \infty)$. Dann gilt:*

(a) *Jedes X_k besitzt eine Dichte, die sogenannte Randdichte*

$$f^{X_k}(x_k) := \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \dots dx_{k-1} dx_{k+1} \dots dx_n, \quad x_k \in \mathbb{R}\,.$$

(b) *Die Zufallsvariablen $X_1, \dots, X_n$ sind genau dann unabhängig, wenn gilt*

$$f^X(x_1, \dots, x_n) = \prod_{k=1}^{n} f^{X_k}(x_k) \text{ für Lebesgue-fast alle } x_1, \dots, x_n \in \mathbb{R}\,.$$

2.22 Beispiel. Sind $f_1, \dots, f_n$ Dichten auf $\mathbb{R}$, so definiert die Produktdichte $f(x) = \prod_{i=1}^n f_i(x_i)$ ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}^n, \mathscr{B}_{\mathbb{R}^n})$. Die Koordinatenprojektionen $X_i : \mathbb{R}^n \to \mathbb{R}$ mit $X_i(x_1, \dots, x_n) = x_i$, $i = 1, \dots, n$, sind Borel-messbar, also Zufallsvariablen. Ihre Verteilung P^{X_i} ist gegeben durch die Dichte f_i, ihre gemeinsame Verteilung $P^{(X_1, \dots, X_n)}$ durch die Dichte f. Wir haben damit einen Wahrscheinlichkeitsraum konstruiert mit unabhängigen Zufallsvariablen $(X_i)_{1 \leqslant i \leqslant n}$, deren Verteilung P^{X_i} jeweils durch f_i bestimmt ist.

Die Konstruktion einer Folge $(X_i)_{i \in \mathbb{N}}$ unabhängiger Zufallsvariablen mit vorgeschriebener Randverteilung ist hingegen mit größerem Aufwand verbunden.

2.23 Definition. Es seien $(\Omega_i, \mathscr{F}_i, P_i)_{i \in I}$, $I \neq \varnothing$ beliebige Indexmenge, Wahrscheinlichkeitsräume. Setze $\Omega := \prod_{i \in I} \Omega_i$ (kartesisches Produkt) und definiere mittels der Koordinatenprojektionen $\pi_i : \Omega \to \Omega_i$, $\pi_i((\omega_j)_{j \in I}) = \omega_i$, über Ω die Produkt-σ-Algebra

$$\mathscr{F} := \bigotimes_{i \in I} \mathscr{F}_i := \sigma\Big(\bigcup_{i \in I} \{ \pi_i^{-1}(A_i) \mid A_i \in \mathscr{F}_i \} \Big).$$

Die Produkt-σ-Algebra $\bigotimes_{i \in I} \mathscr{F}_i$ ist also die kleinste σ-Algebra, so dass alle π_i messbar sind.

Gilt für ein Wahrscheinlichkeitsmaß P auf $\mathscr{F}$

$$\forall\, J \subseteq I \text{ endlich}, A_j \in \mathscr{F}_j : P\Big(\bigcap_{j \in J} \pi_j^{-1}(A_j) \Big) = \prod_{j \in J} P_j(A_j),$$

so heißt P Produktmaß, Schreibweise $P = \bigotimes_{i \in I} P_i$.

2.24 Satz. *Ein solches Produktmaß existiert stets und ist eindeutig.*

2.25 Korollar. *Zu vorgegebenen Wahrscheinlichkeitsmaßen P_i auf $(\Omega_i, \mathscr{F}_i)$, $i \in I$, existiert ein Wahrscheinlichkeitsraum mit einer Familie unabhängiger $(\Omega_i, \mathscr{F}_i)$-wertiger Zufallsvariablen $(X_i)_{i \in I}$, deren Verteilung P_i ist.*

2.26 Definition. Es sei $(X_k)_{k \geqslant 1}$ eine Folge von Zufallsvariablen auf $(\Omega, \mathscr{F}, P)$ mit Werten in $(S_k, \mathscr{S}_k)$. Ein Ereignis $A \in \mathscr{F}$ heißt asymptotisch bezüglich (X_k), falls es für alle $n \geqslant 1$ nur von $(X_k, k \geqslant n)$ abhängt in dem Sinne, dass $A \in \mathscr{A}_X$ gilt. Hierbei ist die asymptotische σ-Algebra $\mathscr{A}_X$ definiert als

$$\mathscr{A}_X := \bigcap_{n \geqslant 1} \sigma\Big(\bigcup_{k \geqslant n} \mathscr{F}^{X_k} \Big).$$

2.27 Beispiel. Es seien X_k, $k \in \mathbb{N}$, reellwertige Zufallsvariablen sowie A das Ereignis, dass $\lim_{n\to\infty} \sum_{k=1}^{n} \frac{X_k}{k}$ existiert. Dann gilt nach dem Cauchy-Kriterium

$$A = \bigcap_{\varepsilon>0,\, \varepsilon\in\mathbb{Q}} \bigcup_{N\in\mathbb{N}} B_N \text{ mit } B_N = \bigcap_{m\geqslant n\geqslant N} \Big\{ \sum_{k=n}^{m} \frac{X_k}{k} \in (-\varepsilon, \varepsilon) \Big\}.$$

Nun ist $B_N \subseteq B_{N+1}$ sowie $B_N \in \sigma(\bigcup_{k\geqslant N} \mathscr{F}^{X_k})$. Daher gilt $\bigcup_{N=1}^{N_{max}} B_N = B_{N_{max}} \in \sigma(\bigcup_{k\geqslant n} \mathscr{F}^{X_k})$ für alle $N_{max} \geqslant n$. Dies impliziert $\bigcup_{N=1}^{\infty} B_N \in \sigma(\bigcup_{k\geqslant n} \mathscr{F}^{X_k})$ für alle $n \in \mathbb{N}$ und somit, dass A ein asymptotisches Ereignis ist. Insbesondere ist $\mathscr{A}_X$ nicht leer.

2.28 Satz (0-1-Gesetz von Kolmogorov)**.** *Es seien $(X_k)_{k\geqslant 1}$ unabhängige Zufallsvariablen auf $(\Omega, \mathscr{F}, P)$. Dann gilt für jedes bezüglich (X_k) asymptotische Ereignis A: $P(A) = 0$ oder $P(A) = 1$.*

2.29 Lemma. *Es seien $(X_i)_{i\in I}$ eine Familie unabhängiger Zufallsvariablen mit Werten in $(S_i, \mathscr{S}_i)$ und $I = I_1 \cup I_2$ eine disjunkte Zerlegung von I. Dann sind die σ-Algebren $\mathscr{F}_1 := \sigma(\bigcup_{i\in I_1} \mathscr{F}^{X_i})$ und $\mathscr{F}_2 := \sigma(\bigcup_{i\in I_2} \mathscr{F}^{X_i})$ unabhängig.*

2.30 Beispiel (Harmonische Reihe mit zufälligen Vorzeichen)**.** Sind $(X_k)_{k\geqslant 1}$ unabhängige $\{-1, 1\}$-wertige Zufallsvariablen, so konvergiert die Reihe $\sum_{k=1}^{\infty} X_k \frac{1}{k}$ entweder mit Wahrscheinlichkeit 1 oder sie divergiert mit Wahrscheinlichkeit 1. Diese Aussage gilt für jede beliebige Randverteilung P^{X_k} der Zufallsvariablen X_k, $k \in \mathbb{N}$.

2.4 Faltung

2.31 Definition. Sind P, Q Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$, so ist die Faltung $P * Q$ definiert als das Wahrscheinlichkeitsmaß(!)

$$(P * Q)(B) = \int_\mathbb{R} P(B - \{x\})\, Q(dx), \quad B \in \mathfrak{B}_\mathbb{R}, \text{ mit } B - \{x\} = \{b - x \,|\, b \in B\}.$$

2.32 Satz. *Es seien X und Y unabhängige reellwertige Zufallsvariablen. Dann besitzt $X + Y$ die Verteilung $P^{X+Y} = P^X * P^Y$.*

2.33 Korollar. *Die Faltung ist kommutativ und assoziativ.*

2.34 Korollar. *Besitzen P und Q Zähldichten p bzw. q auf $\mathbb{Z}$ (auf $\mathbb{N}_0$), so besitzt $P*Q$ die Zähldichte $(p*q)(k) := \sum_{m\in\mathbb{Z}} p(k-m)q(m)$ (auf $\mathbb{N}_0$: $(p*q)(k) := \sum_{m=0}^{k} p(k-m)q(m)$).*

2.35 Beispiel. Ist X Poiss(λ_x)- und Y Poiss(λ_y)-verteilt und sind X und Y unabhängig, so ist $X + Y$ Poiss$(\lambda_x + \lambda_y)$-verteilt.

2.36 Satz. *Es seien X und Y unabhängige reellwertige Zufallsvariablen und X besitze eine Dichte f^X. Dann besitzt $X + Y$ die Dichte*

$$f^{X+Y}(z) = \int_\mathbb{R} f^X(z-y)\, P^Y(dy), \quad z \in \mathbb{R}\,.$$

Falls auch Y eine Dichte besitzt, so gilt

$$f^{X+Y}(z) = f^X * f^Y(z) := \int_\mathbb{R} f^X(z-y) f^Y(y)\, dy, \quad z \in \mathbb{R}\,.$$

2.37 Beispiel. Für $\lambda, p > 0$ definiere die Gamma-Verteilung $\Gamma(\lambda, p)$ über die Dichte

$$f_{\lambda,p}(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} \mathbf{1}_{(0,\infty)}(x), \quad x \in \mathbb{R},$$

mit der Gamma-Funktion $\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt$, so dass $\Gamma(n+1) = n!$ für $n \in \mathbb{N}_0$ und $\Gamma(1/2) = \sqrt{\pi}$. Spezialfälle sind $\Gamma(\lambda, 1) = \operatorname{Exp}(\lambda)$, $\Gamma(1/2, 1/2) = \chi^2(1)$.

Setzt man zusätzlich $\Gamma(\lambda, 0) = \delta_0$, so bildet $(\Gamma(\lambda, p))_{p \geqslant 0}$ für festes $\lambda > 0$ eine Faltungshalbgruppe, d.h. es gilt $\Gamma(\lambda, p_1) * \Gamma(\lambda, p_2) = \Gamma(\lambda, p_1 + p_2)$. Insbesondere ist die Summe von n unabhängigen $\chi^2(1)$-verteilten Zufallsvariablen (äquivalent: die quadrierte Norm $\|Z\|^2$ eines standard-normalverteilten Zufallsvektors $Z \sim N(0, E_n)$ im $\mathbb{R}^n$) gemäß $\chi^2(n) := \Gamma(1/2, n/2)$-verteilt ($\chi^2$-Verteilung mit n Freiheitsgraden). Im Spezialfall $n = 2$ ergibt sich $\chi^2(2) = \operatorname{Exp}(1/2)$, was für die Box-Müller-Methode zur Simulation der Normalverteilung genutzt wird.

3 Erwartungswert, Varianz und Kovarianz

3.1 Erwartungswert und Momente

3.1 Definition. Eine reellwertige Zufallsvariable X auf $(\Omega, \mathscr{F}, P)$ heißt einfach, falls sie nur endlich viele Werte annimmt, d.h. es folgende Darstellung gibt:

$$X = \sum_{i=1}^m \alpha_i \mathbf{1}_{A_i} \text{ mit } m \in \mathbb{N},\ \alpha_i \in \mathbb{R},\ A_i \in \mathscr{F}.$$

Für eine solche Zufallsvariable definieren wir ihren Erwartungswert als

$$\mathbb{E}[X] := \sum_{i=1}^m \alpha_i P(A_i).$$

3.2 Beispiel. Beim Würfeln mit zwei fairen Würfeln ergibt sich für die Augensumme S

$$\mathbb{E}[S] = 2P(S=2) + 3P(S=3) + \dots + 12P(S=12) = 7.$$

3.3 Lemma. *Für eine einfache Zufallsvariable X auf $(\Omega, \mathscr{F}, P)$ gilt:*

(a) $\mathbb{E}[X] = \sum_{x \in X(\Omega)} x P(X = x)$; *insbesondere hängt der Erwartungswert nur von der Verteilung P^X von X ab.*

(b) *Der Erwartungswert ist linear und monoton: ist Y eine weitere einfache Zufallsvariable und sind $\alpha, \beta \in \mathbb{R}$, so gilt*

$$\mathbb{E}[\alpha X + \beta Y] = \alpha\, \mathbb{E}[X] + \beta\, \mathbb{E}[Y];$$

aus $X \leqslant Y$ (d.h. $\forall \omega \in \Omega:\ X(\omega) \leqslant Y(\omega)$) folgt $\mathbb{E}[X] \leqslant \mathbb{E}[Y]$.

(c) *Falls X und Y unabhängige einfache Zufallsvariablen sind, so gilt $\mathbb{E}[X \bullet Y] = \mathbb{E}[X] \bullet \mathbb{E}[Y]$.*

(d) *Für jedes $A \in \mathscr{F}$ gilt $\mathbb{E}[\mathbf{1}_A] = P(A)$.*

3.4 Definition. Es sei $X \geqslant 0$ eine nichtnegative Zufallsvariable. Sind dann X_n einfache nichtnegative Zufallsvariablen mit $X_n(\omega) \uparrow X(\omega)$ für $n \to \infty$ und alle $\omega \in \Omega$, so definiere den <u>Erwartungswert</u>

$$\mathbb{E}[X] := \lim_{n\to\infty} \mathbb{E}[X_n] \ \in [0, +\infty]$$

(man kann zeigen, dass eine solche Folge (X_n) stets existiert und $\mathbb{E}[X]$ nicht von der Auswahl der X_n abhängt).

Betrachte nun auf dem Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, P)$ die Menge der Zufallsvariablen

$$\mathscr{L}^1 := \mathscr{L}^1(\Omega, \mathscr{F}, P) := \{X : \Omega \to \mathbb{R} \quad \text{messbar} \mid \mathbb{E}[|X|] < \infty\}.$$

Dann definiere für $X \in \mathscr{L}^1$ mit $X_+ := \max(X, 0)$, $X_- := \max(-X, 0)$ den <u>Erwartungswert</u> als

$$\mathbb{E}[X] := \mathbb{E}[X_+] - \mathbb{E}[X_-] \in \mathbb{R}\,.$$

Der Erwartungswert $\mathbb{E}[X]$ ist also das <u>Lebesgueintegral</u> von X bezüglich P, und man schreibt $\mathbb{E}[X] = \int X\,dP = \int_\Omega X(\omega)\,P(d\omega)$ sowie $\int_A X\,dP = \int_\Omega X(\omega)\mathbf{1}_A(\omega)\,P(d\omega)$ für $A \in \mathscr{F}$.

3.5 Definition. Ist $(\Omega, \mathscr{F}, \mu)$ ein Maßraum, so heißt eine messbare Funktion $f : \Omega \to [0, \infty)$ <u>μ-Dichte</u> oder genauer <u>μ-Wahrscheinlichkeitsdichte</u>, falls $\int_\Omega f\,d\mu = 1$ gilt. Für Maße μ, ν auf $(\Omega, \mathscr{F})$ heißt ν <u>absolut-stetig</u> bezüglich μ (Notation: $\nu \ll \mu$), falls $\mu(A) = 0$ für $A \in \mathscr{F}$ impliziert $\nu(A) = 0$.

3.6 Lemma. *Ist f μ-Dichte auf $(\Omega, \mathscr{F}, \mu)$, so definiert $P(A) = \int_A f\,d\mu$, $A \in \mathscr{F}$, ein Wahrscheinlichkeitsmaß mit $P \ll \mu$. Ist μ das Zählmaß auf Ω, so sind μ-Dichten gerade Zähldichten; ist μ das Lebesguemaß auf dem $\mathbb{R}^d$, so sind μ-Dichten gerade Dichtefunktionen im $\mathbb{R}^d$.*

3.7 Satz. *Es sei X ein Zufallsvariable mit Werten in $(S, \mathscr{S})$ und $h : S \to \mathbb{R}$ Borel-messbar. Dann gilt*

(a) $h(X) \in \mathscr{L}^1 \iff \int_S |h(x)|\,P^X(dx) < \infty$.

(b) Für $h(X) \in \mathscr{L}^1$ gilt

$$\mathbb{E}[h(X)] = \int_S h(x)\,P^X(dx).$$

(c) Ist X ein Zufallsvektor im $\mathbb{R}^d$ mit Dichte f^X, so gilt $h(X) \in \mathscr{L}^1 \iff \int_{\mathbb{R}^d} |h(x)| f^X(x)\,dx < \infty$. In dem Fall ist

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h(x) f^X(x)\,dx.$$

(d) Ist X diskret verteilt auf $\mathbb{Z}$ mit Zähldichte p^X, so gilt $h(X) \in \mathscr{L}^1 \iff \sum_{k\in\mathbb{Z}} |h(k)| p^X(k) < \infty$. In dem Fall ist

$$\mathbb{E}[h(X)] = \sum_{k\in\mathbb{Z}} h(k) p^X(k).$$

3.8 Beispiel. Für $X \sim N(0,1)$ gilt mit partieller Integration unter Benutzung von $(e^{-x^2/2})' = -xe^{-x^2/2}$

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2\, P^X(dx) = \frac{-1}{\sqrt{2\pi}} \int_{\mathbb{R}} x\big(-xe^{-x^2/2}\big)dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-x^2/2}dx = 1.$$

Alternativ verwende man, dass $Y = X^2$ $\chi^2(1)$-verteilt ist, so dass $\mathbb{E}[X^2] = \int_0^\infty y(2\pi y)^{-1/2}e^{-y/2}dy = 1$.

3.9 Satz. *Für $X \in \mathscr{L}^1(\Omega, \mathscr{F}, P)$ gilt:*

(a) $\mathbb{E}[X] = \int_{\mathbb{R}} x\, P^X(dx)$; insbesondere hängt der Erwartungswert nur von der Verteilung P^X von X ab.

(b) Der Erwartungswert ist linear: ist Y eine weitere Zufallsvariable in $\mathscr{L}^1$ und sind $\alpha, \beta \in \mathbb{R}$, so gilt

$$\mathbb{E}[\alpha X + \beta Y] = \alpha\, \mathbb{E}[X] + \beta\, \mathbb{E}[Y].$$

(c) Der Erwartungswert ist monoton: ist Y eine weitere Zufallsvariable in $\mathscr{L}^1$ mit $X \leqslant Y$, so gilt $\mathbb{E}[X] \leqslant \mathbb{E}[Y]$. Aus $X \leqslant Y$ und $\mathbb{E}[X] = \mathbb{E}[Y]$ folgt $P(X = Y) = 1$.

(d) Falls $X, Y \in \mathscr{L}^1$ unabhängig sind, so gilt $X \bullet Y \in \mathscr{L}^1$ und $\mathbb{E}[X \bullet Y] = \mathbb{E}[X] \bullet \mathbb{E}[Y]$.

3.10 Definition. Wir sagen, dass eine Zufallsvariable X in $\mathscr{L}^p = \mathscr{L}^p(P)$ liegt für $p > 0$, falls $|X|^p \in \mathscr{L}^1$, also $\mathbb{E}[|X|^p] < \infty$ gilt. Für $X \in \mathscr{L}^p$ und $p \in \mathbb{N}$ heißt $\mathbb{E}[X^p]$ das <u>p-te Moment</u> von X; für $X \in \mathscr{L}^p$ und $p > 0$ heißt $\mathbb{E}[|X|^p]$ das <u>p-te absolute Moment</u> von X.

3.11 Satz. *Für $X \in \mathscr{L}^p$ und $Y \in \mathscr{L}^q$ mit $\frac{1}{p} + \frac{1}{q} = 1$ gelten $XY \in \mathscr{L}^1$ und die <u>Hölder-Ungleichung</u>*

$$|\mathbb{E}[XY]| \leqslant \mathbb{E}[|X|^p]^{1/p}\, \mathbb{E}[|Y|^q]^{1/q}.$$

Insbesondere gelten $XY \in \mathscr{L}^1(P)$ für $X, Y \in \mathscr{L}^2(P)$ und die <u>Cauchy-Schwarz-Ungleichung</u>

$$|\mathbb{E}[XY]| \leqslant \mathbb{E}[X^2]^{1/2}\, \mathbb{E}[Y^2]^{1/2}.$$

3.12 Korollar. *Für $0 < p \leqslant q$ gelten $\mathscr{L}^q \subseteq \mathscr{L}^p$ und $\mathbb{E}[|X|^p] \leqslant \mathbb{E}[|X|^q]^{p/q}$ für $X \in \mathscr{L}^q$.*

3.13 Satz. *Für eine Zufallsvariable $X \in \mathscr{L}^2$ gilt die <u>Bias-Varianz-Zerlegung</u>*

$$\forall x \in \mathbb{R}:\ \mathbb{E}[(X-x)^2] = \underbrace{(\mathbb{E}[X]-x)^2}_{\mathit{Bias}^2} + \underbrace{\mathbb{E}[(X-\mathbb{E}[X])^2]}_{\mathit{Varianz}}.$$

Die Funktion $\varphi(x) = \mathbb{E}[(X-x)^2]$ nimmt ihr Minimum auf $\mathbb{R}$ genau bei $x = \mathbb{E}[X]$ an.

3.14 Definition. Es sei $I \subseteq \mathbb{R}$ ein (ggf. auch unendliches) Intervall. Eine Funktion $\varphi : I \to \mathbb{R}$ heißt konvex, falls

$$\forall x, y \in I,\ \alpha \in [0,1] :\ \varphi(\alpha x + (1-\alpha)y) \leqslant \alpha\varphi(x) + (1-\alpha)\varphi(y).$$

3.15 Beispiel. Ist $\varphi : I \to \mathbb{R}$ stetig differenzierbar mit monoton wachsender Ableitung φ' (z.B. wenn $\varphi'' \geqslant 0$ auf I), so ist φ konvex. Insbesondere sind $\varphi(x) = e^{\alpha x}$ für $\alpha \in \mathbb{R}$ beliebig und $\varphi(x) = |x|^p$ für $p > 1$ konvex auf $I = \mathbb{R}$. Auch $\varphi(x) = |x|$ ist konvex auf $\mathbb{R}$.

3.16 Lemma. *Ist $\varphi : I \to \mathbb{R}$ konvex, so existieren für alle $x \in int(I)$ (Inneres von I) die links- und rechtsseitigen Ableitungen $\varphi'(x-)$, $\varphi'(x+)$ mit $\varphi'(x-) \leqslant \varphi'(x+)$. Es gilt weiterhin*

$$\forall x \in int(I),\ y \in I :\ \varphi(y) \geqslant \varphi(x) + \varphi'(x+)(y-x).$$

3.17 Satz. *Es sei $X \in \mathscr{L}^1$ mit Werten in einem offenen Intervall I. Dann gilt $\mathbb{E}[X] \in I$. Ist $\varphi : I \to \mathbb{R}$ konvex und $\varphi(X) \in \mathscr{L}^1$, so gilt die Jensensche Ungleichung*

$$\mathbb{E}[\varphi(X)] \geqslant \varphi(\mathbb{E}[X]).$$

3.18 Beispiel. Für $q > p > 0$ ist $\varphi(x) = |x|^{q/p}$ konvex, und die Jensensche Ungleichung liefert ebenfalls $\mathbb{E}[|X|^p] \leqslant \mathbb{E}[|X|^q]^{p/q}$ für $X \in \mathscr{L}^p \cap \mathscr{L}^q$.

3.2 Varianz, Kovarianz und Korrelation

3.19 Definition. Für eine Zufallsvariable $X \in \mathscr{L}^2$ bezeichnet

$$\operatorname{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$$

die Varianz von X. $\sigma(X) := \sqrt{\operatorname{Var}(X)}$ heißt Standardabweichung von X.

3.20 Satz (Eigenschaften der Varianz)**.** *Für $X, Y \in \mathscr{L}^2$ gilt:*

(a) $\operatorname{Var}(X) = 0 \iff P(X = \mathbb{E}[X]) = 1$;

(b) $\forall a, b \in \mathbb{R} :\ \operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X)$;

(c) $\operatorname{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$;

(d) $\operatorname{Var}(X + Y) \leqslant 2\operatorname{Var}(X) + 2\operatorname{Var}(Y)$;

(e) falls X, Y unabhängig sind, so gilt $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$.

3.21 Satz (Beste lineare Vorhersage)**.** *Es seien X, Y Zufallsvariablen in $\mathscr{L}^2$ sowie*

$$L_X := \{aX + b \,|\, a, b \in \mathbb{R}\} \subseteq \mathscr{L}^2$$

die Menge der auf linear-affinen Funktionen von X basierenden Zufallsvariablen. Dann nimmt der mittlere quadratische Fehler

$$\varphi : L_X \to [0, \infty), \quad \varphi(Z) := \mathbb{E}[(Y - Z)^2]$$

sein Minimum bei $Z = a^*X + b^*$ an mit

$$a^* = \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{\operatorname{Var}(X)}, \quad b^* = \mathbb{E}[Y] - a^* \,\mathbb{E}[X]$$

(a^* beliebig falls $\operatorname{Var}(X) = 0$*). Für* $\operatorname{Var}(X), \operatorname{Var}(Y) > 0$ *gilt*

$$\varphi(a^*X + b^*) = \operatorname{Var}(Y) - \operatorname{Cov}(X,Y)^2/\operatorname{Var}(X) = \operatorname{Var}(Y)(1 - \rho^2(X,Y))$$

mit nachfolgend definierter Kovarianz $\operatorname{Cov}(X,Y)$ *und Korrelation* $\rho(X,Y)$*.*

3.22 Definition. Für Zufallsvariablen $X, Y \in \mathscr{L}^2$ definiert

$$\operatorname{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

die Kovarianz zwischen X und Y. Falls $\sigma(X) > 0$ und $\sigma(Y) > 0$ gilt, heißt

$$\rho(X,Y) := \frac{\operatorname{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$$

die Korrelation zwischen X und Y. Falls $\operatorname{Cov}(X,Y) = 0$ gilt, heißen X und Y unkorreliert.

3.23 Satz (Eigenschaften von Kovarianz und Korrelation)**.** *Für $X, Y, Z \in \mathscr{L}^2$ gilt:*

(a) $\operatorname{Cov}(X,Y) = \operatorname{Cov}(Y,X) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$*,* $\operatorname{Cov}(X,X) = \operatorname{Var}(X)$*;*

(b) $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y)$*;*

(c) $\forall a, b \in \mathbb{R}:\ \operatorname{Cov}(aX + b, Y) = a \operatorname{Cov}(X,Y)$*;*

(d) $\operatorname{Cov}(X + Y, Z) = \operatorname{Cov}(X,Z) + \operatorname{Cov}(Y,Z)$*;*

(e) X, Y unabhängig $\Rightarrow$ X, Y unkorreliert;

(f) $|\operatorname{Cov}(X,Y)| \leqslant \sigma(X)\sigma(Y)$ *und* $\rho(X,Y) \in [-1, +1]$*.*

3.24 Beispiele.

(a) Eine $\operatorname{Bin}(n,p)$-verteilte Zufallsvariable X ergibt sich als $X = \sum_{i=1}^n X_i$ mit einem Bernoullischema (X_i), d.h. $(X_i)_{i=1,\dots,n}$ unabhängig mit $P(X_i = 1) = p$, $P(X_i = 0) = 1-p$. Damit folgt $X \in \mathscr{L}^2$ und $\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = np$, $\operatorname{Var}(X) = \sum_{i=1}^n \operatorname{Var}(X_i) = np(1-p)$. Da Erwartungswert und Varianz nur von der Verteilung abhängen, sagt man, dass die $\operatorname{Bin}(n,p)$-Verteilung Erwartungswert np und Varianz $np(1-p)$ besitzt. Im Fall $p \in \{0,1\}$ gilt also $\operatorname{Var}(X) = 0$ sowie stets $\operatorname{Var}(X) \leqslant n/4$. Die *relative Häufigkeit* von Erfolgen $A = X/n$ erfüllt $\mathbb{E}[A] = p$ (*erwartungstreu*) und $\operatorname{Var}(A) = \frac{p(1-p)}{n}$.

(b) Für eine $N(\mu, \sigma^2)$-verteilte Zufallsvariable X mit $\mu \in \mathbb{R}$, $\sigma > 0$ gilt nach dem Dichtetransformationssatz, dass $Z = (X-\mu)/\sigma$ $N(0,1)$-verteilt ist. Nun sind $X, Z \in \mathscr{L}^2$ und wir schließen $\operatorname{Var}(X) = \operatorname{Var}(\sigma Z + \mu) = \sigma^2 \operatorname{Var}(Z) = \sigma^2$.

(c) Bezeichnen X_1 und X_2 die Augenzahlen beim Wurf zweier Würfel, so ist $S = X_1 + X_2$ die Augensumme und $D = X_1 - X_2$ die Augendifferenz. Es gilt

$$\operatorname{Cov}(S, D) = \operatorname{Var}(X_1) + \operatorname{Cov}(X_2, X_1) - \operatorname{Cov}(X_1, X_2) - \operatorname{Var}(X_2) = 0,$$

und S und D sind unkorreliert (gilt allgemein für $X_1, X_2 \in \mathscr{L}^2$ mit $\operatorname{Var}(X_1) = \operatorname{Var}(X_2)$). Allerdings sind S und D *nicht* unabhängig; denn es gilt beispielsweise $P(S = 2, D = 5) = 0$, aber $P(S = 2)P(D = 5) > 0$. Die beste (nicht nur lineare) Vorhersage von D gegeben S ist gerade konstant $\mathbb{E}[D] = 0$, da D unter jeder Bedingung $\{S = k\}$ symmetrisch um 0 verteilt ist: $P(D = m \,|\, S = k) = P(D = -m \,|\, S = k)$.

3.3 Mehrdimensionale Normalverteilung

3.25 Definition. Es seien $\mu \in \mathbb{R}^d$ ein Vektor sowie $\Sigma \in \mathbb{R}^{d \times d}$ eine symmetrische, positiv semi-definite Matrix. Ein Zufallsvektor X im $\mathbb{R}^d$ ist $N(\mu, \Sigma)$-verteilt, falls $X = \mu + \Sigma^{1/2} Y$ gilt mit einem standard-normalverteilten Zufallsvektor Y im $\mathbb{R}^d$. $N(\mu, \Sigma)$ heißt d-dimensionale Normalverteilung mit Mittelwertvektor μ und Kovarianzmatrix Σ. Im regulären Fall einer invertierbaren Matrix Σ besitzt $N(\mu, \Sigma)$ nach obiger Herleitung die Dichte

$$\varphi_{\mu,\Sigma}(x) = (2\pi)^{-d/2} \det(\Sigma)^{-1/2} \exp\Big(-\frac{1}{2} \langle \Sigma^{-1}(x - \mu), x - \mu \rangle \Big), \quad x \in \mathbb{R}^d.$$

3.26 Beispiel. Ist Σ nicht invertierbar, so auch $\Sigma^{1/2}$, und $Y = \mu + \Sigma^{1/2} Y$ nimmt Werte in einem echten (affinen) Unterraum des $\mathbb{R}^d$ an, dessen Lebesguemaß Null ist. In diesem Fall besitzt $N(\mu, \Sigma)$ also keine Dichte. Im Eindimensionalen gilt $N(\mu, 0) = \delta_\mu$.

3.27 Lemma. *Für einen $N(\mu, \Sigma)$-verteilten Zufallsvektor $X = (X_1, \dots, X_d)$ und $1 \leqslant k, \ell \leqslant d$ gilt*

$$\mathbb{E}[X_k] = \mu_k, \quad \operatorname{Cov}(X_k, X_\ell) = \Sigma_{k\ell}.$$

3.28 Korollar. *Sind $X_1, \dots, X_n$ gemeinsam normalverteilt (d.h. $(X_1, \dots, X_n)$ ist n-dimensional normalverteilt) und sind $X_1, \dots, X_n$ (paarweise) unkorreliert, so sind $X_1, \dots, X_n$ sogar unabhängig.*

3.29 Lemma. *Ist $O \in \mathbb{R}^{d \times d}$ eine orthogonale Matrix, so gilt für einen standard-normalverteilten Zufallsvektor X im $\mathbb{R}^d$, dass auch OX standard-normalverteilt ist.*

3.30 Satz. *Ist X ein $N(\mu, \Sigma)$-verteilter Zufallsvektor im $\mathbb{R}^d$ und ist $A \in \mathbb{R}^{m \times d}$ eine deterministische Matrix, so ist $Y = AX$ ein $N(A\mu, A\Sigma A^\top)$-verteilter Zufallsvektor im $\mathbb{R}^m$.*

Proof. Wir müssen zeigen, dass sich Y darstellen lässt als $Y = A\mu + (A\Sigma A^\top)^{1/2} Z$ mit einer geeigneten Zufallsvariablen $Z \sim N(0, E_m)$. Aus der

Darstellung $X = \mu + \Sigma^{1/2}W$ mit $W \sim N(0, E_d)$ ergibt sich die zu erfüllende Bedingung als

$$A(\mu + \Sigma^{1/2}W) = A\mu + (A\Sigma A^\top)^{1/2}Z, \text{ d.h. } A\Sigma^{1/2}W = (A\Sigma A^\top)^{1/2}Z.$$

Der Satz zur orthogonalen Normalform (z.B. in M. Koecher, Lineare Algebra und Analytische Geometrie, Seite 199) zeigt, dass es orthogonale Matrizen $T_1 \in \mathbb{R}^{m\times m}, T_2 \in \mathbb{R}^{d\times d}$ und eine Diagonalmatrix $D \in \mathbb{R}^{r\times r}$, $r \leqslant \min(m,d)$, mit strikt positiven Diagonaleinträgen gibt, so dass in Blockmatrixnotation (beachte jeweils die Dimensionen!) $A\Sigma^{1/2} = T_1 \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} T_2$ gilt. Dies impliziert

$$\begin{aligned}(A\Sigma A^\top)^{1/2} = (A\Sigma^{1/2}(A\Sigma^{1/2})^\top)^{1/2} &= \Big(T_1 \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}^\top T_1^\top\Big)^{1/2} \\ &= T_1 \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} T_1^\top.\end{aligned}$$

Wir müssen also $Z \sim N(0, E_m)$ finden mit $\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} T_1^\top Z = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} T_2 W$. Setze $Z := T_1\Big(\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} T_2W + W'\Big)$ mit $W' \sim N(0, \begin{pmatrix} 0 & 0 \\ 0 & E_{m-r} \end{pmatrix})$ unabhängig von W, ggf. definiert auf einem größeren Wahrscheinlichkeitsraum (bzw. $W' = 0$ falls $m = r$). Aus dem Lemma folgt $T_2W \sim N(0, E_d)$, weil T_2 orthogonale Matrix ist, und weiter, dass $\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} T_2W$ ein $N(0, \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix})$-verteilter Vektor im $\mathbb{R}^m$ ist (Projektion auf die ersten r-Koordinaten). Daher gilt $\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} T_2W + W' \sim N(0, E_m)$, und es folgt wieder nach dem Lemma $Z \sim N(0, E_m)$.

Schließlich ergibt sich für $(A\Sigma A^\top)^{1/2}Z$

$$T_1 \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \Big(\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} T_2W + W'\Big) = T_1 \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} T_2W = A\Sigma^{1/2}W,$$

wie zu zeigen war. □

3.31 Korollar. *Sind X und Y unabhängig und gemäß $N(\mu_X, \sigma_X^2)$ bzw. $N(\mu_Y, \sigma_Y^2)$ verteilt mit $\mu_X, \mu_Y \in \mathbb{R}$, $\sigma_X, \sigma_Y > 0$, so ist $X + Y$ gemäß $N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$ verteilt. Insbesondere bildet $(N(\mu t, \sigma^2 t))_{t\geqslant 0}$ für jedes $\mu \in \mathbb{R}$, $\sigma \geqslant 0$ eine Faltungshalbgruppe.*

4 Grenzwertsätze

4.1 Gesetze der großen Zahlen

4.1 Satz (Allgemeine Markov-Ungleichung)**.** *Es sei X eine Zufallsvariable und $\varphi : [0,\infty) \to [0,\infty)$ monoton wachsend. Dann gilt für jedes $K > 0$ mit $\varphi(K) > 0$:*

$$P(|X| \geqslant K) \leqslant \frac{\mathbb{E}[\varphi(|X|)]}{\varphi(K)}.$$

4.2 Beispiel. Für $X \in \mathscr{L}^p$ gilt $P(|X| \geqslant K) \leqslant \mathbb{E}[|X|^p]K^{-p}$, $K > 0$.

4.3 Korollar (Tschebyschev-Ungleichung)**.** *Ist X eine Zufallsvariable in $\mathscr{L}^2$, so gilt für jedes $\varepsilon > 0$*

$$P(|X - \mathbb{E}[X]| \geqslant \varepsilon) \leqslant \frac{\operatorname{Var}(X)}{\varepsilon^2}.$$

4.4 Beispiel. Eine faire Münze werde 1000-mal geworfen und X bezeichne die Anzahl der Würfe mit 'Kopf'. Dann gilt

$$P(X \geqslant 550) \leqslant P(|X - 500| \geqslant 50) \leqslant \tfrac{250}{50^2} = 0,1.$$

4.5 Satz (schwaches Gesetz der großen Zahlen)**.** *Es sei $(X_i)_{i\geqslant 1}$ eine Folge unkorrelierter Zufallsvariablen in $\mathscr{L}^2$ mit demselben Erwartungswert $\mu \in \mathbb{R}$ und $\sup_i \operatorname{Var}(X_i) < \infty$. Dann erfüllt das arithmetische Mittel*

$$A_n := \frac{1}{n}\sum_{i=1}^n X_i$$

für jedes $\varepsilon > 0$

$$\lim_{n\to\infty} P(|A_n - \mu| > \varepsilon) = 0.$$

4.6 Korollar. *(Weierstraßscher Approximationssatz) Zur stetigen Funktion $f : [0,1] \to \mathbb{R}$ definiere das zugehörige Bernstein-Polynom n-ten Grades*

$$f_n(x) := \sum_{k=0}^n f\Big(\frac{k}{n}\Big)\binom{n}{k} x^k(1-x)^{n-k}, \quad x \in [0,1].$$

Dann gilt $\lim_{n\to\infty}\|f - f_n\|_\infty = 0$ mit $\|g\|_\infty := \sup_{x\in[0,1]}|g(x)|$.

4.7 Definition. Es seien $(X_n)_{n\geqslant 1}$ und X Zufallsvariablen auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, P)$. Man sagt, dass X_n stochastisch (oder auch in P-Wahrscheinlichkeit) gegen X konvergiert für $n \to \infty$, falls für alle $\varepsilon > 0$ gilt

$$\lim_{n\to\infty} P(|X - X_n| > \varepsilon) = 0.$$

Notation: $X_n \xrightarrow{P} X$.

Man sagt, dass X_n P-fast sicher gegen X konvergiert, kurz $X_n \to X$ P-f.s., falls

$$P(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1.$$

4.8 Beispiel. Im schwachen Gesetz der großen Zahlen gilt $A_n \xrightarrow{P} \mu$.

4.9 Satz. *Fast sichere Konvergenz impliziert stochastische Konvergenz, aber nicht umgekehrt.*

4.10 Satz. *(starkes Gesetz der großen Zahlen) Es sei $(X_i)_{i\geqslant 1}$ eine Folge unkorrelierter Zufallsvariablen in $\mathscr{L}^2$ mit demselben Erwartungswert $\mu \in \mathbb{R}$ und $\sup_i \operatorname{Var}(X_i) < \infty$. Dann konvergiert das arithmetische Mittel $A_n = \frac{1}{n}\sum_{i=1}^n X_i$ fast sicher gegen μ.*

4.11 Definition. Identifiziert man $X, Y \in \mathscr{L}^p(\Omega, \mathscr{F}, P)$ (Zusammenfassung in einer Äquivalenzklasse), wenn $X = Y$ P-fast sicher, d.h. $P(X = Y) = 1$, gilt, so erhält man den Vektorraum $L^p(\Omega, \mathscr{F}, P)$. Mit der Norm $\|X\|_{L^p} = \mathbb{E}[|X|^p]^{1/p}$ wird $L^p(\Omega, \mathscr{F}, P)$ für $p \geqslant 1$ zum Banachraum und mit dem Skalarprodukt $\langle X, Y\rangle = \mathbb{E}[XY]$ wird $L^2(\Omega, \mathscr{F}, P)$ zum Hilbertraum (Beweis in Analysis!). Für eine Folge (X_n) in $\mathscr{L}^p(\Omega, \mathscr{F}, P)$, $p > 0$, und ein $X \in \mathscr{L}^p(\Omega, \mathscr{F}, P)$ sagen wir, dass X_n gegen X in L^p konvergiert, falls $\mathbb{E}[|X_n - X|^p] \to 0$ für $n \to \infty$ gilt.

4.12 Lemma. *Konvergiert (X_n) gegen X in L^p für ein $p > 0$, so auch stochastisch: $X_n \xrightarrow{L^p} X \Rightarrow X_n \xrightarrow{P} X$.*

4.13 Beispiel (Harmonische Reihe mit zufälligen Vorzeichen)**.** Betrachte $S_n = \sum_{k=1}^n X_k \frac{1}{k}$ mit (X_k) unabhängig und $P(X_k = 1) = P(X_k = -1) = 1/2$. Dann gilt $\mathbb{E}[S_n] = 0$, $\operatorname{Var}(S_n) = \sum_{k=1}^n \frac{1}{k^2}$. Wegen $\operatorname{Var}(S_n - S_m) \leqslant \sum_{k \geqslant m} \frac{1}{k^2}$ für $n > m$ und $\lim_{m\to\infty} \sum_{k\geqslant m} \frac{1}{k^2} = 0$ bildet (S_n) eine Cauchyfolge in L^2. Es gilt also $S_n \to S_\infty$ in L^2 und damit auch stochastisch für ein $S_\infty \in L^2$.

4.14 Lemma (Ottaviani-Ungleichung)**.** *Unter den Voraussetzungen des Satzes gilt für $\alpha > 0$*

$$P\Big(\max_{j=1,\dots,n} |S_j| \geqslant 2\alpha\Big) \leqslant \frac{P(|S_n| \geqslant \alpha)}{1 - \max_{j=1,\dots,n} P(|S_n - S_j| \geqslant \alpha)}.$$

4.15 Satz (Lévy's Äquivalenzsatz)**.** *Es seien $(X_i)_{i\geqslant 1}$ eine Folge unabhängiger Zufallsvariablen und $S_n := \sum_{i=1}^n X_i$, $n \geqslant 1$. Dann sind für $n \to \infty$ äquivalent:*

(a) $(S_n)_{n\geqslant 1}$ konvergiert fast sicher.

(b) $(S_n)_{n\geqslant 1}$ konvergiert stochastisch.

Zusatz: andernfalls divergiert $(S_n)_{n\geqslant 1}$ mit Wahrscheinlichkeit Eins.

4.16 Beispiel (Harmonische Reihe mit zufälligen Vorzeichen)**.** $S_n = \sum_{k=1}^n X_k \frac{1}{k}$ mit (X_k) unabhängig und $P(X_k = 1) = P(X_k = -1) = 1/2$ konvergiert sogar P-f.s.

4.2 Konvergenz in Verteilung

4.17 Definition. Die $\mathbb{R}^d$-wertigen Zufallsvektoren $(X_n)_{n\geqslant 1}$ <u>konvergieren in Verteilung</u> gegen den $\mathbb{R}^d$-wertigen Zufallsvektor X, Notation $X_n \xrightarrow{d} X$, falls für jede stetige beschränkte Funktion $\varphi : \mathbb{R}^d \to \mathbb{R}$ gilt

$$\lim_{n\to\infty} \mathbb{E}[\varphi(X_n)] = \mathbb{E}[\varphi(X)].$$

Wahrscheinlichkeitsmaße $(P_n)_{n\geqslant 1}$ auf $(\mathbb{R}^d, \mathfrak{B}_{\mathbb{R}^d})$ <u>konvergieren schwach</u> gegen ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}^d, \mathfrak{B}_{\mathbb{R}^d})$, Notation $P_n \xrightarrow{w} P$, falls für jede stetige beschränkte Funktion $\varphi : \mathbb{R}^d \to \mathbb{R}$ gilt

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} \varphi(x)\, P_n(dx) = \int_{\mathbb{R}^d} \varphi(x)\, P(dx).$$

Man definiert <u>Konvergenz in Verteilung</u> mit einem Wahrscheinlichkeitsmaß P als Grenzwert allgemein durch $X_n \xrightarrow{d} P :\iff P^{X_n} \xrightarrow{w} P$.

4.18 Beispiel. Ist X ein Zufallsvektor und $a_n \in \mathbb{R}$, $b_n \in \mathbb{R}^d$ mit $a_n \to a$, $b_n \to b$, so gilt $a_n X + b_n \xrightarrow{d} aX + b$: Stetigkeit von φ impliziert $\varphi(a_n X(\omega) + b_n) \to \varphi(aX(\omega) + b)$ für alle $\omega \in \Omega$. Nun ist $\|\varphi\|_\infty$ eine integrierbare Majorante und majorisierte (dominierte) Konvergenz zeigt $\mathbb{E}[\varphi(a_n X + b_n)] \to \mathbb{E}[\varphi(aX + b)]$. Beachte, dass $P_n \xrightarrow{w} P$ <u>nicht</u> impliziert $P_n(B) \to P(B)$ für alle $B \in \mathfrak{B}_{\mathbb{R}^d}$: setze $a_n = a = 0$ oben und schließe $\delta_{b_n} \xrightarrow{d} \delta_b$, während für $b_n \neq b$ gilt $\delta_{b_n}(\{b\}) = 0 \neq 1 = \delta_b(\{b\})$.

4.19 Satz. *Konvergiert X_n gegen X stochastisch, so auch in Verteilung: $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$.*

4.20 Beispiel. Die Umkehrung gilt nicht: für X $N(0,1)$-verteilt ist auch $Y = -X$ $N(0,1)$-verteilt, so dass $X_n \xrightarrow{d} Y$ für $X_n := X$ gilt, während $P(|X_n - Y| > \varepsilon) = P(2|X| > \varepsilon)$ für $\varepsilon > 0$ nicht gegen Null konvergiert.

4.21 Satz. *Für reellwertige Zufallsvariablen sind äquivalent:*

(a) $X_n \xrightarrow{d} X$

(b) Die Verteilungsfunktionen erfüllen $F^{X_n}(x) \to F^X(x)$ für alle $x \in \mathbb{R}$, an denen F^X stetig ist (<u>Stetigkeitspunkte</u> von F^X).

4.22 Beispiele.

(a) Sind X_n $N(0, \sigma_n^2)$-verteilt mit $\sigma_n \to 0$, so gilt $X_n \xrightarrow{d} \delta_0 = N(0,0)$.

(b) Sind X_n, X diskrete Zufallsvariablen mit Werten in $\mathbb{Z}$, so gilt $X_n \xrightarrow{d} X$ genau dann, wenn für die Zähldichten $p^{X_n}(k) \to p^X(k)$ für alle $k \in \mathbb{Z}$ gilt. Der Poissonsche Grenzwertsatz besagt insbesondere $\mathrm{Bin}(n, p_n) \xrightarrow{w} \mathrm{Poiss}(\lambda)$ für $np_n \to \lambda > 0$.

(c) Sind $U_1, \ldots, U_n$ unabhängige $U([0,1])$-verteilte Zufallsvariablen, so besitzt $X_n = n \min(U_1, \ldots, U_n)$ die Verteilungsfunktion $F^{X_n}(x) = 1 - (1 - x/n)_+^n$ für $x \geqslant 0$. Es folgt $n \min(U_1, \ldots, U_n) \xrightarrow{d} \mathrm{Exp}(1)$.

4.23 Satz. *(Auswahlsatz von Helly) Ist (P_n) eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$ mit Verteilungsfunktionen (F_n), so existiert eine Teilfolge (n_k) und eine monoton wachsende rechtsstetige Funktion $F : \mathbb{R} \to [0,1]$ mit $\lim_{k \to \infty} F_{n_k}(x) = F(x)$ für alle Stetigkeitspunkte von F.*

4.24 Beispiele.

(a) Sind P_n Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$ mit Verteilungsfunktionen F_n, so folgt aus $P_n \xrightarrow{w} P$, dass $F_n(x) \to F(x)$ an den Stetigkeitspunkten x von F gerade für die Verteilungsfunktion F von P gilt.

(b) Ist $P_n = U([n, n+1])$ mit Verteilungsfunktionen F_n, so gilt $\lim_{n \to \infty} F_n(x) = 0$ sowie $\lim_{n \to -\infty} F_n(x) = 1$ für alle $x \in \mathbb{R}$. Für $P_n = N(0, n)$ gilt $\lim_{n \to \infty} F_n(x) = 1/2$ für alle $x \in \mathbb{R}$. Die Funktion F im Satz von Helly ist hier jeweils keine Verteilungsfunktion.

4.25 Definition. Eine Folge von Wahrscheinlichkeitsmaßen (P_n) auf $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$ heißt (gleichgradig) straff, falls für jedes $\varepsilon > 0$ ein $K_\varepsilon > 0$ existiert mit $\sup_{n\geqslant 1} P_n([-K_\varepsilon, K_\varepsilon]^\complement) \leqslant \varepsilon$.

4.26 Beispiel. Es gelte $P_n \xrightarrow{w} P$ und F bezeichne die Verteilungsfunktion von P. Dann gibt es für $\varepsilon > 0$ ein $x > 0$ mit $F(-x) \leqslant -\varepsilon/4$, $F(x) \geqslant 1 - \varepsilon/4$ und F ist stetig bei x und $-x$. Dann folgt $P_n((-x, x]) \to F(x) - F(-x) \geqslant 1 - \varepsilon/2$. Wähle nun $N \in \mathbb{N}$, so dass $P_n([-x, x]) \geqslant 1 - \varepsilon$ für alle $n \geqslant N$ sowie $y \geqslant x$ mit $P_n([-y, y]) \geqslant 1 - \varepsilon$ für $n = 1, \ldots, N - 1$ (möglich wegen σ-Stetigkeit). Dann gilt $P_n([-y, y]^\complement) \leqslant \varepsilon$ für alle n. (P_n) ist also straff.

4.27 Korollar. *Ist (P_n) eine straffe Folge von Wahrscheinlichkeitsmaßen, so gibt es eine Teilfolge (n_k) und ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$, so dass $P_{n_k} \xrightarrow{w} P$ gilt.*

4.3 Charakteristische Funktionen und Zentrale Grenzwertsätze

4.28 Definition. Für eine reellwertige Zufallsvariable X bezeichnet

$$\varphi^X(u) := \mathbb{E}[e^{iuX}] = \mathbb{E}[\cos(uX)] + i\,\mathbb{E}[\sin(uX)], \quad u \in \mathbb{R},$$

die charakteristische Funktion von X. Entsprechend ist für ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$

$$\varphi^P(u) := \int_{\mathbb{R}} e^{iux} P(dx) = \int_{\mathbb{R}} \cos(ux)\, P(dx) + i \int_{\mathbb{R}} \sin(ux)\, P(dx), \quad u \in \mathbb{R},$$

die charakteristische Funktion von P.

4.29 Beispiele.

(a) $\varphi^{\mathrm{Bin}(n,p)}(u) = (pe^{iu} + 1 - p)^n$;

(b) $\varphi^{\mathrm{Poiss}(\lambda)}(u) = \exp(\lambda(e^{iu} - 1))$;

(c) $\varphi^{N(0,1)}(u) = e^{-u^2/2}$, $\varphi^{N(\mu,\sigma^2)}(u) = e^{iu\mu - \sigma^2 u^2/2}$.

4.30 Lemma. *Eine charakteristische Funktion φ erfüllt $\varphi(0) = 1$, $\sup_u |\varphi(u)| \leqslant 1$ und ist gleichmäßig stetig auf $\mathbb{R}$.*

4.31 Lemma. *Es gilt $\varphi^{P_1 * P_2}(u) = \varphi^{P_1}(u)\varphi^{P_2}(u)$ bzw. $\varphi^{X_1+X_2}(u) = \varphi^{X_1}(u)\varphi^{X_2}(u)$ für unabhängige Zufallsvariablen X_1, X_2.*

4.32 Lemma. *Für eine Zufallsvariable $X \in \mathscr{L}^m(\Omega, \mathcal{F}, P)$ mit $m \in \mathbb{N}$ gilt $\varphi^X \in C^m(\mathbb{R})$ mit Ableitungen $(\varphi^X)^{(k)}(0) = i^k\,\mathbb{E}[X^k]$, $k = 0, \ldots, m$. Weiterhin existiert eine Funktion $r_m : \mathbb{R} \to \mathbb{C}$ mit $\sup_{u\in\mathbb{R}} |r_m(u)| \leqslant 3\,\mathbb{E}[|X|^m]$ und $r_m(u) \to 0$ für $u \to 0$, so dass gilt*

$$\varphi^X(u) = \sum_{k=0}^{m} \frac{(iu)^k}{k!}\,\mathbb{E}[X^k] + \frac{(iu)^m}{m!} r_m(u).$$

4.33 Satz. *(Eindeutigkeitssatz) Für $a < b$ gilt die Inversionsformel*

$$P((a,b)) + \tfrac{1}{2}P(\{a,b\}) = \frac{1}{2\pi}\lim_{U\to\infty}\int_{-U}^{U}\frac{e^{-iua}-e^{-iub}}{iu}\varphi^P(u)\,du.$$

Insbesondere sind zwei Wahrscheinlichkeitsmaße mit derselben charakteristischen Funktion identisch.

4.34 Satz. *(Stetigkeitssatz von Lévy) Sind (P_n) Wahrscheinlichkeitsmaße auf $(\mathbb{R},\mathfrak{B}_{\mathbb{R}})$ mit charakteristischen Funktionen (φ_n) und gilt $\lim_{n\to\infty}\varphi_n(u) = \psi(u)$ für alle $u \in \mathbb{R}$ und eine bei $u = 0$ stetige Funktion ψ, so ist $\psi = \varphi^P$, die charakteristische Funktion eines Wahrscheinlichkeitsmaßes P auf $(\mathbb{R},\mathfrak{B}_{\mathbb{R}})$, und es gilt $P_n \xrightarrow{w} P$.*

4.35 Beispiele.

(a) Für $p_n \in [0,1]$ mit $\lim_{n\to\infty} np_n = \lambda > 0$ folgt $\varphi^{\mathrm{Bin}(n,p_n)} \to \varphi^{\mathrm{Poiss}(\lambda)}$ punktweise. Der Stetigkeitssatz von Lévy impliziert daher den Poissonschen Grenzwertsatz $\mathrm{Bin}(n,p_n) \xrightarrow{w} \mathrm{Poiss}(\lambda)$.

(b) Ist S_n $\mathrm{Bin}(n,p)$-verteilt mit $p \in (0,1)$, so ist $S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}}$ zentriert mit Varianz 1 (*standardisiert*). Es gilt mit $q = 1-p$ für $n \to \infty$

$$\varphi^{S_n^*}(u) = \big(pe^{iuq/\sqrt{npq}} + qe^{-iup/\sqrt{npq}}\big)^n = \big(1 - u^2/(2n) + o(1/n)\big)^n,$$

so dass $\varphi^{S_n^*}$ punktweise gegen $\varphi^{N(0,1)}$ konvergiert. Es folgt $S_n^* \xrightarrow{d} N(0,1)$.

(c) Für unabhängige $U([-1,1])$-verteilte Zufallsvariablen X_j, $j \in \mathbb{N}$, gilt $\mathbb{E}[X_j] = 0$, $\mathrm{Var}(X_j) = 1/3$. Die *standardisierte Summe* $S_n^* = \sqrt{\frac{3}{n}}\sum_{j=1}^n X_j$ hat charakteristische Funktion

$$\varphi^{S_n^*}(u) = \Big(\frac{\sin(\sqrt{3}u/\sqrt{n})}{\sqrt{3}u/\sqrt{n}}\Big)^n = \big(1 - u^2/(2n) + o(1/n)\big)^n.$$

Es folgt wiederum $S_n^* \xrightarrow{d} N(0,1)$ für $n \to \infty$.

4.36 Satz. *(Zentraler Grenzwertsatz) Ist $(X_i)_{i\geqslant 1}$ eine Folge unabhängiger und identisch verteilter Zufallsvariablen (i.i.d.=independent and identically distributed) in $\mathscr{L}^2$ mit $\mu = \mathbb{E}[X_i]$, $\sigma^2 = \mathrm{Var}(X_i) > 0$, so erfüllt ihre standardisierte Summe*

$$S_n^* := \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i - \mu}{\sigma} \xrightarrow{d} N(0,1).$$

Insbesondere gilt für $a < b$ also $\mathbb{P}(a < S_n^* \leqslant b) \to \Phi(b) - \Phi(a)$ mit der Verteilungsfunktion Φ der Standardnormalverteilung $N(0,1)$.

4.37 Lemma (*Continuous mapping theorem*)**.** *Konvergiert X_n gegen X fast sicher (bzw. stochastisch bzw. in Verteilung) und ist $g : \mathbb{R} \to \mathbb{R}$ stetig, so konvergiert auch $g(X_n)$ gegen $g(X)$ fast sicher (bzw. stochastisch bzw. in Verteilung).*

4.38 Beispiel. Im zentralen Grenzwertsatz folgt aus $S_n^* \xrightarrow{d} N(0,1)$ mit $g(x) = \sigma x$, dass $g(S_n^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \xrightarrow{d} N(0, \sigma^2)$ gilt.

4.39 Definition. Eine Familie $(X_j^{(n)})_{1 \leqslant j \leqslant n, n \geqslant 1}$ von Zufallsvariablen in $\mathscr{L}^2$ bildet ein <u>standardisiertes Dreiecksschema</u>, falls für jedes $n \in \mathbb{N}$ die Zufallsvariablen $X_1^{(n)}, \ldots, X_n^{(n)}$ („in einer Zeile") unabhängig sind sowie $\mathbb{E}[X_j^{(n)}] = 0$, $j = 1, \ldots, n$, und $\sum_{j=1}^n \operatorname{Var}(X_j^{(n)}) = 1$ gilt. $(X_j^{(n)})_{1 \leqslant j \leqslant n, n \geqslant 1}$ genügt der <u>Lindeberg-Bedingung</u>, falls gilt

$$\forall \delta > 0: \lim_{n \to \infty} \mathbb{E}\left[(X_j^{(n)})^2 \mathbf{1}(|X_j^{(n)}| \geqslant \delta)\right] = 0.$$

4.40 Beispiele.

(a) Ist $(Y_j)_{j \geqslant 1}$ eine Folge unabhängiger, identisch verteilter (i.i.d.) Zufallsvariablen in $\mathscr{L}^2$ mit $\mu = \mathbb{E}[Y_j]$, $\sigma^2 = \operatorname{Var}(Y_j) > 0$, so bildet $X_j^{(n)} = (Y_j - \mu)/(\sigma\sqrt{n})$ ein standardisiertes Dreiecksschema, das der Lindeberg-Bedingung genügt.

(b) Ist $(X_j^{(n)})_{1 \leqslant j \leqslant n, n \geqslant 1}$ ein standardisiertes Dreiecksschema und gilt für ein $p > 2$ die <u>Lyapunov-Bedingung</u>

$$\lim_{n \to \infty} \sum_{j=1}^n \mathbb{E}\left[|X_j^{(n)}|^p\right] = 0,$$

so genügt $(X_j^{(n)})_{1 \leqslant j \leqslant n, n \geqslant 1}$ auch der Lindeberg-Bedingung.

Ist insbesondere $(Y_j)_{j \geqslant 1}$ eine Folge unabhängiger und standardisierter (d.h. $\mathbb{E}[Y_j] = 0$, $\operatorname{Var}(Y_j) = 1$) Zufallsvariablen, die in L^p beschränkt ist (d.h. $\sup_{j \geqslant 1} \mathbb{E}[|Y_j|^p] < \infty$) für ein $p > 2$, so ist $X_j^{(n)} = Y_j/\sqrt{n}$ ein standardisiertes Dreiecksschema, das der Lyapunov- und somit auch der Lindeberg-Bedingung genügt:

$$\sum_{j=1}^n \mathbb{E}\left[|X_j^{(n)}|^p\right] = n^{-p/2} \sum_{j=1}^n \mathbb{E}[|Y_j|^p] \leqslant n^{1-p/2} \sup_{j \geqslant 1} \mathbb{E}[|Y_j|^p] \xrightarrow{n \to \infty} 0.$$

(c) In der harmonischen Reihe mit zufälligen Vorzeichen $S_n = \sum_{j=1} \varepsilon_j / j$ mit $(\varepsilon_j)_{j \geqslant 1}$ unabhängig, $P(\varepsilon_j = 1) = P(\varepsilon_j = -1) = 1/2$, bilden $X_j^{(n)} = \varepsilon_j/(j\sigma_n)$ mit $\sigma_n^2 = \sum_{j=1}^n j^{-2} \to \sigma_\infty^2 = \pi^2/6$ ein standardisiertes Dreiecksschema, das der Lypunov-Bedingung *nicht* genügt; denn für $\delta \in (0, \sigma_\infty^{-1})$ erhält man

$$\sum_{j=1}^n \mathbb{E}[(X_j^{(n)})^2 \mathbf{1}(|X_j^{(n)}| \geqslant \delta)] = \sum_{j=1}^n j^{-2} \sigma_n^{-2} \mathbf{1}(j\sigma_n\delta \leqslant 1) \geqslant \sigma_\infty^{-2} > 0.$$

4.41 Satz (Zentraler Grenzwertsatz nach Lindeberg, 1922). *Ist $(X_j^{(n)})_{1 \leqslant j \leqslant n, n \geqslant 1}$ ein standardisiertes Dreiecksschema, das der Lindeberg-Bedingung genügt, so gilt für (die Zeilensummen) $S_n^* = \sum_{j=1}^n X_j^{(n)}$ und $n \to \infty$*

$$S_n^* \xrightarrow{d} N(0,1).$$

4.42 Definition. Es seien $X_1, \dots, X_n$ unabhängige, identisch verteilte Zufallsvariablen (*Beobachtungen*) mit Werten in $\mathbb{R}$. Dann heißt das Wahrscheinlichkeitsmaß $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ empirische Verteilung oder empirisches Maß sowie seine Verteilungsfunktion $F_n(x) := \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{(-\infty,x]}(X_i)$, $x \in \mathbb{R}$, empirische Verteilungsfunktion.

4.43 Satz. *Für alle $x \in \mathbb{R}$ gilt $\lim_{n\to\infty} F_n(x) = F^X(x)$ P-fast sicher mit $F^X(x) = P(X_i \leqslant x)$. Für alle $x \in \mathbb{R}$ mit $F^X(x) \in (0,1)$ gilt*

$$\sqrt{n}\big(F_n(x) - F^X(x)\big) \xrightarrow{d} N\big(0, F^X(x)(1 - F^X(x))\big).$$

4.44 Satz (Glivenko-Cantelli)**.** *Die empirische Verteilungsfunktion konvergiert gleichmäßig gegen die wahre Verteilungsfunktion:*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F^X(x)| = 0 \quad P\text{-f.s.}$$

5 Einführung in die Schätztheorie

5.1 Grundlagen

5.1 Definition. Ein statistisches Modell ist ein Tripel $(\Omega, \mathscr{F}, (P_\vartheta)_{\vartheta\in\Theta})$ bestehend aus einem Messraum $(\Omega, \mathscr{F})$ (dem Stichprobenraum) und einer Familie $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ von Wahrscheinlichkeitsmaßen auf $\mathscr{F}$. Die nichtleere Menge Θ heißt Parametermenge und jedes $\vartheta \in \Theta$ Parameter.

5.2 Beispiel (Normalverteilungsmodell)**.** Für ein n-fach wiederholtes physikalisches Experiment, wo die Messfehler als $N(0,\sigma^2)$-verteilt für ein $\sigma > 0$ angenommen werden können, ist $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n}, (P_\vartheta)_{\vartheta\in\Theta})$ mit $\Theta = \{(\mu,\sigma^2) \,|\, \mu \in \mathbb{R}, \sigma > 0\}$ und $P_\vartheta = P_{(\mu,\sigma^2)} = N(\mu,\sigma^2)^{\otimes n} = N(\mu\mathbf{1}, \sigma^2 E_n)$ ein adäquates statistisches Modell für die Messergebnisse.

5.3 Definition. Es sei $(\Omega, \mathscr{F}, (P_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell sowie $g : \Theta \to \mathbb{R}$. Für jedes $\vartheta \in \Theta$ wird $g(\vartheta)$ abgeleiteter Parameter genannt. Jede messbare Funktion $\hat{g} : \Omega \to \mathbb{R}$ heißt Schätzer von $g(\vartheta)$. Für eine Realisierung (konkrete Beobachtung, Stichprobe) $\omega \in \Omega$ ist $\hat{g}(\omega)$ der zugehörige Schätzwert.

5.4 Beispiel (Normalverteilungsmodell)**.** Im Normalverteilungsmodell ist häufig $g(\vartheta) = g(\mu,\sigma^2) = \mu$ der einzig interessierende Parameter. Schätzer von μ sind $\hat{\mu}_1(\omega) = 15$, $\hat{\mu}_2(\omega) = \omega_1$, $\hat{\mu}_3(\omega) = \frac{1}{n}\sum_{i=1}^n \omega_i$, $\hat{\mu}_4(\omega) = $Median$(\omega_1, \dots, \omega_n)$ für $\omega \in \Omega = \mathbb{R}^n$.

5.5 Definition. Der mittlere quadratische Fehler MSE (*mean squared error*) eines Schätzers $\hat{g}$ von $g(\vartheta)$ ist gegeben durch

$$R(\hat{g}, \vartheta) := \mathbb{E}_\vartheta[(\hat{g} - g(\vartheta))^2] \in [0,\infty], \quad \vartheta \in \Theta.$$

Liegt $\hat{g}$ in $\mathscr{L}^1(P_\vartheta)$, so heißt

$$B(\hat{g}, \vartheta) := \mathbb{E}_\vartheta[\hat{g} - g(\vartheta)], \quad \vartheta \in \Theta$$

Verzerrung oder Bias von $\hat{g}$. Gilt $B(\hat{g}, \vartheta) = 0$ für alle $\vartheta \in \Theta$, so ist $\hat{g}$ ein erwartungstreuer Schätzer von $g(\vartheta)$.

5.6 Beispiele.

(a) Im Normalverteilungsmodell mit $g(\vartheta) = \mu$ und $\vartheta = (\mu, \sigma^2)$ gilt $R(\hat{\mu}_1, \vartheta) = (15 - \mu)^2$, $R(\hat{\mu}_2, \vartheta) = \sigma^2$, $R(\hat{\mu}_3, \vartheta) = \sigma^2 n^{-1}$ und $R(\hat{\mu}_4, \vartheta)$ ist nicht explizit berechenbar. Die Schätzer $\hat{\mu}_2$, $\hat{\mu}_3$ und $\hat{\mu}_4$ (Letzterer aus Symmetriegründen) sind erwartungstreu. Der MSE von $\hat{\mu}_2$ ist stets größer als der MSE von $\hat{\mu}_3$, aber je nach Parameterwert ϑ kann der MSE von $\hat{\mu}_1$ kleiner oder auch viel größer als der MSE von $\hat{\mu}_3$ sein.

(b) Im Bernoulli-Schema mit $\Omega = \{0,1\}^n$, $\mathscr{F} = \mathscr{P}(\Omega)$ und der Verteilung P_p des Bernoullischemas mit Erfolgswahrscheinlichkeit $p \in \Theta = (0,1)$ ist die relative Häufigkeit $\hat{p}(\omega) = \frac{1}{n}\sum_{i=1}^n \omega_i$ ein typischer Schätzer von $g(p) = p$. $\hat{p}$ ist erwartungstreu mit MSE $R(\hat{p}, p) = p(1-p)n^{-1}$. Der MSE ist für $p = 1/2$ mit $(4n)^{-1}$ maximal und fällt für $p \to 0$, $p \to 1$ gegen Null.

Dies macht sich der Schätzer $\hat{p}_\alpha = (1-\alpha)\hat{p} + \alpha\frac{1}{2}$ mit $\alpha \in [0,1]$ zunutze, indem er $\hat{p}$ in Richtung $\frac{1}{2}$ verzerrt. $\hat{p}_\alpha$ besitzt den Bias $B(\hat{p}_\alpha, p) = \alpha(\frac{1}{2} - p)$ und die Varianz $\mathrm{Var}_p(\hat{p}_\alpha) = (1-\alpha)^2 p(1-p)n^{-1}$. Der MSE $R(\hat{p}_\alpha, p) = \frac{(1-\alpha)^2}{4n} + (\alpha^2 - (1-\alpha)^2 n^{-1})(p - \frac{1}{2})^2$ hängt für $\alpha_0 = (\sqrt{n}+1)^{-1}$ nicht von p ab. Es gilt $R(\hat{p}_{\alpha_0}, p) = \frac{1}{4(n+2\sqrt{n}+1)}$, und $\hat{p}_{\alpha_0}$ hat kleineren maximalen MSE (über $p \in (0,1)$) als $\hat{p}$, ist aber nicht erwartungstreu.

5.2 Cramér-Rao-Ungleichung und Fisher-Information

5.7 Definition. Ein statistisches Modell $(\Omega, \mathscr{F}, (P_\vartheta)_{\vartheta\in\Theta})$ heißt dominiert von einem Maß μ auf $\mathscr{F}$, falls jedes P_ϑ eine μ-Dichte $p_\vartheta : \Omega \to [0,\infty)$ besitzt. Fasst man p_ϑ als Zufallsvariable auf, so heißt die (zufällige) Funktion $L(\vartheta) = L(\vartheta, \omega) = p_\vartheta(\omega)$ auf Θ Likelihood-Funktion. Mit $\ell(\vartheta) = \log(L(\vartheta))$ wird die Loglikelihood-Funktion bezeichnet.

5.8 Beispiele.

(a) Besitzen im Fall $(\Omega, \mathscr{F}) = (\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n})$ die P_ϑ Lebesguedichten f_ϑ, so gilt $L(\vartheta, x) = f_\vartheta(x)$, $\vartheta \in \Theta$, $x \in \mathbb{R}^n$. Bei Produktdichten $f_\vartheta(x) = \prod_{i=1}^n f_{\vartheta,i}(x_i)$ ist die Loglikelihood-Funktion additiv: $\ell(\vartheta, x) = \sum_{i=1}^n \log(f_{\vartheta,i}(x_i))$. Im Normalverteilungsmodell $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n}, (N(\mu,\sigma^2)^{\otimes n})_{\mu\in\mathbb{R},\sigma>0}$ ergibt sich $\ell(\mu, \sigma) = -\frac{n}{2}\log(2\pi\sigma) - \sum_{i=1}^n (x_i - \mu)^2/(2\sigma^2)$, wobei $x = (x_1, \ldots, x_n) \in \Omega = \mathbb{R}^n$ den Versuchsausgang angibt.

(b) Im diskreten Fall mit Ω abzählbar und $\mathscr{F} = \mathscr{P}(\Omega)$ ist die Likelihood-Funktion bezüglich dem Zählmaß $L(\vartheta)$ gerade die Zähldichte p_ϑ von P_ϑ. Im Binomialmodell $(\mathrm{Bin}(n,p))_{p\in(0,1)}$ ergibt sich $L(p) = \binom{n}{x} p^X (1-p)^{n-X}$, $\ell(p) = n\log(1-p) + \log\binom{n}{x} + x\log(p/(1-p))$, wobei $x \in \Omega = \{0, \ldots, n\}$ den Versuchsausgang angibt.

5.9 Lemma (Chapman-Robbins-Ungleichung)**.** *Es sei $(\Omega, \mathscr{F}, (P_\vartheta)_{\vartheta\in\Theta})$ ein dominiertes statistisches Modell mit Likelihood-Funktion $L(\vartheta)$. Ist $\hat{g}$ ein erwartungstreuer Schätzer von $g(\vartheta)$ für $g : \Theta \to \mathbb{R}$, so gilt für alle $\vartheta, \vartheta_0 \in \Theta$ mit $P_\vartheta \neq P_{\vartheta_0}$ und $P_\vartheta \ll P_{\vartheta_0}$*

$$R(\hat{g}, \vartheta_0) = \mathrm{Var}_{\vartheta_0}(\hat{g}) \geqslant \frac{(g(\vartheta) - g(\vartheta_0))^2}{\mathrm{Var}_{\vartheta_0}(L(\vartheta)/L(\vartheta_0))}.$$

Im Fall $L(\vartheta)/L(\vartheta_0) \notin \mathscr{L}^2(P_{\vartheta_0})$ sei die rechte Seite als Null definiert.

5.10 Beispiel. Im vom Lebesguemaß dominierten statistischen Modell $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}}, (\mathrm{Exp}(\vartheta))_{\vartheta>0})$ gilt $L(\vartheta, x)/L(\vartheta_0, x) = \frac{\vartheta}{\vartheta_0} e^{-(\vartheta-\vartheta_0)x}$ für $\vartheta, \vartheta_0 > 0$ und P_{ϑ_0}-fast alle $x \in \mathbb{R}$ (z.B. $x \geqslant 0$). Für $g(\vartheta) = \vartheta$ erhalten wir für erwartungstreue Schätzer $\hat{\vartheta}$ von ϑ:

$$R(\hat{\vartheta}, \vartheta_0) \geqslant \sup_{\vartheta>\vartheta_0/2} \frac{(\vartheta-\vartheta_0)^2}{\int_0^\infty (\vartheta\vartheta_0^{-1} e^{-(\vartheta-\vartheta_0)x} - 1)^2 \vartheta_0 e^{-\vartheta_0 x} dx} = \sup_{\vartheta>\vartheta_0/2} \vartheta_0(2\vartheta - \vartheta_0) = \infty.$$

Es gibt also keinen erwartungstreuen Schätzer von ϑ mit endlicher Varianz!

Für $g(\vartheta) = \vartheta^{-1}$ und erwartungstreue Schätzer $\hat{g}$ hingegen folgt

$$R(\vartheta_0, \hat{\vartheta}) \geqslant \sup_{\vartheta>\vartheta_0/2} \frac{2\vartheta - \vartheta_0}{\vartheta^2 \vartheta_0} = \vartheta_0^{-2}.$$

Die Identität $\hat{g} = X$ als Schätzer erreicht genau diese Schranke, ist daher erwartungstreuer Schätzer von ϑ^{-1} mit minimalem MSE (also minimaler Varianz). Analog gilt im Produktmodell von unabhängigen Beobachtungen $X_1, \dots, X_n \sim \mathrm{Exp}(\vartheta)$, dass $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ erwartungstreuer Schätzer von ϑ^{-1} mit minimaler Varianz $n^{-1}\vartheta^{-2}$ ist. Das Supremum in der Chapman-Robbins-Ungleichung wird in beiden Fällen für $\vartheta \to \vartheta_0$ erreicht, was auf eine differentielle Ungleichung, die Cramér-Rao-Ungleichung, führt.

5.11 Definition. Ein dominiertes statistisches Modell $(\Omega, \mathscr{F}, (P_\vartheta)_{\vartheta\in\Theta})$ mit $\Theta \subseteq \mathbb{R}$ offen und Loglikelihood-Funktion ℓ heißt Cramér-Rao-regulär bei $\vartheta_0 \in \Theta$, falls

(a) $P_\vartheta \ll P_{\vartheta_0}$ für alle ϑ in einer Umgebung von ϑ_0 gilt;

(b) $\vartheta \mapsto \ell(\vartheta, \omega)$ für P_{ϑ_0}-fast alle $\omega \in \Omega$ bei ϑ_0 differenzierbar ist mit Ableitung $\dot{\ell}(\vartheta_0) \in \mathscr{L}^2(P_{\vartheta_0})$;

(c) $\lim_{\vartheta\to\vartheta_0} \mathbb{E}_{\vartheta_0}\Big[\Big(\frac{\exp(\ell(\vartheta)-\ell(\vartheta_0))-1}{\vartheta-\vartheta_0} - \dot{\ell}(\vartheta_0)\Big)^2\Big] = 0$ gilt.

Die zufällige Funktion $\dot{\ell}$ heißt Score-Funktion und $I(\vartheta_0) := \mathbb{E}_{\vartheta_0}[\dot{\ell}(\vartheta_0)^2]$ Fisher-Information des Modells bei ϑ_0.

5.12 Satz (Cramér-Rao-Ungleichung)**.** *Ist das statistische Modell $(\Omega, \mathscr{F}, (P_\vartheta)_{\vartheta\in\Theta})$ mit $\Theta \subseteq \mathbb{R}$ offen bei $\vartheta_0 \in \Theta$ Cramér-Rao-regulär mit Fisher-Information $I(\vartheta_0) > 0$ und ist $g : \Theta \to \mathbb{R}$ bei ϑ_0 differenzierbar, so gilt für jeden erwartungstreuen Schätzer $\hat{g}$ von $g(\vartheta)$*

$$R(\hat{g}, \vartheta_0) = \mathrm{Var}_{\vartheta_0}(\hat{g}) \geqslant \frac{g'(\vartheta_0)^2}{I(\vartheta_0)}.$$

5.13 Beispiel. Im Normalverteilungsmodell $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n}, (N(\mu, \sigma^2)^{\otimes n})_{\mu\in\mathbb{R}}$ mit $\sigma > 0$ bekannt ist für jeden Parameter $\mu \in \mathbb{R}$ das Modell Cramér-Rao-regulär mit Score-Funktion $\dot{\ell}(\mu) = -\sum_{i=1}^n (X_i - \mu)/\sigma^2$ und Fisher-Information $I(\mu) = n/\sigma^2$. Für erwartungstreue Schätzer $\hat{\mu}$ von $g(\mu) = \mu$ gilt daher

$\mathrm{Var}_\mu(\hat{\mu}) \geqslant \sigma^2/n$. Das Stichprobenmittel $\bar{X}$ ist also erwartungstreuer Schätzer mit minimaler Varianz.

Auch im Binomialmodell lässt sich nachrechnen, dass das Stichprobenmittel (die relative Häufigkeit) $\hat{p} = X/n$ erwartungstreuer Schätzer von minimaler Varianz für p ist.