Stochastik I
Gliederung zur Vorlesung
im Sommersemester 2012

Markus Reiß
Humboldt-Universität zu Berlin

Vorläufige Version vom 11. Juli 2012

Inhaltsverzeichnis

Markus Reiß

Vorlesung
Stochastik I
Sommersemester 2009

Ein paar Literaturempfehlungen

- Hans-Otto Georgii, *Stochastik*, de Gruyter: exzellentes Lehrbuch inkl. Maßtheorie, verfügbar als E-Book: `http://www.reference-global.com/isbn/978-3-11-019349-7`

- 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 char. Funktionen und Konvergenzresultate

- 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 Definition. Mit Ω werde die nichtleere Menge der möglichen Versuchsausgänge oder Ergebnismenge 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 $P : \mathscr{F} \to [0, 1]$ mit

(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 $\mathscr{F}$.

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

1.3 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.*

1.4 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 $\mathscr{S} = \mathfrak{B}_{\mathbb{R}^d}$ 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 meinen wir stets die zu P^X gehörige Größe.

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.

1.2 Diskrete Verteilungen

1.5 Definition. Ist Ω eine 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.

1.6 Lemma.

(a) *Ist $(\Omega, \mathscr{F}, P)$ ein diskreter Wahrscheinlichkeitsraum, so ist P eindeutig durch seine Zähldichte $p : \Omega \to [0, 1]$ 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.

1.7 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 $0 \leqslant n \leqslant N$, $0 \leqslant W \leqslant N$

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

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 $p \in [0,1]$

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

Multinomialverteilung: Länge $n \in \mathbb{N}$, Klassenzahl $r \in \mathbb{N}$, Erfolgswahrscheinlichkeiten $p_1, \dots, p_r \in [0,1]$ mit $\sum_i p_i = 1$

$$p_{Mult(n,r,p_1,\dots,p_r)}(k) = \frac{n!}{k_1! \cdots k_r!} p_1^{k_1} \cdots p_r^{k_r}, \quad k = (k_1, \dots, k_r) \in \{0, 1, \dots, n\}^r.$$

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

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

Poissonverteilung: Parameter $\lambda > 0$

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

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

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

1.9 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 $\mathbb{R}^d$

1.10 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.*

1.11 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 $\mathscr{F} = \sigma(\mathscr{E})$.

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

1.13 Satz.

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

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

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

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

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

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

1.15 Korollar.

(a) *Jede stetige Funktion $g : S \to T$ zwischen metrischen Räumen (S,d_S) und (T,d_T) ist Borel-messbar, 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.*

1.16 Definition. Es sei Ω eine nichtleere Menge. Dann heißt $\mathscr{A} \subseteq \mathscr{P}(\Omega)$ Algebra 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 Prämaß 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.

1.17 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 $A \in \mathscr{A}$.*

1.18 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.*

1.19 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.

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

1.21 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.

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

1.23 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.*

1.24 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 $\mathbb{R}^d$.

1.25 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 $\mathfrak{B}_{\mathbb{R}}$.

1.26 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.

1.27 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 $\mu \in \mathbb{R}$, $\sigma > 0$.

$\chi^2(1)$-Verteilung: $f_{\chi^2(1)}(x) = \frac{1}{\sqrt{2\pi x}} e^{-x/2} \mathbf{1}_{\mathbb{R}^+}(x)$

1.28 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 $P_f(B) = \int_B f(x)dx$.

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

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

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

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

1.30 Satz (Dichtetransformationssatz)**.** *Ist X eine reellwertige Zufallsvariable mit Dichte f^X sowie $Y = \varphi(X)$ für $\varphi : \mathbb{R} \to \mathbb{R}$ injektiv derart, dass die Inverse $\varphi^{-1} : \varphi(R) \to \mathbb{R}$ differenzierbar ist, so besitzt Y ebenfalls eine Dichte und zwar $f^Y(y) = f^X(\varphi^{-1}(y))|(\varphi^{-1})'(y)|\mathbf{1}(y \in \varphi(\mathbb{R}))$.*

1.31 Korollar. *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 $f^Y(y) = |a|^{-1} f^X(a^{-1}(y - b))$.*

1.32 Lemma. *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 $f^Y(y) = |\det(A)|^{-1} f^X(A^{-1}(y - b))$.*

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 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 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.

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

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

2.2 Unabhängige Ereignisse und Lemma von Borel-Cantelli

2.4 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

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

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

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

$$\liminf_{n \to \infty} A_n := \bigcup_{m \geqslant 1} \bigcap_{n \geqslant m} A_n = \{\omega \in \Omega \,|\, \omega \in A_n \text{ für alle, bis auf endlich viele } n\}.$$

2.6 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 $P(\limsup_{n \to \infty} A_n) = 1$.

2.7 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.

2.8 Lemma. *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.*

2.3 Unabhängige Zufallsvariablen

2.9 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 $\mathscr{F}^{X_i} = \{X_i^{-1}(A) \,|\, A \in \mathscr{S}_i\}$, $i \in I$, unabhängig sind.

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

2.11 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*

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

2.12 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*

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

2.13 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 ist auch die Familie $(g_i(X_i))_{i \in I}$ unabhängig.*

2.14 Definition. Es seien $(\Omega_k, \mathscr{F}_k, P_k)$, $1 \leqslant k \leqslant n$, Wahrscheinlichkeitsräume. Setze $\Omega := \Omega_1 \times \cdots \times \Omega_n$ und definiere über Ω die Produkt-σ-Algebra

$$\mathscr{F} := \mathscr{F}_1 \otimes \cdots \otimes \mathscr{F}_n := \sigma(\{A_1 \times \cdots \times A_n \,|\, A_1 \in \mathscr{F}_1, \dots, A_n \in \mathscr{F}_n\}).$$

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

$$\forall\, A_1 \in \mathscr{F}_1, \dots, A_n \in \mathscr{F}_n : P(A_1 \times \cdots \times A_n) = \prod_{k=1}^{n} P_k(A_k),$$

so heißt P Produktmaß, Schreibweise $P = P_1 \otimes \cdots \otimes P_n$.

2.15 Lemma. *Ist $(\Omega_1 \times \cdots \times \Omega_n, \mathscr{F}_1 \otimes \cdots \otimes \mathscr{F}_n, P_1 \otimes \cdots \otimes P_n)$ ein Produkt-Wahrscheinlichkeitsraum, so sind die Koordinatenabbildungen $\pi_k(\omega) = \pi_k(\omega_1, \dots, \omega_n) := \omega_k$, $k = 1, \dots, n$, unabhängige $(\Omega_k, \mathscr{F}_k)$-wertige Zufallsvariablen auf dem Produkt-Wahrscheinlichkeitsraum mit Verteilung $P^{\pi_k} = P_k$.*

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

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

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

$$\forall\, J \subseteq I \text{ endlich}, A_i \in \mathscr{F}_i : P\Big(\bigcap_{i\in J} \pi_i^{-1}(A_i)\Big) = \prod_{i\in J} P_i(A_i),$$

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

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

2.18 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.*

2.19 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

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

2.20 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 $P(A) = 1$.*

2.21 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.*

2.4 Faltung

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

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

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

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

2.25 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 $\mathbb{N}_0$: $(p*q)(k) := \sum_{m=0}^{k} p(k-m)q(m)$).*

2.26 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

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

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 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 $\mathbb{E}[\mathbf{1}_A] = P(A)$.

3.3 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 Erwartungswert

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

(man kann zeigen, dass dies 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

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

Man schreibt auch $\mathbb{E}[X] = \int X\,dP = \int_\Omega X(\omega)\,P(d\omega)$ sowie $\int_A X\,dP = \int_\Omega X(\omega)\mathbf{1}_A(\omega)\,P(d\omega)$ für $A \in \mathscr{F}$.

3.4 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 und monoton: ist Y eine weitere Zufallsvariable in $\mathscr{L}^1$ und sind $\alpha, \beta \in \mathbb{R}$, so gilt

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

aus $X \leqslant Y$ folgt $\mathbb{E}[X] \leqslant \mathbb{E}[Y]$.

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

3.5 Korollar.

(a) Ist X eine Zufallsvariable mit abzählbarem Wertebereich $X(\Omega) \subseteq \mathbb{R}$, so gilt $X \in \mathscr{L}^1$ genau dann, wenn $\sum_{x \in X(\Omega)} |x| P(X = x)$ endlich ist. In diesem Fall gilt für den Erwartungswert

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x\,P(X = x).$$

(b) Ist X eine Zufallsvariable mit Dichte $f^X : \mathbb{R} \to [0, \infty)$, so gilt $X \in \mathscr{L}^1$ genau dann, wenn $\int_\mathbb{R} |x| f^X(x)\,dx$ endlich ist. In diesem Fall gilt für den Erwartungswert

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

3.6 Satz. *Es seien X ein Zufallsvektor mit Dichte $f^X : \mathbb{R}^d \to [0, \infty)$ sowie $h : \mathbb{R}^d \to \mathbb{R}$ Borel-messbar. Dann gilt:*

$$h(X) \in \mathscr{L}^1 \iff \int_{\mathbb{R}^d} |h(x)| f^X(x)\,dx < \infty.$$

In diesem Fall erhalten wir

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

3.7 Definition. Wir sagen, dass eine Zufallsvariable X in $\mathscr{L}^p$ 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.

3.8 Lemma. *Für $0 < p \leqslant q$ gilt $\mathscr{L}^q \subseteq \mathscr{L}^p$.*

3.2 Varianz, Kovarianz und Korrelation

3.9 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.

3.10 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 $\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$.

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

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

die Menge der auf linearen Funktionen von X basierenden Zufallsvariablen. Dann nimmt die mittlere quadratische Abweichung

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

ihr 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* $\operatorname{Var}(X) > 0$ *gilt*

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

mit nachfolgend definierter Korrelation $\rho(X, Y)$.

3.12 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, so 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.

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

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

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

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

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

(e) falls X,Y unabhängig sind, so sind X,Y unkorreliert;

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

3.3 Mehrdimensionale Normalverteilung

3.14 Definition. Es seien $\mu \in \mathbb{R}^d$ 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.

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

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

3.16 Lemma. *Ist $\Sigma \in \mathbb{R}^{d\times d}$ symmetrisch und strikt positiv definit, so besitzt die $N(\mu,\Sigma)$-Verteilung eine Dichte im $\mathbb{R}^d$, nämlich*

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

3.17 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.*

3.18 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.*

3.19 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 $Z := T_1\Big(\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} T_2 W + W'\Big)$ 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 $T_2 W \sim N(0, E_d)$, weil T_2 orthogonale Matrix ist, und weiter, dass $\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} T_2 W$ 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 $\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} T_2 W + W' \sim N(0, E_m)$, und es folgt wieder nach dem Lemma $Z \sim N(0, E_m)$.

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

$$T_1 \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \Big(\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} T_2 W + W'\Big) = T_1 \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} T_2 W = A\Sigma^{1/2} W,$$

wie zu zeigen war. □

3.20 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.*

4 Einführung in statistische Tests

4.1 Hypothesentests

4.1 Definition. Ein statistisches Modell ist ein Tripel $(\mathcal{X}, \mathscr{F}, (P_\vartheta)_{\vartheta\in\Theta})$ bestehend aus einer Menge $\mathcal{X}$ mit einer σ-Algebra $\mathscr{F}$ (dem Stichprobenraum) und einer Familie $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ von Wahrscheinlichkeitsmaßen auf $\mathscr{F}$. Die mindestens zwei-elementige Menge Θ heißt Parametermenge und jedes $\vartheta \in \Theta$ Parameter.

4.2 Definition. Aufbau eines Testverfahrens:

(a) Wahl eines statistischen Modells $(\mathcal{X}, \mathscr{F}, (P_\vartheta)_{\vartheta\in\Theta})$

(b) Formulierung von Hypothese und Alternative: $\Theta = \Theta_0 \dot\cup \Theta_1$
$\vartheta \in \Theta_0$: ϑ entspricht der Hypothese H_0
$\vartheta \in \Theta_1$: ϑ entspricht der Alternative H_1

(c) Wahl eines Irrtumsniveaus $\alpha \in (0,1)$ für den Fehler erster Art, sich bei Vorliegen der Hypothese für die Alternative zu entscheiden.

(d) Konstruktion eines (randomisierten) Tests $\varphi : \mathcal{X} \to [0,1]$ zum Niveau α:
$\varphi(x) = 0$: Entscheidung für H_0,
$\varphi(x) = 1$: Entscheidung für H_1,
$\varphi(x) \in (0,1)$: Entscheidung mit Wahrscheinlichkeit $\varphi(x)$ für H_1,
$\sup_{\vartheta\in\Theta_0} \mathbb{E}_\vartheta[\varphi] \leqslant \alpha$.

(e) Durchführen des Experiments

4.3 Definition. Die Funktion $G_\varphi : \Theta \to [0,1]$ mit $G_\varphi(\vartheta) = \mathbb{E}_\vartheta[\varphi]$ heißt Gütefunktion des Tests φ. Ist φ ein Test vom Niveau α, so gilt $G_\varphi(\vartheta_0) \leqslant \alpha$ für alle $\vartheta_0 \in \Theta_0$. Für $\vartheta_1 \in \Theta_1$ bezeichnet $\beta_\varphi(\vartheta_1) = 1 - G_\varphi(\vartheta_1)$ die Wahrscheinlichkeit für den Fehler 2. Art der Entscheidung für H_0, obwohl $\vartheta_1 \in \Theta_1$ vorliegt.

4.4 Definition. Ein Test φ von $H_0 : \vartheta \in \Theta_0$ gegen $H_1 : \vartheta \in \Theta_1$ heißt gleichmäßig bester Test zum Niveau α, falls φ ein Test zum Niveau α ist und für jeden anderen Test ψ zum Niveau α gilt:

$$\forall\, \vartheta_1 \in \Theta_1 : \ \mathbb{E}_{\vartheta_1}[\varphi] \geqslant \mathbb{E}_{\vartheta_1}[\psi].$$

4.2 Neyman-Pearson-Tests

4.5 Definition. Der Likelihood-Quotient von $\mathbb{P}_1$ bezüglich $\mathbb{P}_0$ ist im diskreten Fall mit Zähldichten $p_1(x)$ und $p_0(x)$ gegeben durch

$$R(x) := \begin{cases} p_1(x)/p_0(x), & \text{falls } p_0(x) > 0, \\ +\infty, & \text{falls } p_0(x) = 0, \\ \text{beliebig}, & \text{falls } p_0(x) = p_1(x) = 0. \end{cases}$$

Im Fall von Dichten $f_0(x)$ und $f_1(x)$ im $\mathbb{R}^d$ ist $R(x)$ entsprechend definiert, indem p_0, p_1 jeweils durch f_0, f_1 ersetzt werden.

Jeder Test φ der Form

$$\varphi(x) = \begin{cases} 1, & \text{falls } R(x) > c, \\ 0, & \text{falls } R(x) < c, \\ \gamma, & \text{falls } R(x) = c \end{cases}$$

mit beliebigem $c \geqslant 0$ und $\gamma \in [0,1]$ heißt ein Neyman-Pearson-Test.

4.6 Satz. *Für das Testen von $H_0 : \vartheta = 0$ gegen $H_1 : \vartheta = 1$ gilt:*

(a) Ist φ^* ein Neyman-Pearson-Test, so gilt $\mathbb{E}_1[\varphi^*] \geqslant \mathbb{E}_1[\varphi]$ für jeden beliebigen Test φ mit $\mathbb{E}_0[\varphi] \leqslant \mathbb{E}_0[\varphi^*]$.

(b) Für jedes Niveau $\alpha \in (0,1)$ existiert ein Neyman-Pearson-Test φ^* mit exakt $\mathbb{E}_0[\varphi^*] = \alpha$.

(c) Ein (gleichmäßig) bester Test zum Niveau α ist gegeben durch einen Neyman-Pearson-Test φ^* mit $\mathbb{E}_0[\varphi^*] = \alpha$.

5 Grenzwertsätze

5.1 Gesetze der großen Zahlen

5.1 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 $\varphi(K) > 0$:*

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

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

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

5.3 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 $\sup_i \mathrm{Var}(X_i) < \infty$. Dann erfüllt das arithmetische Mittel*

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

für jedes $\varepsilon > 0$

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

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

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

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

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

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

Man sagt, dass X_n <u>P-fast sicher</u> gegen X <u>konvergiert</u>, falls

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

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

5.7 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 $\sup_i \mathrm{Var}(X_i) < \infty$. Dann konvergiert das arithmetische Mittel $A_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ fast sicher gegen μ.*

5.8 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 $\mathcal{L}^p(\Omega, \mathscr{F}, P)$, $p > 0$, und ein $X \in \mathcal{L}^p(\Omega, \mathscr{F}, P)$ 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.

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

5.10 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.

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

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

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

5.2 Konvergenz in Verteilung

5.12 Definition. Die $\mathbb{R}^d$-wertigen Zufallsvariablen $(X_n)_{n\geqslant 1}$ konvergieren in Verteilung gegen die $\mathbb{R}^d$-wertige Zufallsvariable 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 $(\mathbb{R}^d, \mathfrak{B}_{\mathbb{R}^d})$ konvergieren schwach 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 Konvergenz in Verteilung mit einem Wahrscheinlichkeitsmaß P als Limes allgemein durch $X_n \xrightarrow{d} P :\iff P^{X_n} \xrightarrow{w} P$.

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

5.14 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 (Stetigkeitspunkte von F^X).*

5.15 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.*

5.16 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 $\sup_{n\geqslant 1} P_n([-K_\varepsilon, K_\varepsilon]^\complement) < \varepsilon$.

5.17 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.*

5.3 Charakteristische Funktionen und Zentraler Grenzwertsatz

5.18 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.

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

5.20 Satz. *(Eindeutigkeitssatz) Zwei Wahrscheinlichkeitsmaße mit derselben charakteristischen Funktion sind identisch.*

5.21 Satz. *(Stetigkeitssatz von Lévy) Sind (P_n) Wahrscheinlichkeitsmaße 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 $P_n \xrightarrow{w} P$.*

5.22 Satz. *(Zentraler Grenzwertsatz) Es sei $(X_i)_{i\geqslant 1}$ 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 $N(0,1)$.

5.23 Satz. *Für alle $n \geqslant 1$ und $p \in (0,1)$ gilt folgende Fehlerabschätzung im Poissonschen Grenzwertsatz:*

$$\sum_{k\geqslant 0} |\mathrm{Bin}_{n,p}(k) - \mathrm{Poiss}_{np}(k)| \leqslant 2np^2.$$

6 Einführung in die Schätztheorie

6.1 Grundlagen

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

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

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

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

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

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 $g(\vartheta)$.

6.3 Lemma (Bias-Varianz-Zerlegung)**.** *Für jeden Schätzer $\hat{g}$ von $g(\vartheta)$ mit $R(\hat{g}, \vartheta) < \infty$ gilt*

$$R(\hat{g}, \vartheta) := |B(\hat{g}, \vartheta)|^2 + \mathbb{E}_\vartheta[|\hat{g} - \mathbb{E}_\vartheta[\hat{g}]|^2].$$

6.2 Cramér-Rao-Ungleichung und ML-Schätzer

6.4 Satz (Cramér-Rao-Ungleichung)**.** *Im statistischen Modell $(\mathbb{R}^d, \mathfrak{B}_{\mathbb{R}^d}, (P_\vartheta)_{\vartheta\in\Theta})$ seien $\Theta \subseteq \mathbb{R}$ offen, $g : \Theta \to \mathbb{R}$ differenzierbar und $\hat{g}$ ein erwartungstreuer Schätzer von $g(\vartheta)$. Weiterhin besitze jedes P_ϑ eine Dichte f_ϑ, so dass $\frac{d}{d\vartheta} f_\vartheta(x)$ für Lebesgue-fast alle $x \in \mathbb{R}^d$ existiert und folgende Vertauschungen erlaubt sind:*

$$\frac{d}{d\vartheta} \int_{\mathbb{R}^d} h(x) f_\vartheta(x)\, dx = \int_{\mathbb{R}^d} h(x) \frac{d}{d\vartheta} f_\vartheta(x)\, dx \quad \text{für } h(x) = 1,\ h(x) = \hat{g}(x).$$

Dann folgt

$$\forall \vartheta \in \Theta :\ R(\hat{g}, \vartheta) \geqslant \frac{g'(\vartheta)^2}{I(\vartheta)} \quad \text{mit } I(\vartheta) := \mathbb{E}_\vartheta\Big[\Big(\frac{\frac{d}{d\vartheta} f_\vartheta}{f_\vartheta}\Big)^2\Big],$$

sofern die Fisher-Information $I(\vartheta)$ endlich ist.

Ein vollkommen analoges Resultat gilt im Fall von Verteilungen P_ϑ mit Zähldichten p_ϑ.

6.5 Lemma. *Im Produktmodell $(\mathcal{X}^n, \mathcal{F}^{\otimes n}, (P_\vartheta^{\otimes n})_{\vartheta\in\Theta})$ gilt unter den Voraussetzungen im Satz für die Fisher-Information I_n bei n Beobachtungen $I_n(\vartheta) = nI_1(\vartheta)$.*

6.6 Definition. Ist $(\mathcal{X}, \mathcal{F}, (P_\vartheta)_{\vartheta\in\Theta})$ ein diskretes statistisches Modell mit Zähldichten p_ϑ, so heißt $L(\vartheta, x) = p_\vartheta(x)$ Likelihood-Funktion. Entsprechend definiert man in einem Modell mit Lebesgue-Dichten f_ϑ die Likelihood-Funktion als $L(\vartheta, x) = f_\vartheta(x)$. Mit $\ell(\vartheta, x) = \log(L(\vartheta, x))$ wird die Loglikelihood-Funktion bezeichnet. Man schreibt auch nur $L(\vartheta), \ell(\vartheta)$ für die entsprechenden Zufallsvariablen.

6.7 Definition. Gilt für einen Schätzer $\hat\vartheta$, dass $L(\hat\vartheta(x), x) = \max_{\vartheta\in\Theta} L(\vartheta, x)$ oder äquivalent $\ell(\hat\vartheta(x), x) = \max_{\vartheta\in\Theta} \ell(\vartheta, x)$ für alle $x \in \mathcal{X}$ erfüllt ist, so nennt man $\hat\vartheta$ Maximum-Likelihood-Schätzer (MLE).

6.8 Lemma (Plug-in-Prinzip)**.** *Ist $g : \Theta \to \Theta'$ bijektiv, so ist $g(\hat\vartheta^{MLE})$ mit dem MLE $\hat\vartheta^{MLE}$ für $\vartheta \in \Theta$ Maximum-Likelihood-Schätzer von $g(\vartheta) \in \Theta'$.*

6.3 Likelihood-Quotienten-Tests

6.9 Definition. In einem statistischen Modell mit Likelihoodfunktion $L(\vartheta, x)$ betrachte das Testproblem $H_0 : \vartheta \in \Theta_0$ gegen $H_1 : \vartheta \in \Theta_1$ mit $\Theta = \Theta_0 \dot\cup \Theta_1$. Dann heißt ein Test der Form

$$\varphi(x) := \mathbf{1}\Big(\frac{\sup_{\vartheta\in\Theta_1} L(\vartheta, x)}{\sup_{\vartheta\in\Theta_0} L(\vartheta, x)} > c_\alpha\Big)$$

Likelihood-Quotienten-Test. Mit den Maximum-Likelihood-Schätzern $\hat\vartheta_0, \hat\vartheta_1$ für Parametermengen Θ_0 bzw. Θ_1 gilt $\varphi = \mathbf{1}(L(\hat\vartheta_1) > c_\alpha L(\hat\vartheta_0))$.

6.10 Lemma. *Im Multinomialmodell mit Stichprobenraum $\mathcal{X} = \{k \in \{0, \dots, n\}^r \,|\, k_1 + \cdots + k_r = n\}$, Potenzmenge als σ-Algebra und Parametermenge $\Theta = \{\vartheta \in (0,1)^r \,|\, \vartheta_1 + \cdots + \vartheta_r = 1\}$ für $r \in \mathbb{N}$ ist der Likelihoodquotienten gegeben durch*

$$\varphi(k) := \mathbf{1}\Big(k_1 \log(\frac{k_1}{n\vartheta_{0,1}}) + \cdots + k_1 \log(\frac{k_r}{n\vartheta_{0,r}}) > c_\alpha\Big)$$

mit geeignetem kritischen Wert $c_\alpha > 0$.

Der Test φ wird durch den χ^2-Test $\tilde\varphi(k) = \mathbf{1}(V^2(k) > c'_\alpha)$ mit Pearsons χ^2-Statistik $V^2(k) = \sum_{i=1}^r \frac{(k_i - n\vartheta_{0,i})^2}{n\vartheta_{0,i}}$ und $c'_\alpha > 0$ geeignet für $n \to \infty$ approximiert.

6.11 Satz. *Sind $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ unabhängige, identisch verteilte Beobachtungen mit $\mu \in \mathbb{R}$ und $\sigma > 0$ unbekannt ($n \geqslant 2$), so ist der Likelihood-Quotienten-Test für $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$ ($\sigma > 0$ beliebig) gegeben durch den zweiseitigen t-Test der Form*

$$\varphi(x) = \mathbf{1}(|T_{n-1}(x)| > C_\alpha)$$

mit $T_{n-1} := \frac{\bar X - \mu_0}{\hat\sigma}$, wobei $\bar X := \frac{1}{n}\sum_{i=1}^n X_i$, $\hat\sigma^2 = \frac{1}{(n-1)}\sum_{i=1}^n (X_i - \bar X)^2$. Unter der Hypothese H_0 ist T_{n-1} gemäß einer Student-t-Verteilung mit $n-1$ Freiheitsgraden verteilt, das heißt mit Dichte

$$f(x) = \frac{\Gamma(n/2)}{\Gamma((n-1)/2)\sqrt{2\pi}}\Big(1 + \frac{x^2}{n-1}\Big)^{-n/2}.$$