Stochastik I Übersicht

Prof. Dr. Dirk Becherer

Stochastik I
Dirk Becherer

(Vorlesung für Studierende der Studiengänge Mathematik Bachelor Mono und Master Statistik) [1]

this version: 12. Juli 2019

Inhaltsverzeichnis

[1] Diese Übersicht soll einen Überblick über den Stoff geben aber kann nicht den Besuch der Vorlesung und Übungen ersetzten. Sie basiert auf Mitschriften von Marlene Kretschmer, Moana-Rose Marcello, Steffen Miels und Jacqueline Rosar zu meiner Vorlesung, denen ich für Ihr Engagement danke. Es gibt f.s. noch Druckfehler. Hinweise hierzu gern per Email an mich.

1 Einführung

Motivation: Beispiele und der Satz von Vitali

Stochastik	
Wahrscheinlichkeit (Modellierung, Schließen aus dem Modell)	Statistik (Schließen aus Daten auf Modelle)

Bemerkung
Ziel ist die Erklärung eines Wahrscheinlichkeitsraumes $(\Omega, \mathcal{F}, P)$, wobei
Ω die Menge der möglichen Ausgänge / Ergebnisse,
$\mathcal{F}$ eine Ereignismenge und
P das Wahrscheinlichkeitsmaß, welches dem Ereignis $A \in \mathcal{F}$ eine Wahrscheinlichkeit $P[A] \in [0,1]$ zuschreibt, ist.

Bereits im Beispiel einer nicht-endlichen Münzwurffolge führt die Definition eines geeigneten Wahrscheinlichkeitsraumes zu nicht-trivialen Problemen, deren Lösung Maßtheorie erfordert.

Beispiel

Einfaches Würfeln

$$\Omega = \{1,2,3,4,5,6\}$$

$$w \in \Omega\,, w = i : \text{Würfel zeigt Augenzahl } i$$

n-maliges Würfeln

$$\Omega = \{1,\dots,6\}^n$$

$$\Omega \ni w = (w_1,\dots,w_n) : \text{in Wurf } j \text{ wird Augenzahl } w_j \text{ geworfen}$$

$$A = \{w \in \Omega \mid w_j = w_i \quad \forall i,j\} : \text{Alle Würfel zeigen die gleiche Zahl}$$

Wir erwarten für die Gleichverteilung auf Ω, dass

$$P[A] = \frac{|A|}{|\Omega|} = \frac{6}{6^n} = 6^{-(n-1)}.$$

Das entspricht dem Laplace'schen Wahrscheinlichkeitsmaß, das jedem der endlich vielen Ausgänge $w \in \Omega$ die Wahrscheinlichkeit $P[\{w\}] = \frac{1}{|\Omega|}$ zuordnet.

einfacher Münzwurf

$$\Omega = \{0,1\}$$

$$w = 0 \leftrightarrow \text{Zahl}$$

$$w = 1 \leftrightarrow \text{Kopf}$$

n-facher Münzwurf

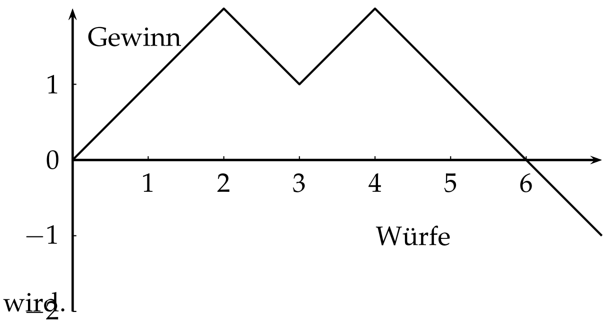
$$\Omega = \{0,1\}^n$$

∞-facher Münzwurf

$$\Omega = \{0,1\}^{\mathbb{N}}$$

$$w = (w_i)_{i \in \mathbb{N}}$$

Kumulativer Gewinn $\sum_{i=1}^{n}(2\omega_i - 1)$ ausi $n = 1,2,\ldots$ Münzwurfwetten, wenn jeweils 1 € auf "Kopf" gewettet wird.



Aktienindexentwicklung in stetiger Zeit

$$\Omega = C([0,T],\mathbb{R})$$

$$w(t) : \text{Aktienindex zur Zeit } t \leq T$$

$$A = \{w \in \Omega \mid w(t) \geq 9500\ \forall t \in [0,T]\} \leftrightarrow \text{Dax bleibt über 9500}$$

Definition 1.0 (Potenzmenge)
*Die **Potenzmenge** 2^{Ω} ist die Menge aller Teilmengen von Ω.*

Beispiel

$$\Omega = \{0,1\}$$

$$2^{\Omega} = \{\emptyset, \Omega, \{1\}, \{0\}\}$$

Bemerkung
Das Ziel ist es, jedem Ereignis A aus einem geeigneten Mengensystem $\mathcal{F} \subset 2^{\Omega}$ eine Wahrscheinlichkeit $P[A] \in [0,1]$ zuzuordnen, so dass P als Abbildung

$$P : \quad \mathcal{F} \to [0,1]$$

"schöne" Eigenschaften hat.

Frage: Warum definieren wir nicht solche P i.A. gleich auf 2^{Ω}?

Satz 1.1 (Satz von Vitali)
Sei $\Omega = \{0,1\}^{\mathbb{N}}$. Es gibt keine Abbildung $P:\ 2^{\Omega} \to [0,1]$ mit den folgenden Eigenschaften:

1. $P[\Omega] = 1$
2. *Sind $A_k \in 2^{\Omega}$ disjunkte Teilmengen von Ω, $k \in \mathbb{N}$, so gilt, dass*[2]
$$P[\biguplus_{k=1}^{\infty} A_k] = \sum_{k=1}^{\infty} P[A_k].$$
3. *$\forall A \in 2^{\Omega}$ gilt: $P[A] = P[T_k(A)]\ \forall k$, wobei T_k einen "Flip" der k-ten Koordinate beschreibt:*
$T_k:\ \Omega \to \Omega,\ (w_i)_{i\in\mathbb{N}} \mapsto\ (w_1, \dots, w_{k-1}, 1-w_k, w_{k+1}, \dots)$

Satz 1.2
Sei $\Omega := [0,1) = \mathbb{R} \mod 1$. Es gibt keine Abbildung $P: 2^{\Omega} \to [0,1)$ so dass:

i) $P[\Omega] = 1$

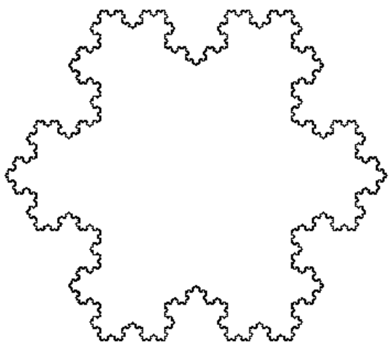
ii) $P[\biguplus_{k\in\mathbb{N}} A_k] = \sum_{k\in\mathbb{N}} P[A_k]$ *für A_k disjunkt*

iii) $P[A] = P[T_x(A)] \qquad \forall A \in 2^{\Omega}, x \in \mathbb{R}$
wobei $T_x : \Omega \to \Omega, \quad a \mapsto a + x \mod 1$
und $T_x(A) := \{T_x(a) \mid a \in A\} = \{y \in \Omega \mid y - x \mod 1 \in A\}$

Bemerkung
Forderungen i)-iii) in Satz 1 bzw Satz 2 sind im Prinzip zwar "vernünftig" und wünschenswert; was sich allerdings als nicht möglich herausstellt, ist jedoch, diese Forderungen auf der Potenzmenge zu realisieren! Ziel ist darum, ein Maßmit diesen Eigenschaften auf einer möglichst großen σ-Algebra zu definieren, welche offenbar nicht die Potenzmenge sein kann.

Beispiele dafür, dass abzählbare Additivität wünschenswert ist:

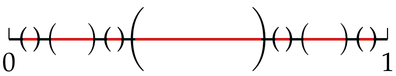
Beispiel Koch'sche Schneeflocke



Fläche $= a^2 \times 2 \times \frac{\sqrt{3}}{4}(1 + \frac{1}{3} + \frac{4}{3^3} + \frac{4^2}{3^5} + \cdots)$, wobei die Reihe konveriert.

[2]wobei wir mit der Notation $\uplus$ eine disjunkte Vereinigung bezeichnen

Beispiel Cantormenge



$\mathcal{C} := [0,1]$ ohne offenes, inneres Drittel,
ohne offene, innere Drittel der verbleibenden abgeschlossenen Intervalle,
ohne offene, innere Drittel der wiederum verbleibenden abgeschlossenen Intervalle,
...etc.(rot markiert)
Intuitiv erwarten wir, falls Intervalle die kanonische Länge haben, dass $\mathcal{C}$ die Länge $\leq (\frac{2}{3})^n \quad \forall n \in \mathbb{N}$, also die Länge 0 hat.
Die Länge von $\mathcal{C}^c$ kann man auch mit abzählbarer Additivität ausrechen

$$\frac{1}{3} + \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \left(\frac{2}{3}\right)^2 + \cdots = \frac{1}{3}\left(1 + \frac{2}{3} + \left(\frac{2}{3}\right)^2 + \cdots\right) = 1$$

was natürlich mit $1 - 0 = 1-$ Länge von C^c übereinstimmt.

Bemerkung: $\mathcal{C}$ ist überabzählbar, aber trotzdem eine Nullmenge bzgl Lebesgue Maß.

2 Wahrscheinlichkeitsräume

Axiomatische Grundlagen für $(\Omega, \mathcal{F}, P)$ und erste Eigenschaften

Definition 2.3
*Sei $\Omega \neq \emptyset$ eine nichtleere Menge. Ein Mengensystem $\mathcal{F} \subset 2^\Omega$ heißt σ-**Algebra** (auf Ω), falls*

(i) $\Omega \in \mathcal{F}$

(ii) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$, *wobei $A^c = \Omega \setminus A$ das Komplement von A in Ω bezeichnet.*

(iii) $A_k \in \mathcal{F}, k \in \mathbb{N} \implies \bigcup_{k \in \mathbb{N}} A_k \in \mathcal{F}$ *(Abgeschlossenheit bzgl abzählbarer Vereinigungen)*

Das Paar $(\Omega, \mathcal{F})$ heißt messbarer Raum oder Ereignisraum. Ein Element $A \in \mathcal{F}$ heißt messbar oder auch Ereignis.

Definition 2.4
*Sei $(\Omega, \mathcal{F})$ ein Ereignisraum. Eine Funktion $P: \mathcal{F} \to [0,1]$ heißt **Wahrscheinlichkeitsmaß**, falls*

(i) $P[\Omega] = 1$

(ii) $A_k \in \mathcal{F},\ k \in \mathbb{N},$ *disjunkt* $\implies P[\dot{\bigcup}_{k \in \mathbb{N}} A_k] = \sum_{k \in \mathbb{N}} P[A_k]$ *(σ-Additivität)*

*Wir bezeichnen das Tripel $(\Omega, \mathcal{F}, P)$ (also einen messbaren Raum $(\Omega, \mathcal{F})$ mit Wahrscheinlichkeitsmaß P) als **Wahrscheinlichkeitsraum**.*

Bemerkung
(i) und (ii) nennt man Kolmogorov'sche Axiome (Andrej N. Kolmogorov, 1903-1987, Axiomatisierung der W.theorie. Schüler u.a.: Albert Shiryeev, s. Literaturliste!)

Bemerkung
Sei $\mathcal{F}$ eine σ-Algebra auf Ω, dann gilt:

(iv) $\emptyset \in \mathcal{F}$

(v) $A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F}$, d.h. $\mathcal{F}$ ist abgeschlossen bzgl. endlichen Vereinigungen

(vi) $A, B \in \mathcal{F} \implies A \cap B \in \mathcal{F}$, d.h. $\mathcal{F}$ ist auch abgeschlossen bzgl. endlichen Schnitten, denn nach De Morganscher Regel (gilt für beliebige Indexmengen I !):

$$\bigcup_{i \in I} A_i = \left(\bigcap_{i \in I} A_i^c \right)^c$$

(vii) und natürlich auch abgeschlossen bzgl. abzählbar vielen Schnitten $A_k \in \mathcal{F}, k \in \mathbb{N} \implies \bigcap\limits_{k \in \mathbb{N}} A_k \in \mathcal{F}$

Beispiel
Sei $\Omega \neq \emptyset$. Dann ist 2^{Ω} eine σ-Algebra.

Beispiel
$\mathcal{F} = \{\emptyset, \Omega\}$ ist eine σ-Algebra.

Lemma 2.5 (Definition und Lemma)
Sei $\Omega \neq \emptyset, \quad \mathcal{G} \subset 2^{\Omega}$. Dann existiert eine kleinste σ-Algebra $\mathcal{F}$ auf Ω mit $\mathcal{G} \subset \mathcal{F}$. Wir schreiben: $\mathcal{F} = \sigma(\mathcal{G})$ für diese von $\mathcal{G}$ erzeugte (oder generierte σ-Algebra.

Definition 2.6
*Für einen topologischen Raum Ω (z.B. einen metrischen Raum (Ω, d)) bezeichnet $\mathcal{B}(\Omega)$ die **Borel σ-Algebra**, welche von den offenen Mengen erzeugt wird.*

Beispiel
$(\Omega, d) = (\mathbb{R}^n, d)$ mit $d(x, y) = |x - y|$, wir definieren $\mathcal{B}^n := \mathcal{B}(\mathbb{R}^n)$ auf $\mathbb{R}^n$

Bezeichne

$$\begin{aligned}
\mathcal{G} &:= \text{das System der offenen Mengen von } \mathbb{R}^n \\
\mathcal{A} &:= \text{das System der abgeschlossenen Mengen von } \mathbb{R}^n \\
\mathcal{K} &:= \{\prod_{i=1}^{n} [a_i, b_i] \mid a_i < b_i \text{ und } a_i, b_i \in \mathbb{Q}\}
\end{aligned}$$

Lemma 2.7
Es gilt: $\mathcal{B}^n \equiv \sigma(\mathcal{G}) = \sigma(\mathcal{K}) = \sigma(\mathcal{A})$

Lemma 2.8
Es gilt: $\mathcal{B}^1$ *wird auch erzeugt von* $\Phi_i,\ i = 1, \ldots, 4$, *d.h.*

$$\mathcal{B}^1 = \sigma(\Phi_i)$$

mit

$$\begin{aligned}\Phi_1 &= \{(-\infty, b) \mid b \in \mathbb{R}\}\\ \Phi_2 &= \{(-\infty, b] \mid b \in \mathbb{R}\}\\ \Phi_3 &= \{(a, b] \mid a < b \wedge a, b \in \mathbb{R}\}\\ \Phi_4 &= \{[a, b] \mid a < b \wedge a, b \in \mathbb{R}\}\end{aligned}$$

Beweis:
Übung bzw. Maßtheorie $\square$

Bemerkung
Analoge Aussagen gelten für $\mathcal{B}^n = \sigma(\Phi_i)$ mit z.B. $\Phi_1 = \{\prod_{i=1}^n(-\infty, b_i) \mid b_i \in \mathbb{R}\}$, etc.

Definition 2.9
Sei $(\Omega, \mathcal{F})$ *ein messbarer Raum,* $\Omega' \subset \Omega$, *dann ist*

$$\mathcal{F}' = \{A \cap \Omega' \mid A \in \mathcal{F}\}$$

eine σ*-Algebra auf* Ω' *(Übung) und heißt* ***Spur-σ-Algebra*** *(von* Ω' *auf* $\mathcal{F}$*).*

Beispiel
Borel-Mengen von $\Omega' \subset \mathbb{R}^n$ sind definiert als die Elemente der Spur-σ-Algebra von Ω' und $\mathcal{F} = \mathcal{B}^n$

Ein wichtiger Typ von σ-Algebren betrifft Messbarkeitsstrukturen auf Produkträumen, welche als (kartesisches) Produkt

$$\Omega = \times_{i \in I} E_i = \prod_{i \in I} E_i$$

über eine (beliebige) Indexmenge $I \neq \emptyset$ mit messbaren Räumen $(E_i, \mathcal{E}_i)$ gegeben sind.

Zur Defnition von Produkt-σ-Algebren betrachten wir die (kanonischen) Projektionsabbildungen und die sogenannten Zylindermengen:

$$\pi_i : \Omega \to E_i, \quad \omega = (\omega_i)_{i \in I} \mapsto \omega_i$$

$$\mathcal{G} := \{\pi_i^{-1}(A_i) \mid A_i \in \mathcal{E}_i,\ i \in I\} = \{\pi^{-1}(\mathcal{E}_i) \mid i \in I\}.$$

Definition 2.10
*Die **Produkt-σ-Algebra** auf Ω für $(E_i, \mathcal{E}_i)$, $i \in I$, ist definiert als*

$$\bigotimes_{i \in I} \mathcal{E}_i := \sigma(\mathcal{G}) = \sigma(\pi \,:\, i \in I)$$

und ist offenbar die kleinste σ-Algebra, bzgl. welcher alle Projektionen π messbare Abbildungen sind. Letztere schreiben wir wir als $\sigma(\pi_i \mid i \in I)$.

Notationen
Falls $(E_i, \mathcal{E}_i) = (E, \mathcal{E})$ $\forall i$ schreiben wir auch kürzer

$$\mathcal{E}^I = \mathcal{E}^{\otimes I} = \bigotimes_{i \in I} \mathcal{E}_i$$

und, falls I endlich ist ($|I| = n$ mit $n \in \mathbb{N}$), auch lediglich

$$\mathcal{E}^I = \bigotimes_{i=1}^{n} \mathcal{E}_i = \mathcal{E}^n$$

Bemerkung

- Es gilt: $\mathcal{B}^n \equiv \mathcal{B}(\mathbb{R}^n) = \bigotimes_{i=1}^n \mathcal{B}^1 \equiv (\mathcal{B}^1)^n$ (Übung); mit der Konvention $\mathcal{B} = \mathcal{B}^1$ also $\mathcal{B}^n \equiv \mathcal{B}(\mathbb{R}^n) = (\mathcal{B})^n$;
- Es gilt $\bigotimes_{i \in I} \mathcal{E}_i = \sigma(\pi_i \mid i \in I) = \{\pi^{-1}(\Phi_i) \mid i \in I\}$ falls Φ_i, $i \in I$, Erzeugendensysteme der $\mathcal{E}_i$ sind. (Übung!)

Es gilt: $\mathcal{B}^n \equiv \mathcal{B}(\mathbb{R}^n) = \bigotimes_{i=1}^n \mathcal{B}^1 \equiv (\mathcal{B}^1)^n$ (Übung); mit der Konvention $\mathcal{B} = \mathcal{B}^1$, also $\mathcal{B}^n \equiv \mathcal{B}(\mathbb{R}^n) = (\mathcal{B})^n$, also ist unsere Notation doppeldeutig aber dabei korrekt.

Satz 2.11 (Elementare Eigenschaften von Wahrscheinlichkeitsmaßen)
für jedes W.Maß $P : \mathcal{F} \to [0,1]$ (und weitgehend analog allgemeiner für Maße $\mu : \mathcal{F} \to [0,\infty]$) auf $(\Omega, \mathcal{F})$ gilt:

(i) $P[\emptyset] = 0$ *(Notiz: für allg. Maße μ ist diese Eigenschaft* nicht Folgerung sondern Teil der Definition, vgl [Klenke, Def.1.28])

(ii) $P[A \cup B] + P[A \cap B] = P[A] + P[B], \quad A, B \in \mathcal{F}$ (**endliche Additivität**)

(iii) $A, B \in \mathcal{F}$ und $A \subset B \implies P[A] \leq P[B]$ (**Monotonie**)

(iv) $A_i \in \mathcal{F},\ i \in \mathbb{N} \implies P[\bigcup_{i=1}^{\infty} A_i] \leq \sum_{i=1}^{\infty} P[A_i]$ (**σ-Subadditivität**)

(v) $A_n \in \mathcal{F},\ n \in \mathbb{N}$ mit $A_n \uparrow A$, d.h. $A_n \subset A_{n+1}$ $\forall n$ und $\bigcup_{n \in \mathbb{N}} A_n = A$, gilt:

$$\lim_{n \to \infty} P[A_n] = P[A] = P[\bigcup_{n=1}^{\infty} A_n]$$

und mit $A_n \downarrow A$, d.h. $A_n \supset A_{n+1}$ $\forall n$ und $\bigcap_{n \in \mathbb{N}} A_n = A$ (sowie[3] $P(A_n) < \infty$ für $n \geq N$), dann gilt: $P(A_n) < \infty$

$$\lim_{n \to \infty} P[A_n] = P[A] = P[\bigcap_{n=1}^{\infty} A_k]$$

(**σ-Stetigkeit** von unten bzw. oben von P)

[3] die Bedingung $P(A_n) < \infty$ für große n ist für W.Maße P überflüssig, da automatisch, aber i.A. nötig für nichtendliche Maßes μ.

(vi) $A_n \in \mathcal{F},\ n \in \mathbb{N}$, so dass $A_n \to A$, d.h. $\mathbb{1}_{A_n}(\omega) \to \mathbb{1}_A(\omega) \quad \forall \omega \in \Omega$, dann gilt:

$$P[A_n] \to P[A]$$

wobei $\mathbb{1}_A(\omega) := \begin{cases} 1, & \omega \in A \\ 0, & \text{sonst} \end{cases}$ die **Indikatorfunktion** einer Menge ist.

Wichtiger einfacher Spezialfall von Wahrscheinlichkeitsräumen:Ω abzählbar und $\mathcal{F} = 2^\Omega \equiv 2^\Omega$

Definition 2.12
*Falls Ω abzählbar ist und $\mathcal{F} = 2^\Omega$, so nennen wir $(\Omega, \mathcal{F}, P)$ **diskret**.*

Satz 2.13
Sei Ω abzählbar und seien $\rho(\omega) \geq 0$, $\omega \in \Omega$, Elemente in $\mathbb{R}_+$ mit $\sum\limits_{\omega\in\Omega} \rho(\omega) = 1$
Dann existiert genau ein Wahrscheinlichkeitsmaß P auf $(\Omega, 2^\Omega)$ mit $P(A) = \sum\limits_{\omega\in A} \rho(w) \quad \forall A \in 2^\Omega$
Ein solches ρ nennen wir "Zähldichte", die $\rho(\omega),\ \omega \in \Omega$, nennt man auch "Wahrscheinlichkeitsgewichte" der $\omega \in \Omega$

Diskrete Produkträume:

Seien $(E_i, \mathcal{E}_i)$ diskrete, abzählbare (endlich $|E_i| < \infty$ oder abzählbar unendlich, mit $\mathcal{E}_i = 2^{E_i}$) Maßräume mit Zähldichten ρ_i, für $i \in I = \{1, \cdots, n\}$, und bezeichne P_i jeweils das Wahrscheinlichkeitsmaß zur Zähldichte ρ_i.

Definition 2.14
Das Wahrscheinlichkeitsmaß P, das durch die Zähldichte

$$\rho(\omega) = \prod_{i=1}^n \rho_i(\omega_i) \qquad \forall \omega = (\omega_1, \cdots, \omega_n) \in \Omega := \bigtimes_{i=1}^n E_i \equiv \prod_{i=1}^n E_i$$

*auf $\mathcal{F} := \bigotimes_{i=1}^n \mathcal{E}_i = 2^\Omega$ gegeben ist (Übung), heißt **diskretes Produktmaß**. Wir schreiben[4]:*

$$P = \bigotimes_{i=1}^n \mathcal{P}_i = \prod_{i=1}^n \rho_i$$

Falls $(E_i, \mathcal{E}_i, \rho_i) = (E_1, \mathcal{E}_1, \rho_1)$ für $i = 1, \ldots n$ gilt, schreiben wir auch: $P = P_1^{\otimes n} = \rho_1^{\otimes n}$

Beispiel (n-facher Wurf mit evtl. "unfairer" Münze)

$$E = \{0, 1\}, \quad \mathcal{E} = 2^E, \quad \rho_1(1) = 1 - \rho_1(0) =: p \in [0, 1]$$
$$\Omega = E^n = \{0, 1\}^n$$

für P mit Zähldichte $\rho = \prod\limits_{i=1}^n \rho_1 = \rho_1^{\otimes n}$ liefert:

$$P[\{\omega\}] = \rho(\omega) = p^{\left(\sum_{i=1}^n \omega_i\right)} \cdot (1-p)^{\left(\sum_{i=1}^n (1-\omega_i)\right)}$$

.

[4](wobei wir bequemerweise für W.Maße auf diskreten W.Räumen die Zähldichte π mit dem entsprechenden W.Maß P nach Satz 2.13identifizieren)

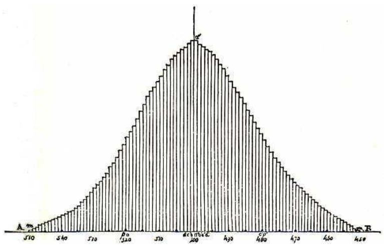
Bemerkung
Man spricht für dieses P auf $2^{\{0,1\}^n}$ von einem "Bernoulli-Schema".

Frage
Wahrscheinlichkeit für k-Erfolge in n Münzwürfen jeweils mit Erfolgswahrscheinlichkeit $p \in [0,1]$

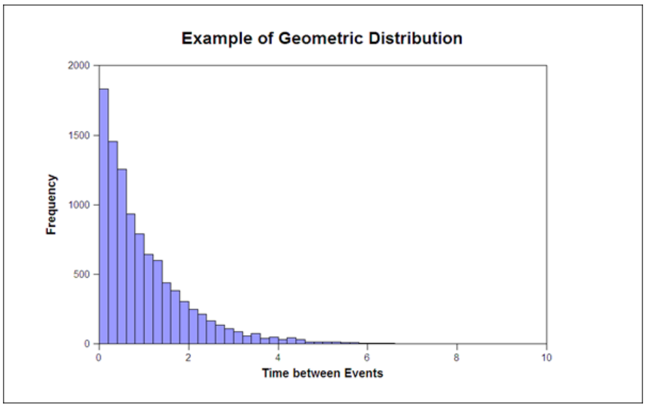
$$\begin{aligned} A &= \{\omega \in \Omega \mid \sum_{i=1}^{n} \omega_i = k\} \\ \implies P(A) &= \sum_{\omega \in A} \underbrace{P(\{\omega\})}_{p^k(1-p)^{n-k}} \\ &= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\ &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} =: \mathrm{Bin}_{n,p}(k) \end{aligned}$$

Beispiel : Beispiele für diskrete Verteilungen sind

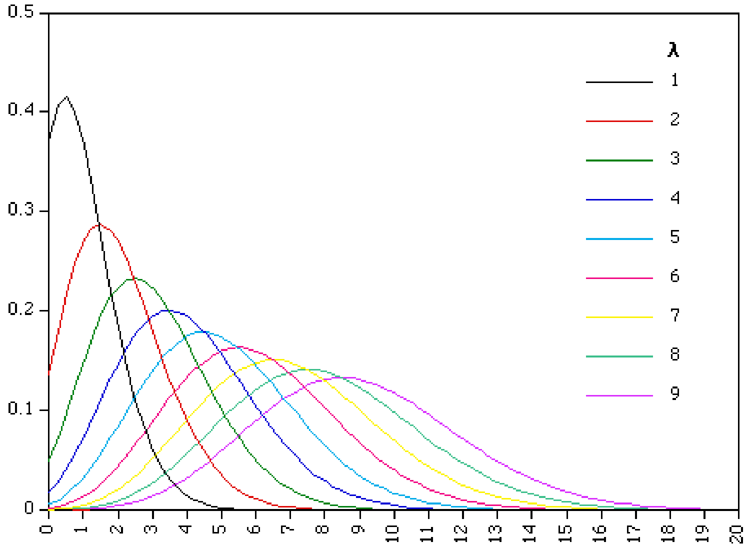
- *Binomialverteilung*
  auf $\Omega = \{0, 1, \dots, n\}$, $\quad \mathcal{F} = 2^\Omega$ mit Zähldichte $\rho(k) = \mathrm{Bin}_{n,p}(k)$, $\quad k \in \Omega$, für Parameter $n \in \mathbb{N}$, $\quad p \in [0,1]$



- *Geometrische Verteilung*
  auf $\Omega = \mathbb{N}_0 = \{0, 1, \dots\}$, $\quad p \in (0,1]$, $\quad \mathcal{F} = 2^\Omega$, $\quad \rho(k) = p(1-p)^k$, $\quad k \in \mathbb{N}_0$
  Schreibe "Wartezeitverteilung" in diskreter Zeit: Verteilung der Wahrscheinlichkeiten auf ersten Erfolg ("Kopf") bei Wurf unfairer Münze.



- *Poisson-Verteilung*
  $\Omega = \mathbb{N}_0$, $\quad \mathcal{F} = 2^\Omega$, Parameter $\lambda > 0$ ("Intensität", $\lambda \in \mathbb{R}_+$),
  Zähldichte $\rho(k) := \mathrm{Poiss}_\lambda(k) := \mathrm{e}^{-\lambda} \cdot \dfrac{\lambda^k}{k!}$, $\quad k \in \Omega$



(Hinweis: Hier wurden diskrete Histogrammdaten stetig interpoliert.)

Satz 2.15 (Poisson'scher Grenzwertsatz)
Sei $p_n \in (0,1]$ *mit* $\lim_{n\to\infty} n \cdot p_n = \lambda > 0$*. Dann gilt* $\forall k \in \mathbb{N}_0$*:*

$$\lim_{n\to\infty} \mathrm{Bin}_{n,p_n}(k) = \mathrm{Poiss}_\lambda(k)$$

Definition 2.16
Eine ***Zufallsvariable*** X *auf einem messbaren Raum* $(\Omega, \mathcal{F})$ *ist eine* $(\mathcal{F} - \mathcal{F}')$*-messbare Abbildung* $X : \Omega \to \Omega'$ *in einen messbaren Raum* $(\Omega', \mathcal{F}')$*, d.h. es gilt* $X^{-1}(A') \in \mathcal{F} \quad \forall A' \in \mathcal{F}'$*, in kompakterer Notation* $\quad X^{-1}(\mathcal{F}') \subset \mathcal{F}$*.*

Definition 2.17 (Definition und Lemma)
Sei $X : (\Omega, \mathcal{F}, P) \to (\Omega', \mathcal{F}')$ *Zufallsvariable auf dem Wahrscheinlichkeitsraum* $(\Omega, \mathcal{F}, P)$*. Dann heißt*

$$P_X := P \circ X^{-1}$$

mit $P_X(A') := P[X \in A'] = P \circ X^{-1}(A')$ *das* ***Bildmaß*** *bzw. die* ***Verteilung*** *von* X *(unter* P*) und ist ein Wahrscheinlichkeitsmaß auf dem Bildraum* $(\Omega', \mathcal{F}')$

Definition 2.18
Für $\Omega \neq \emptyset$ *ist* $\mathcal{A} \subset 2^\Omega$ *eine* ***Algebra****, falls*

(i) $\Omega \in \mathcal{A}$

(ii) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$ *(Abgeschlossenheit unter Komplementbildung)*

(iii) $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$ *(also Abgeschlossenheit bzgl. endlicher Vereinigungen)*

Definition 2.19

(i) Eine Abbildung $\mu : \mathcal{A} \to [0, \infty]$ heißt ***Prämaß*** *auf Algebra $\mathcal{A}$, falls $\mu(\emptyset) = 0$ und für*

$A_n \in \mathcal{A},\ n \in \mathbb{N}$, *disjunkt mit* $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A} \implies \mu(\biguplus_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mu(A_n)$

(ii) Ein solches μ heißt ***Maß*** *(auf $\mathcal{A}$), falls $\mathcal{A}$ eine σ-Algebra ist.*

(iii) Ein Maß μ heißt ***σ-endlich***, *falls*

$$\exists A_n \in \mathcal{A} \text{ mit } A_n \uparrow \Omega \quad (\text{ d.h. } A_n \subset A_{n+1} \text{ und } \bigcup_{n\in\mathbb{N}} A_n = \Omega) \quad \text{und mit } \mu(A_n) < \infty \quad \forall n$$

(iv) Ein Maß heißt ***Wahrscheinlichkeitsmaß***, *falls $\mu(\Omega) = 1$*

Wir nennen ein Mengensystem $\mathcal{A} \subset 2^\Omega$

(i) "abgeschlossen unter monoton wachsenden Limites", falls für $A_k \in \mathcal{A}$ mit $A_k \subset A_{k+1} \quad \forall k \implies A_k \uparrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$

(ii) "abgeschlossen bzgl. (monotoner) Differenzen", falls für $A, B \in \mathcal{A}$ mit $A \subset B \implies B \setminus A \in \mathcal{A}$

Satz 2.20 (Satz über monotone Klassen (Version für Mengensysteme))
Sei $\mathcal{C}$ ein Mengensystem ($\mathcal{C} \subset 2^\Omega$) abgeschlossen bzgl. endl. Durchschnitten und mit $\Omega \in \mathcal{C}$.

Sei $\mathcal{A}$ das kleinste Mengensystem, welches $\mathcal{C}$ enthält und (im Sinne von obigen i) und ii)) abgeschlossen bzgl. wachsender Limites und bzgl. Differenzen ist. Dann gilt: $\mathcal{A} = \sigma(\mathcal{C})$

Bemerkung

1. $\mathcal{A}$ ist wohldefiniert: (Übung)
2. Es gibt auch eine Version des Satzes für Funktionensysteme: Übung
3. in der Literatur wird eine solche Aussage teils auch mit "Dynkin"- und "π"-Systemen" formuliert (vgl. z.B. Klenke): Mit Thm.1.19 und Def.1.10 aus Klenke sieht man leicht: Mit $\Omega \in C \in \mathcal{A}$ und i) und ii) folgt: $\mathcal{A} = \delta(\mathcal{C})$ ist das von $\mathcal{C}$ erzeugte Dynkinsystem (aka λ-System), $\mathcal{C}$ is durchschnittsstabil, d.h. ein π-System in der Sprache von Klenke, und der vorige Satz ist equivalent zum $\pi - \lambda$-Theorem 1.19 in Klenke, S.7.

Korollar 2.21
Seien P, Q Wahrscheinlichkeitsmaße auf einem messbaren Raum $(\Omega, \mathcal{F})$, die auf einem durchschnitts-stabilen Erzeugersystem $\mathcal{E}$ von $\mathcal{F} = \sigma(\mathcal{E})$ übereinstimmen (kurz: auf einem $\cap$-stabilen Erzeuger von $\mathcal{F}$).
Dann gilt:

$$P = Q \qquad (\text{auf } \mathcal{F})$$

Satz 2.22 (Fortsetzungssatz von Caratheodory - Berlin 1917)
Für jedes σ-endliche Prämaß μ auf einer Algebra $\mathcal{A} \subset 2^\Omega$ existiert ein (eindeutiges) Maß $\bar{\mu}$ auf der σ-Algebra $\mathcal{F} = \sigma(\mathcal{A})$, welches auf $\mathcal{A}$ mit μ übereinstimmt (d.h. $\mu(A) = \bar{\mu}(A) \quad \forall A \in \mathcal{A}$). Wir sagen dazu "$\bar{\mu}$ setzt μ fort nach $\mathcal{F}$". Zudem ist ein solches $\bar{\mu}$ σ-endlich.

Lemma 2.23 (Eindeutigkeitssatz)
Seien $\mu,\ \nu$ σ-endliche Maße auf einem Maßraum $(\Omega, \mathcal{F})$, wobei $\mathcal{F} = \sigma(\Phi)$ für einen $\cap$-stabilen Erzeuger Φ, auf dem die Maße übereinstimmen ($\mu = \nu$ auf Φ, d.h. $\mu(A) = \nu(A)\ \forall A \in \Phi$). Es gebe $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}$ mit $A_n \uparrow \Omega$ und $\mu(A_n) = \nu(A_n) < \infty$.
Dann gilt:

$$\mu = \nu \qquad (\text{auf } \mathcal{F})$$

Lemma 2.24
Sei $\mu : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ ein Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, endlich auf Kompakta. Dann ist

$$G(x) := \begin{cases} \mu((0,x]), & \text{für } x > 0 \\ 0, & \text{für } x = 0 \\ -\mu((x,0]), & \text{für } x < 0 \end{cases}$$

monoton wachsend und rechtsstetig.

Beispiel für das Lebesgue-Maß ist $G(x) = x$

Definition 2.25
*Für ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ist seine (kumulative) **Verteilungsfunktion**[5] F gegeben durch*

$$F(x) := P\left[(-\infty, x]\right], \quad x \in \mathbb{R}.$$

Falls die Verteilungsfunktion absolutstetig (bzgl. des Lebesmaßes ist), d.h. $F(x) = \int_{-\infty}^{x} \rho(y)dy$ (Lebesgue-Integral) $\forall x$ mit einer messbaren Funktion $\rho \geq 0$, so nennen wir ρ die Dichtefunktion von P (oder F).

Bemerkung $F(x) = G(x) - G(-\infty)$ mit $G(-\infty) := \lim_{x \to \infty} G(x) > -\infty$.

Korollar 2.26
Jede Verteilungsfunktion F eines Wahrscheinlichkeitsmaßes ist rechtsstetig, monoton wachsend und erfüllt

$$\text{“}F(+\infty)\text{”} := \lim_{x \to \infty} F(x) = 1$$
$$\text{“}F(-\infty)\text{”} := \lim_{x \to -\infty} F(x) = 0$$

Wir werden zeigen, dass umgekehrt diese Eigenschaften nicht nur notwendig sondern auch hinreichend dafür sind, dass eine Fkt. F VF. eines W.Maßes P ist.

Satz 2.27
Sei $F : \mathbb{R} \to \mathbb{R}$ monoton wachsend, rechtsstetig. Dann existiert ein (eindeutiges)σ-endliches Maß μ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit

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

Beispiel

(i) für $F(x) = x$ liefert dies das **Lebesguemaß** λ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

(ii) $F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$ liefert die Gleichverteilung U auf $([0,1], \mathcal{B}([0,1]))$, d.h. λ eingeschränkt auf $[0,1]$

Beispiele (für Verteilungsfunktionen von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$)

(i)

$$F(x) := \begin{cases} 1, & x \geq a \\ 0, & x < a \end{cases} \quad a \in \mathbb{R}$$

$F(x)$ entspricht also dem ‘Dirac’sche Punktmaß” auf $a \in \mathbb{R}$ $\delta_a(A) = \begin{cases} 1, & a \in A \\ 0, & sonst \end{cases}$

[5]engl. Abkürzung “CDF” - cumulative distribution function

(ii) $\mu = p \cdot \delta_{a_1} + q \cdot \delta_{a_2} + (1-q-p) \cdot \delta_{a_3}$ für $p, q \geq 0$ mit $1-(p+q) \geq 0$.

Definition 2.28 (Definition & Lemma)

(i) Ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, dessen Verteilungsfunktion F absolutstetig ist, ist durch eine ***Dichte-Funktion***[6] *f beschrieben, mit welcher die Verteilungsfunktion F die Darstellung*

$$F(x) = \overset{\text{(Lebesgue-Integral)}}{\int_{-\infty}^{x} f(y)\mathrm{d}y} \quad , x \in \mathbb{R}$$

hat. Die Dichte ist eine messbare Funktion $f : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Um eine Verteilungsfunktion F darzustellen, muss die Dichtefunktion f (f.ü.) nicht-negativ und endlich sein, sowie integrierbar sein (mit $\int_{\mathbb{R}} f(y)dy = 1$).

(ii) Umgekehrt: Gilt

$$F(x) = \int_{-\infty}^{x} f(y)\mathrm{d}y \quad \forall x \in \mathbb{R}$$

für f nicht-negativ, messbar, und integrierbar mit

$$\int_{-\infty}^{\infty} f(y)\mathrm{d}y = 1,$$

*dann ist F die Verteilungsfunktion eines Wahrscheinlichkeitsmaßes P auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$,
denn F ist monoton, rechtsstetig mit $F(x) \to 1$ für $x \to \infty$ und $F(x) \to 0$ für $x \to -\infty$ (Übung).*

Bemerkung (und allgemeinere Definition). Ein Wahrscheinlichkeitsmaß auf $\mathbb{R}^d$ (in Dimension $d \geq 1$) hat genau dann eine Dichte[7] f, falls

$$P[B] = \int_{\mathbb{R}^d} 1_B(y) f(y) P(dy) \quad \text{for all } B \in \mathcal{B}(\mathbb{R}^d).$$

Da ein Wahrscheinlichkeitsmaß durch seine Werte auf einem durchschnittsabgeschlossenen Erzeuger der σ-Algebra eindeutig bestimmt ist, ist es äquivalent, die obige Gleichheit für alle B aus einem solchen Erzeugersystem zu fordern (z.B. für alle $B = (\infty, x]$, $x \in \mathbb{R}$, falls $d = 1$, bzw. für alle $B = \bigtimes_{i=1}^{d} (\infty, x_i]$, $x \in \mathbb{R}^d$, in Dimension $d \geq 1$).

Beispiele (für Wahrscheinlichkeitsmaße mit Dichten) (i) *Normalverteilung, Parameter $\mu \in \mathbb{R}$, $\sigma > 0$, mit Dichte*:

$$\varphi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Auch: "Gauß-Verteilung"

$\varphi \geq 0$, messbar, da stetig und $\int_{\mathbb{R}} \varphi(x)\mathrm{d}x = 1$, (Übung, ggf. Hinweis $(\int \varphi dx)^2 = 1$).

Normierung ergibt

$$\begin{aligned} \int_{\mathbb{R}} \varphi_{\mu,\sigma}(x)\mathrm{d}x &\overset{y=\frac{x-\mu}{\sigma}}{=} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \mathrm{d}y \\ &= \int_{\mathbb{R}} \varphi_{0,1}(y)\mathrm{d}y \end{aligned}$$

für die Dichte $\varphi_{0,1}$ der Standard-Normalverteilung $\mathcal{N}(0,1)$

[6] Hierbei handelt es sich um die Radon-Nikodym-Dichte bzgl. des Lebesguemaßes, während die sog. Zähldichte auf einem diskreten Wahrscheinlichkeitsraum die R.N.-Dichte bzgl. des Zählmaßist.

[7] Genauer: eine Radon-Nikodym Dichte bzgl. des (d-dimensionalen) Lebesguemaßes.

(ii) *Exponentialverteilung* (Spezialfall der Gamma-Verteilung mit $r = 1$)

mit Parameter $\alpha > 0$ hat Dichte
$$f(x) = \gamma_{\alpha,1}(x) = \mathbb{1}_{(o,\infty)}(x) \cdot \alpha \cdot \mathrm{e}^{-\alpha x}$$

(iii) *Gammaverteilung*

Hat Parameter $\alpha, r > 0$, und Dichte

$$\gamma_{\alpha,r} = \mathbb{1}_{(0,\infty)}(x) \cdot \frac{\alpha^r \cdot x^{r-1}}{\Gamma(r)} \cdot \mathrm{e}^{-\alpha x} = \begin{cases} \frac{\alpha^r x^{r-1}}{\Gamma(r)} \cdot \mathrm{e}^{-\alpha x}, & \text{für } x > 0 \\ 0, & x \leq 0 \end{cases}$$

wobei
$$\Gamma(r) := \int_0^\infty y^{r-1} \cdot \mathrm{e}^{-y} \mathrm{d}y, \quad r > 0$$

die Gamma-Funktion (aus der Analysis) ist, mit $\Gamma(r) = (r-1)!$ via $\Gamma(r) = (r-1) \cdot \Gamma(r-1)$ mit partieller Integration für $r \in \mathbb{N}$.

Anwendungsbeispiel zur Motivation der Gammaverteilung: Wahrscheinlichkeit von mindestens r Versicherungsschäden im Zeitintervall $[0, t]$, wenn Anzahl der Einzelschäden durch eine Poissonverteilung mit Parameter $\lambda = \alpha t$ gegeben ist.

Eine Erinnerung an zentrale Resultate zum Maßintegral $\int_\Omega f(\omega)\mu(\mathrm{d}\omega)$:
(lediglich Überblick, genaueres siehe Maß- & Integrationstheorie, vgl. 1.Kapitel im Buch von A.Klenke)

Betrachte messbaren Raum $(\Omega, \mathcal{F})$ mit (evtl endl. oder W.-)Maß $\mu : \mathcal{F} \to [0, \infty]$ und erkläre "$\int_\Omega f \mathrm{d}\mu = \int_\Omega f(\omega)\mu(\mathrm{d}\omega)$" für eine messbare, numerische Fkt. $f : \Omega \to \bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ in drei Schritten:

(i) für $f \geq 0$, elementare Funktion: $f = \sum_{k=1}^N \alpha_k \cdot \mathbb{1}_{A_k}, \quad A_k \in \mathcal{F}, a_k \in [0, \infty]$, erkläre

$$\int f \mathrm{d}\mu := \sum_{k=1}^N \alpha_k \mu(A_k)$$

(ii) für $f : (\Omega, \mathcal{F}) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ und $f \geq 0$, f messbar, d.h.

$$f^{-1}(\mathcal{B}(\bar{\mathbb{R}})) \subset \mathcal{F} \iff f^{-1}(B) \in \mathcal{F} \quad \forall B \in \mathcal{B}(\bar{\mathbb{R}}) \iff f^{-1}((-\infty, c]) \in \mathcal{F} \quad \forall c \in \mathbb{R},$$

erkläre:
$$\int f \mathrm{d}\mu := \sup_{\substack{g \text{ elementar} \\ 0 \leq g \leq f}} \int g \mathrm{d}\mu \quad \in \mathbb{R} \cup \{\infty\}$$

(iii) für f $\bar{\mathbb{R}}$-wertig und messbar mit $\int |f^+| \mathrm{d}\mu < \infty$ oder[8] $\int |f^-| \mathrm{d}\mu < \infty$ erkläre :

$$\int f \mathrm{d}\mu := \int f^+ \mathrm{d}\mu - \int f^- \mathrm{d}\mu \quad \in \bar{\mathbb{R}}$$

mit $f^+ = \max(0, f)$ und $f^- = \max(0, -f)$ sowie $f = f^+ - f^-$.

[8](Es genügt auch, wenn ein Summand endlich ist, um das ggf nicht-endliche Integral zu definieren)

Ist das Integral endlich definiert, so heißt die Funktion f integrierbar (bzgl. μ).
Wir schreiben dann $f \in L^1(\mu) \equiv L^1(\Omega, \mathcal{F}, \mu)$ mit L^1-Norm $||f||_{L^1(\mu)} = \int |f| \mathrm{d}\mu < \infty$.
Für $p \in [1, \infty)$ heißt f "p-integrierbar" falls $f \in L^p(\mu)$, d.h. $|f|^p \in L^1(\mu)$ mit L^p-Norm $||f||_{L^p(\mu)} = (\int |f|^p \mathrm{d}\mu)^{1/p} < \infty$.

Bemerkung

a) Konvergenzsätze für $\displaystyle\int f \mathrm{d}\mu$ Maßintegral: Sei $(f_n)_{n\in\mathbb{N}}$ eine Folge messbarer Funktionen.

   (i) **Lemma von Fatou**

   Wenn $f_n \geq 0 \quad \forall n \in \mathbb{N}$, gilt

$$\int_\Omega \liminf_{n\to\infty} f_n \mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n \mathrm{d}\mu$$

   (ii) **Satz über monotone Konvergenz**

   Falls $0 \leq f_1 \leq f_2 \leq \cdots$, so gilt

$$\lim_{n\to\infty} \int f_n \mathrm{d}\mu = \int \left(\lim_{n\to\infty} f_n \right) \mathrm{d}\mu$$

   (iii) **Konvergenzsatz von Lebesgue**

   Falls $(f_n)_{n\in\mathbb{N}}$ eine gleichgradig integrierbare Folge in $L^1(\mu)$ ist
   (hinreichend dafür ist inbesondere die Existenz einer integrierbaren Majorante[9] $g \in L^1(\mu)$ mit $|f_n| \leq g \; \forall n$)

$$(\text{d.h.} \quad \mu\left(\{\omega \in \Omega \mid |f_n(\omega)| > g(\omega) \text{ für mindestens ein } n\}\right) = 0 \; \forall n)$$

   und falls $f_n \underset{n\to\infty}{\longrightarrow} f_\infty$ μ-fast überall gilt (d.h. $\mu(\{\omega \in \Omega \mid \lim_{n\to\infty} f_n(\omega) \neq f^\infty(\omega)\}) = 0$) so folgt

$$||f_n - f_\infty||_{L^1(\mu)} = \int |f_n - f_\infty| \to 0 \text{ und } \int f_n \mathrm{d}\mu \underset{n\to\infty}{\longrightarrow} \int f_\infty \mathrm{d}\mu.$$

b) **Bezüge zum Riemann Integral**: Eine Funktion $f : [a, b] \to \mathbb{R}$, welche auf einem kompaktem Intervall $[a, b]$ Riemann-integrierbar ist, ist dort auch Lebesgue-integrierbar ist und das Riemann-Integral stimmt mit dem Leb-Integral überein. Bei uneigentlichen Integralen muß man mehr aufpassen: Hier braucht Gleichheit nicht zu gelten, selbst wenn das R-Integral endlich definiert ist. Es gilt jedoch z.B.: Für (nichtnegative) $f : \mathbb{R} \to [0, \infty)$ Riemann-integrierbar auf $\mathbb{R}$ gilt, dass f Lebesgue-integrierbar ist und das Riemann-Integral gleich dem $\displaystyle\int_\mathbb{R} f \mathrm{d}\lambda$ (Lebesgue-Integral bzgl. des Lebesgue-Maßes $\mu = \lambda$) ist (vgl. Walter, Analysis 2, 9.7 + 7.20 .

   In dieser Vorlesung benutzen wir Notationen wie $\int f(y)dy$ und meinen damit grundsätzlich das Lebesgue-Integral.

c) **Hauptsatz der Differential- und Integral-Rechnung**: In einer geeigneten Formulierung gilt auch ein Haupsatz der Differential- und Integralrechnung für das Lebesgue Integral, welcher insbesondere einen Integranden f, der auf allen Kompakta integrierbar ist, als f.ü. existierende Ableitung der Funktion $F(t) := \int_a^t f(x)dx$ identifiziert; siehe zB Walter, Analysis 2, 9.23.

Satz 2.29 (Definition und Satz)
Sei μ ein positives Maße auf einem messbaren Raum $(\Omega, \mathcal{F})$.

[9] In diesem oft nützlichen aber weniger allgemeinen Fall spricht man vom "Satz von der majorisierten (oder dominierten) Konvergenz"

1. *Sei auch ν positives Maße auf $(\Omega, \mathcal{F})$. Falls es eine messbare Funktion ρ gibt, so dass gilt*

$$\begin{aligned} \nu(A) &= \int 1_A \rho \, d\mu \quad \forall A \in \mathcal{F} & (1) \\ \Leftrightarrow \int f \, d\nu &= \int f \rho \, d\mu \quad \forall \quad \text{messbaren Funktionen } f \geq 0 \end{aligned}$$

   *so nennen wir ρ die **Radon-Nikodym-Dichte** von ν bzgl. μ, schreiben $\rho = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ und sagen "ν is absolutstetig zu μ mit R.N.-Dichte ρ.".*

2. *Falls μ ein Maßund $\rho \geq 0$ eine messbare Funktion ist, so wird durch (1) ein positives Maß ν definiert. Falls $\int \rho \, d\mu < \infty$ gilt, ist ν dabei ein endliches Maß. Falls $\int \rho \, d\mu = 1$ gilt, ist ν ein Wahrscheinlichkeitsmaß. Die R.N.-Dichte ρ von von ν bzgl. μ ist μ-f.ü. eindeutig.*

Beispiel
- Sei μ Zählmaß auf einem abzählbaren Ω mit $\mathcal{F} = 2^{\Omega}$. Sei $\rho : \Omega \to \mathbb{R}_+$ eine Zähldichte auf Ω. Dann ist

$$P(A) := \sum_{\omega \in A} \rho(\omega) = \int 1_A \rho d\mu$$

  ein W.Maß und die Zähldichte ρ ist offenbar die Radon-Nykodym Dichte von P bzgl dem Zählmaß.
- Sei λ das Lebesguemaß auf $(\mathbb{R}, \mathcal{B})$ und $\rho \geq 0$ messbar mit $\int \rho \, d\lambda = 1$ die Dichte einer (stetigen) Wahrscheinlichkeitsverteilung P auf $\mathbb{R}$ mit Verteilungsfunktion $F(x) = P((-\infty, x])$. Dann ist

$$P(A) := \int 1_A \rho \, d\lambda \ \forall A \in \mathcal{B}.$$

  ein (stetiges) W.Maß auf $\mathbb{R}$ mit der Dichte ρ und Verteilungsfunktion $F(x) = \int_{-\infty}^{x} \rho(y) dy \equiv \int_{-\infty}^{x} \rho(y) \lambda(dy)$, $x \in \mathbb{R}$, ist. Die Dichte ρ einer stetigen W.Verteilung P ist also die Radon-Nykodym Dichte von P bzgl dem Lebesguemaß.

Beispiel konkrete Beispiele
Verteilungen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ können durchaus weder diskret noch stetig sein:

a) z.B. $P = \frac{1}{2}\delta_1 + \frac{1}{2} \cdot \underbrace{\mathcal{U}_{[0,1]}}_{\text{Gleichverteilung auf } [0,1]}$ Verteilungsfunktion F von P:

b) auch stetige Verteilungsfunktionen F auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ brauchen keine Dichte haben. Betrachte als F zB. die stetige Verteilungsfunktion, welche nur "gleichverteilt" auf Cantormenge (Lebesgue-Nullmenge!) wächst, und auf ihrem Komplement konstant ist ($\to$ Übung).

3 Bedingte Wahrscheinlichkeit und Unabhängigkeit

Motivation:"für beliebige Wahrscheinlichkeiten "
betrachten die Studie eines (billigen, schnellen, gut verträglichen) Tests auf eine Krankheit, welche anderweitig (aufwendig oder teuer) feststellbar ist.

Test mit 1000 Versuchsprobanden

	negativ	positiv	Σ
krank	1	9	10
gesund	970	20	990
Σ	971	29	1000

Frage: Diagnose bei Testergebnis ”positiv “(d.h. Test signalisiert “krank”)?

- Anteil der gesunden unter den positiv getesteten: $\frac{20}{29} \approx 69\%$
- Anteil der tatsächlich kranken unter den positiv getesteten: $\frac{9}{29} \approx 31\%$

Anderseits gilt unter der Bedingung, dass ein negatives Testergebnis vorliegt:

- Anteil der gesunden unter den negativ getesteten: $\frac{970}{971} \approx 99,9\%$
- Anteil der kranken unter den negativ getesteten: $\frac{1}{971} \approx 0,1\%$

Diskutiere: Wie gut ist dieser Test?

Definition 3.30
Seien $(\Omega, \mathcal{F}, P)$ ein W.raum, $A, B \in \mathcal{F}, P[B] > 0$. Dann heißt

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$

*die **bedingte Wahrscheinlichkeit** von A gegeben B, oder auch: die Wahrscheinlichkeit von A unter der Bedingung B.*

Bemerkung: Mit der Konvention $P[B]P[A|B] := 0$, falls $P[B] > 0$ ist, gilt allgemein $P[B]P[A|B] = P[A \cap B]$.

Satz 3.31
Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, $B \in \mathcal{F}$ und $P[B] > 0$.

(i) *$\mathcal{F} \ni A \mapsto Q(A) := P[A|B]$ ist ein Wahrscheinlichkeitsmaß auf $\mathcal{F}$,*

(ii) ***Formel totaler Wahrscheinlichkeit**:*
Sei $B = \dot{\bigcup}_{i \in I} B_i$ für disjunkte $B_i \in \mathcal{F}$, mit I abzählbar. Dann gilt

$$P[A \cap B] = \sum_{i \in I} P[B_i] \cdot P[A|B_i],$$

(iii) *(Bayes-formel) Seien $A \in \mathcal{F}$ mit $P(A) > 0$, I abzählbar, disjunkte $(B_i)_{i \in I} \in \mathcal{F}$ und $\Omega = \dot{\bigcup}_{i \in I} B_i$. Dann gilt*

$$P[B_j|A] = \frac{P[B_j] \cdot P[A|B_j]}{\sum_{i \in I} P[B_i] \cdot P[A|B_i]}. \tag{2}$$

Lemma 3.32 (Multiplikationsformel)
$(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, $A_1, \ldots, A_n \in \mathcal{F}$ mit $P\left[\bigcap_{i=1}^{n-1} A_i\right] > 0$, dann ist

$$P[A_1 \cap \cdots \cap A_n] = P[A_1] \cdot P[A_2|A_1] \cdot \ldots \cdot P[A_n|A_1 \cap \cdots \cap A_{n-1}]$$

Bemerkung
Wahrscheinlichkeitsmaße auf endlichen Produkträumen können im Baumschema dargestellt werden: Interpretation als mehrstufige Experimente

a) Beispiel "Patient "
Pfadwahrscheinlichkeiten:

$$
\begin{aligned}
(k,+) &= \frac{10}{1000} \cdot \frac{9}{10} = \frac{9}{1000} \\
(k,-) &= \frac{1}{1000} \\
(g,+) &= \frac{20}{1000} \\
(g,-) &= \frac{970}{1000}
\end{aligned}
$$

b) $2-$maliges Ziehen aus einer Urne mit R roten und B blauen Kugeln ohne Zurücklegen.

Definition 3.33 (Unabhängigkeit)
Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und I eine (beliebige) Indexmenge.

a) ***Zwei Ereignisse*** *$A, B \in \mathcal{F}$ heißen (paarweise) unabängig, falls $P[A \cap B] = P[A] \cdot P[B]$.*

b) ***Eine Familie*** *$A_i \in \mathcal{F}$, $i \in I$,* ***von Ereignissen*** *heißt unabhängig, falls $\forall$ endlichen Teilmengen $J \subset I$ gilt:*

$$P\left[\bigcap_{i\in J} A_i\right] = \prod_{i\in J} P[A_i].$$

c) ***Eine Familie von Mengensystemen*** *$\mathcal{A}_i \subset \mathcal{F}$, $i \in I$, heißt unabhängig, falls jede Auswahl $A_i \in \mathcal{A}_i$ eine unabhängige Familie $(A_i)_{i\in I}$ von Ereignissen liefert.*

d) ***Eine Familie von Zufallsvariablen*** *$Y_i : (\Omega, \mathcal{F}) \to (\Omega_i, \mathcal{F}_i)$, $i \in I$, heißt unabhängig, falls die σ-Algebren $\sigma(Y_i) = Y_i^{-1}(\mathcal{F}_i)$, $i \in I$, unabhängig sind.*

Bemerkung (zur Definition) a) Sind $(A_i)_{i\in I}$ unabhängig, dann folgt: A_i und A_j sind paarweise unabhängig $\forall i \neq j \in I$. Die umgekehrte Implikation gilt <u>nicht</u>.

b) $(A_i)_{i\in I}$ unabhängig $\Leftrightarrow$ $(\{\emptyset, A_i, A_i^c, \Omega\})_{i\in I}$ sind unabhängige σ-Algebren (Übung)

c) $(A_i)_{i\in I}$ unabhängig $\Leftrightarrow$ $(\mathbb{1}_{A_i})_{i\in I}$ unabhängig

d) A und B (mit $P[B] > 0$) sind unabhängig $\Leftrightarrow$ $P[A|B] = P[A]$

e) Sind die Mengensysteme $\mathcal{A}_i$, $i \in I$, durchschnittsstabil, so gilt:
$\mathcal{A}_i$, $i \in I$, sind unabhängig $\Leftrightarrow$ $\sigma(\mathcal{A}_i)$, $i \in I$, sind unabhängig

f) Alle Teile der Definition lassen sich als Spezialfälle der Unabhängigkeit von σ-Algebren (wie in d)) auffassen, indem Sie die geeignet gegnerierten σ-Algebren betrachten. In diesem Sinne ist dann auch z.B. die Unabhängigkeit einer Zufallsvariablen von einer σ-Algebra (oder von einem Ereignis) zu verstehen.

Satz 3.34
Situation von Def.3.33 (d). Sei $\mathcal{E}_i$ ein $\cap$-stabiler Erzeuger von $\mathcal{F}_i = \sigma(\mathcal{E}_i)$, $i \in I$. Für $J \subset I$, J endlich, $B_i \in \mathcal{E}_i$ mit $i \in J$ gelte:

$$P\left[\bigcap_J Y_i^{-1}(B_i)\right] = \prod_{i\in J} P\left[Y_i \in B_i\right] \tag{1}$$

(das heißt $(\{Y_i^{-1}(E_i) \mid E_i \in \mathcal{E}_i\})_{i\in I}$ sind unabhängig).
Dann sind $(Y_i)_{i\in I}$ ***unabhängig****.*

Korollar 3.35
Sei $(Y_i)_{i\in\{1,\ldots,n\}}$ eine endliche Familie von Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$. Dann gilt:

(i) diskreter Fall:

falls $Y_i : (\Omega, \mathcal{F}) \to (\Omega_i, \mathcal{F}_i)$ mit Ω_i abzählbar, $\mathcal{F}_i = 2^{\Omega_i}$ $\forall i$, so sind $(Y_i)_{i=1,\ldots,n}$ unabhängig genau dann, wenn $\underbrace{P[Y_i = \omega_i, i = 1,\ldots,n]}_{P\circ Y^{-1}(\{(\omega_1,\ldots,\omega_n)\})} = \prod_{i=1}^{n} \underbrace{P[Y_i = \omega_i]}_{P\circ Y_i^{-1}(\{w_i\})} \quad \forall w_i \in \Omega_i$

(ii) reellwertiger Fall:

$Y_i : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), \forall i$ reellwertige Zufallsvariablen, so sind $(Y_i)_{i=1,\ldots,n}$ unabhängig genau dann, wenn $\underbrace{P[Y_i \le c_i, i = 1,\ldots,n]}_{\text{gemeinsame Verteilungsfunktion von } \mathbb{R}^n\text{-wertigen } Y=(Y^1,\ldots,Y^n)} = \prod_{i=1}^{n} \underbrace{P[Y_i \le c_i]}_{\text{Verteilungsfunktion von } Y_i} \quad \forall c_i \in \mathbb{R}$

(iii) reellwertiger und absolutstetiger Fall $\subset (ii)$

Y_i wie in (ii) und zudem so, dass $P_{Y_i} := P \circ Y_i^{-1}$ absolutstetige Verteilung auf $(\mathbb{R}, \mathcal{B})$ mit Dichtefunktion $f_i(\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ sind $\forall i = 1,\ldots,n$. Dann gilt: $(Y_i)_{i=1,\ldots,n}$ unabhängig genau dann, wenn $Y = (Y^1, \ldots Y^n) : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}^n(\mathbb{R}^n))$ eine absolutstetige Verteilung $P \circ Y^{-1}$ auf $(\mathbb{R}^n, \mathcal{B}^n(\mathbb{R}^n))$ mit Dichte $f_Y(y) = f_Y(y_1,\ldots,y_n) = \prod_{i=1}^n f_i(y_i)$ hat.

Beispiel
Y_1, Y_2 sind unabhängig und jeweils $\mathcal{N}(0,1)$-verteilt $\Leftrightarrow Y = (Y_1, Y_2)$ hat eine absolutstetige Verteilung mit Dichte

$$\rho_y(y) = \rho_1(y_1)\rho_2(y_2) = \frac{1}{2\pi}\exp(-\frac{y_1^2 + y_2^2}{2})$$

Diese Verteilung von Y ist die Standard Normalverteilung mit den Parametern Mittelwertvektor $0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ und Kovarianzmatrix $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Notation: $Y \sim \mathcal{N}(0, I_2)$

Bemerkung

1. Die gemeinsame Verteilung einer Familie $Y_i = (\Omega, \mathcal{F}) \to (\Omega_i, \mathcal{F}_i), i \in I$, von Zufallsvariablen ist die Verteilung $P \circ Y^{-1}$ der Zufallsvariablen $Y = (Y_i)_{i\in I} : (\Omega, \mathcal{F}) \to (\bigtimes_{i\in I} \Omega_i, \otimes_{i\in I}\mathcal{F}_i)$
2. Jede mehrdimensionale Zufallsvariable $Y = (Y_i)_{i\in I}$ induziert eine Verteilung $P \circ Y^{-1}$ auf dem Produktraum $(\times_{i\in I}\Omega_i, \otimes_{i\in I}\mathcal{F}_i)$. Dies ist die Verteilung $P \circ Y^{-1}$ von Y, d.h. die gemeinsame Vertelung der Y_i, $i \in I$.
3. Die gemeinsame Verteilung induziert die (ein-dimensionalen, bzw. allgemeiner die endlich-dimensionalen) Randverteilungen der jeweiligen Y_i (bzw. der $(Y_{j_1}, \ldots Y_{j_n})$ für endliche $J = \{j_1, \ldots, j_n\} \subset I, n \in \mathbb{N}$). Insbesondere induziert jede endlich dimensionale Verteilung von $Y = (Y_1, \ldots, Y_n)$ ihre Randverteilungen:

$$\begin{aligned} P \circ Y_i^{-1}(B_i) &= P[Y_i \in B_i] \\ &= P[Y_1 \in \Omega_1, \ldots, Y_i \in B_i, \ldots, Y_n \in \Omega_n] \\ &= P \circ Y^{-1}(\Omega_1 \otimes \cdots \otimes B_i \otimes \cdots \otimes \Omega_n) \end{aligned}$$

Satz 3.36
Sei $(\Omega_i, \mathcal{F}_i, P_i), i \in I$, eine Familie von Wahrscheinlichkeitsräumen. Dann existiert genau ein Wahrscheinlichkeitsmaß P auf $\Omega := \times_{i\in I}\Omega_i$ und mit $\mathcal{F} := \otimes_{i\in I}\mathcal{F}_i := \sigma(\pi_i,\ i \in I)$ (mit $\mathcal{F}$, die kleinste σ-Algebra auf dem Produktraum bezüglich welcher alle Koordinatenprojektionen $\pi_i : \omega = (\omega_i)_{i\in I} \mapsto \omega_i,\ i \in I$ messbar sind), so dass für alle endlichen Teilmengen $J \subset I$ gilt:

$$P(\bigcap_{i\in J} \pi_i^{-1}(A_i)) = \prod_{i\in J} P_i(A_i) \quad \forall A_i \in \mathcal{F}_i$$

Dieses P heißt Produktmaß. Notation: $P = \otimes_{i\in I} P_i$

Beweis: (Für MathematikerInnen)
(Maßtheorie. Siehe auch im Georgii Satz 3.26 für konstruktive Beweise für einfachere aber wichtige Fälle abzählbarer Produkte.) Für (allgemeinere) Beweise mittels Kolmogorovs Konsistenzsatz s. zB. Stochastik-2 VL, oder in Terence Taos Buch für reguläre Maße auf topologischen Räumen, oder die W.-theoriebücher von A.Klenke (Ch.14) oder H.Bauer. Die Konstruktion von Ionescu-Tuscea (vgl. Klenke, oder Gänssler/Stute) zeigt eine allgemeinere Aussage mit stochastischen Kernen für abzählbare (semi-direkte) Produkte. Im folgenden nutzen wir ein vereinfachtes ähnlicher Argument für zunächst abzählbare Produkte von Maßen, welches sich unter Nutzung der Struktur produktmessbarer Mengen dann jedoch auf beliebige Produkte verallgemeinert:

Bezeichne $\pi_J^{J'} : \bigtimes_{j\in J'} \Omega_j \to \bigtimes_{j\in J} \Omega_j$ für $J \subset J' \subset I$ die kanonische Projektion. Und $\pi_J := \pi_J^I$. Für diesen Beweis bezeichne zudem $\pi_n := \pi_{\{1,\ldots,n\}}$ für $\in I$.

Die Eindeutigkeit von P folgt daraus, dass die Definition P auf dem $\cap$-stabilen Erzeugersystem der (endlich dimensionalen) Zylindermengen $\pi_J^{-1}(A)$ mit endlichem $J \subset I$, $A = \bigtimes_{j\in J} A_j$, $A_j \in \mathcal{F}_j$, festlegt.

Zur Existenz:

1. Für endliches I ($|I| < \infty$) ist die Aussage bereits bekannt (vgl. z.B. Klenke Th.1.61) aus den Grundlagen der Maßtheorie oder Analysis-3 VL.
2. Für abzählbar unendliche $I = \mathbb{N}$ argumentieren wir wie folgt:

Definiere $P(A) := (\bigotimes_1^n P_i)(A_n')$ für $A \in \bigotimes_{i\in\mathbb{N}} \mathcal{F}_i$ der Form $A = A_n' \times \bigtimes_{i=n+1}^{\infty} \Omega_i$ mit $A_n' \in \bigotimes_{i=1}^n \mathcal{F}_i$, $n \in \mathbb{N}$. Mengen solcher Form $A = \pi_n^{-1}(A_n')$ bilden eine Algebra

$$\mathcal{A} := \bigcup_{n\in\mathbb{N}} \pi_n^{-1}\left(\{A_n' \ :\ A_n' \in \bigotimes_{i=1}^n \mathcal{F}_i\right).$$

Definiere nun $P : \mathcal{A} \to [0,1]$ durch $P(A_n) := (\bigotimes_{i=1}^n P_i)\,(A_n')$.

Beh.: Dann ist P ein Prämaß.

Dass P Inhalt ist, ist klar (additiv, $P(\emptyset) = 0$). Es genügt also, für σ-Additivität zu zeigen, dass aus $\mathcal{A} \ni A_n \downarrow \emptyset$ folgt, dass $P(A_n) \to 0$. Wir zeigen hierfür : für fallenden $A_n \downarrow$ mit $\lim_n P(A_n) \downarrow \delta > 0$ gilt $\bigcap_n A_n \neq \emptyset$.

O.b.d.A. habe A_n die Form $A_n = A_n' \times \bigtimes_{i\geq n+1} \Omega_i = \pi_n^{-1}(A_n')$ mit $A_n' \in \bigotimes_{i=1}^n \mathcal{F}_i$.

Aus $A_n \supseteq A_{n+1}$ folgt $A_n' \times \Omega_{n+1} \supseteq A_{n+1}'$.

Definiere für $n > m \geq 1$ die Funktionen

$$h_{m,n}(\omega_1, \ldots, \omega_m) := \int_{\Omega_{m+1}} \cdots \int_{\Omega_n} 1_{A_n'}(\omega_1, \ldots, \omega_m, \omega_{m+1}, \ldots, \omega_n) P_n(d\omega_n) \cdots P_{m+1}(d\omega_{m+1}).$$

Die Funktionenfolge $h_{m,n}$, $n \geq 1$, ist fallend in n. Sei

$$h_m := \lim_{n\to\infty} h_{m,n} = \inf_{n>m} h_{m,n}$$

der montone (punktweise) Limes. Mit monotoner Konvergenz folgt

$$\int_{\Omega_1} \cdots \int_{\Omega_m} h_m(\omega_1, \ldots, \omega_m) P_1(d\omega_1) \cdots P_m(d\omega_m) = \lim_{n\to\infty} \int_{\Omega_1} \cdots \int_{\Omega_m} h_{m,n}(\omega_1, \ldots, \omega_m) P_1(d\omega_1) \cdots P_m(d\omega_m) \tag{3}$$

$$= \lim_{n\to\infty} P(A'_n) \ \geq \delta \ > 0\,. \tag{4}$$

Hieraus folgt induktiv dass $\bigcap_n A_n \neq \emptyset$ gilt, wie folgt:

Induktionanfang ($m = 1$): Es gibt ein $\bar\omega_1 \in \Omega_1$ mit $h_1(\bar\omega_1) \geq \delta > 0$.
Induktionsvorrausetzung: Es gebe $\bar\omega_i \in \Omega_i$, $1 \leq i \leq m$ mit $h_m(\bar\omega_1, \ldots, \bar\omega_m) \geq \delta > 0$.
Induktionsschritt ($m \to m+1$): Dann folgt mit dem Satz der monotonen Konvergenz

$$\int_{\Omega_{m+1}} h_{m+1}(\bar\omega_1, \ldots, \bar\omega_m, \omega_{m+1}) P_{m+1}(d\omega_{m+1}) = \int_{\Omega_{m+1}} \inf_{n>m+1} h_{m+1,n}(\bar\omega_1, \ldots, \bar\omega_m, \omega_{m+1}) P_{m+1}(d\omega_{m+1}) \tag{5}$$

$$= \inf_{n>m+1} \int_{\Omega_{m+1}} h_{m+1,n}(\bar\omega_1, \ldots, \bar\omega_m, \omega_{m+1}) P_{m+1}(d\omega_{m+1}) \tag{6}$$

$$= \inf_{n>m+1} h_{m,n}(\bar\omega_1, \ldots, \bar\omega_m) \ \geq \delta > 0\,. \tag{7}$$

Dann gibt es $\bar\omega_{m+1} \in \Omega_{m+1}$ derart, dass $h_{m+1}(\bar\omega_1, \ldots, \bar\omega_m, \bar\omega_{m+1}) \geq \delta > 0$.

Für jedes m ist daher $(\bar\omega_1, \ldots, \bar\omega_m) \in A'_m$, mithin $(\bar\omega_i)_{i\in\mathbb{N}} \in A_n$ für alle n und somit $(\bar\omega_i)_{i\in\mathbb{N}} \in \bigcap_n A_n \neq \emptyset$.

Also ist P ein Prämaß.

Mit dem Fortsetzungssatzes von Caratheodory folgern wir nun: Es existiert eine (eindeutige) Fortsetzung von P zu einem Maß auf der Produkt-σ-Algebra.

3. Die Existenz für den Falle einer überabzählbaren Indexmenge I folgt nun aus dem bereits gezeigten, wie folgt:

   In der Tat (siehe z.B. [Klenke, Def.14.9 und Exercise 14.1.1] oder [Gänssler/Stute, Satz 1.3.11]) ist jedes $A \in \mathcal{F} := \otimes_{i\in I}\mathcal{F}_i$ von der Form $A = A'_J \times \bigtimes_{i\notin J} \Omega_i = \pi_J^{-1}(A'_J)$ für ein abzählbares $J \subset I$ und $A'_J \in \otimes_{j\in J}\mathcal{F}_j$. Mit der (widerspruchsfreien, wieso nämlich?) Definition $P(A) := P_J(A'_J)$ mit $P_J := \bigotimes_{j\in J} P_j$ (aus vorigem Teil) folgt die Behauptung.

□

Bemerkung
Insbesondere gilt für die Randverteilung der i-ten Koordinate unter P, dass

$$P[\pi_i^{-1}(A_i)] = P_i(A_i), \quad A_i \in \mathcal{F}_i$$

d.h. die P_i sind die Randverteilung der kanonischen Koordinatenprojektion π_i.

Korollar 3.37

1. *zu gegebenen Wahrscheinlichkeitsmaßen P_i auf $(\Omega_i, \mathcal{F}_i)_{i\in I}$ existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit unabhängigen Zufallsvariablen $(X_i)_{i\in I}$, $X_i : (\Omega, \mathcal{F}) \to (\Omega_i, \mathcal{F}_i)$, so dass $P \circ X_i^{-1} = P_i$ gilt.*
2. *Zufallsvariable $Y = (Y_i)_{i\in I}$ auf einem W.raum sind unabhängig genau dann wenn ihre gemeinsame Verteilung das Produktmaß der Einzelverteilungen ist: $P_Y = \otimes_I P_{Y_i}$.*

Bedingte Verteilungen aus gemeinsamen Verteilungen

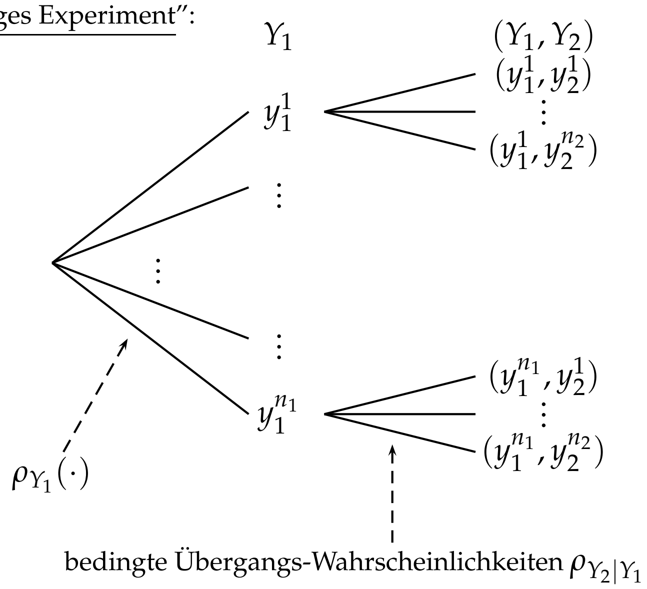
a) Diskreter Fall:
$\Omega = \Omega_1 \times \Omega_2, \mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2, Y = (Y_1, Y_2)$ diskrete Zufallsvariable mit Zähldichte $\rho(y_1, y_2)$. Die bedingte Verteilung von Y_2 gegeben $Y_1 = y_1$ ist beschrieben durch die bedingte Zähldichte:

$$\rho_{Y_2|Y_1}(y_2|y_1) = \frac{\rho(y_1, y_2)}{\sum\limits_{y_2 \in \Omega_2} \rho(y_1, y_2)} = \frac{\rho(y_1, y_2)}{\rho_{Y_1}(y_1)}$$

jene ist definiert für P_{Y_1}-fast alle y_1 [an welchen $\rho_{Y_1}(y_1) > 0$].

Visualisierungen/Veranschaulichungen

i) "mehr- bzw. zwei-stufiges Experiment":



ii) Kontingenztafel, z.B.

$Y_1 \setminus Y_2$	A	B	Randverteilung ↓
A	10%	80%	90%
B	1%	9%	10%
Randverteilung →	11%	89%	100%

während man die Randverteilungen der einzelnen Koordinaten durch aufsummieren (integrieren) über die jeweiligen Zeilen bzw. Spalten bestimmt, erhalten wir die bedingten Verteilungen durch Re-normalisierung der Wahrscheinlichkeiten auf der gegebenen Spalten bzw. Zeilen: Gegeben dass Y_1 in A ist, ist zB. die WS. dass Y_2 in B ist 8/9.

b) Absolutstetiger Fall:
Sei $Y = (Y_1, Y_2)$ absolutstetig mit Dichte $f(y_1, y_2)$ (bzgl. Leb.maß). Die bedingte Dichte von Y_2 gegeben $Y_1 = y_1$ ist

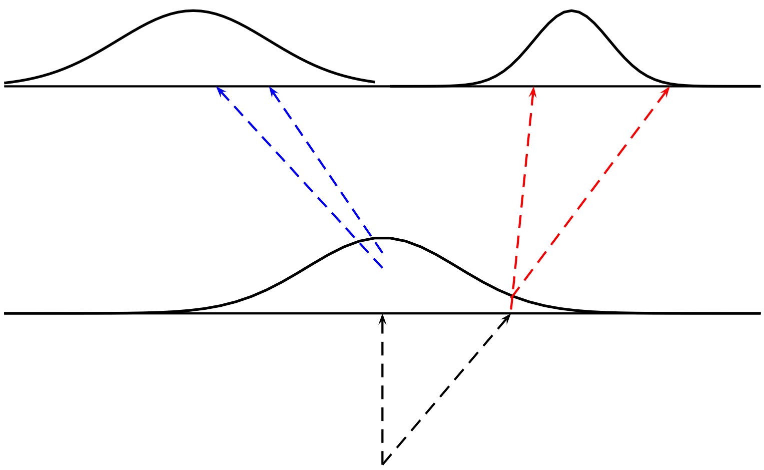
$$f_{Y_2|Y_1}(y_2|y_1) := \frac{f(y_1, y_2)}{\int_{\Omega_2} f(y_1, y_2) \mathrm{d}y_2} \equiv \frac{f(y_1, y_2)}{f_{Y_1}(y_1)}, \quad y_1 \in \Omega_1,\ y_2 \in \Omega_2, \text{mit } f_{Y_1}(y_1)\text{greater}0 \ (:= 0 \text{ sonst})$$

(ist definiert für P_{Y_1} -fast alle y_1). Dann gilt z.B.

$$P[Y_1, Y_2 \in A \times B] = \int_{\Omega_1 \times \Omega_2} \underbrace{\mathbb{1}_{A\times B}}_{\mathbb{1}_A \mathbb{1}_B} f(y_1, y_2) \mathrm{d}y_1 \mathrm{d}y_2 = \int_{\Omega_1} \mathbb{1}_A f_{Y_1}(y_1) \left(\int_{\Omega_2} \mathbb{1}_B f_{Y_2|Y_1}(y_2|y_1) \mathrm{d}y_2 \right) \mathrm{d}y_1$$

Visualisierung/Veranschaulichung : Man stelle sich vor, dass die 2te Koordinate eines mehrstufigen Experimentes gemäß einer bedingten Dichte $f_{Y2|Y1}(\cdot|y_1)$ gezogen wird, welche mit der Realisierung des Wertes y_1 für die erste Koordinate Y_1 variiert:

i) mehr- bzw. zwei-stufiges Experiment



Ziehe in erster Stufe des Experimentes Merkmal Y_1 gemäß Verteilung P_{Y_1} mit Dichte f_{Y_1}, und anschließend, bedingt auf die Realisierung von $Y_1 = y_1$, das zweite Merkmal Y_2 gemäß der (bedingten) Verteilung $P_{Y_2|Y_1}$ mit bedingter Dichte $y_2 \mapsto f_{Y_2|Y_1}(y_2|y_1)$.

ii) Das Analogon zur Kontingenztafel bei diskreten Zufallsvariablen ist hier die gemeinsame Verteilung, die durch einen Plot der gemeinsamen Dichtefunktion visualisiert werden kann.

Bemerkung Analoge Aussagen gelten für den "n-stufigen Fall" in Dimension $n \geq 2$.

Satz 3.38
Seien $(Y_i)_{i\in I}$ eine Familie von unabhängigen Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$, $Y_i : (\Omega, \mathcal{F}) \mapsto (\Omega_i, \mathcal{F}_i)$ messbare Abbildung mit $I = \bigcupdot_{k\in K} I_k$, I_k disjunkt, und seien

$$\varphi_k : (\bigtimes_{i\in I_k} \Omega_i, \bigotimes_{i\in I_k} \mathcal{F}_i) \mapsto (\tilde{\Omega}_k, \tilde{\mathcal{F}}_k), \quad k \in K, \quad \textit{messbare Abbildungen.}$$

Dann sind $\tilde{Y}_k := \varphi_k((Y_i)_{i\in I_k})$, $k \in K$, unabhängig.

4 Asymptotische Ereignisse

$(\Omega, \mathcal{F}, P)$ Wahrscheinlichkeitsraum, $(Y_k)_{k\in\mathbb{N}}$ Folge von Zufallsvariablen $Y_k : (\Omega, \mathcal{F}) \to (\Omega_k, \mathcal{F}_k)$

Definition 4.39
*Ein Ereignis $A \in \mathcal{F}$ heißt **asymptotisch** bzgl. $(Y_k)_{k\in\mathbb{N}}$, falls*

$$\forall n \in \mathbb{N}\ \exists B_n \in \bigotimes_{k\geq n} \mathcal{F}_k \textit{ so dass: } A = ((Y_k)_{k\geq n})^{-1}(B_n)$$

Wir schreiben $\mathcal{A}((Y_k)_{k\in\mathbb{N}})$ für das System der asymptotischen Ereignisse A bzgl. $(Y_k)_{k\in\mathbb{N}}$.

Bemerkung

1. $\mathcal{A}((Y_k)_{k\in\mathbb{N}})$ ist eine σ-Algebra, und
2. $A \in \mathcal{A}((Y_k)_{k\in\mathbb{N}})$ gilt g.d.w. $A \in \sigma(Y_k : k \geq n)$ für alle n (zeigen!).

Beispiel

a) $A = \bigcap_{k\in\mathbb{N}} \bigcup_{l\geq k}\{Y_l \in A_l\}$ für $A_l \in \mathcal{F}_l\ \forall l \in \mathbb{N}$ ist asymptotisch (Übung)

b) $A = \left\{ \lim\limits_{N\to\infty} \left(\frac{1}{N} \sum\limits_{i=1}^{N} Y_i \right) \text{ existiert und nimmt Werte im Intervall } [a,b] \text{ an} \right\}$ ist asymptotisch.

Satz 4.40 (0-1 Gesetz von Kolmogorov)
Sei $(Y_k)_{k\in\mathbb{N}}$ Folge von unabhängigen Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$. Dann hat jedes asymptotische Ereignis $A \in \mathcal{A}((Y_k)_{k\in\mathbb{N}})$ die Wahrscheinlichkeit 1 oder 0.

Satz 4.41 (Borel-Cantelli-Lemma)
In $(\Omega, \mathcal{F}, P)$ sei $(A_k)_{k\in\mathbb{N}}$ Folge von Ereignissen mit $A := \limsup_{k\to\infty} A_k \equiv \bigcap\limits_{n\in\mathbb{N}} \bigcup\limits_{k\geq n} A_k$ dann gilt:

(i)

$$\sum_{k\in\mathbb{N}} P(A_k) < \infty \implies P(A) = 0$$

(ii) sind die $A_k, k \in \mathbb{N}$ zudem unabhängige Ereignisse mit

$$\sum_{k\in\mathbb{N}} P(A_k) = \infty \implies P(A) = 1$$

5 Erwartungswert und Varianz

Kenngrößen reellwertiger Zufallsvariablen.

Definition 5.42
*X Zufallsvariable auf $(\Omega, \mathcal{F}, P)$ mit Werten in $(\mathbb{R}, \mathcal{B})$. Ist $X \geq 0$ oder $X \in L^1(P)$ (d.h. $\int_\Omega |X|^1 dP < \infty$) so heißt $E[X] = \int X dP$ der **Erwartungswert** von X.*

Bemerkung Für $\mathbb{R}^n$-wertige Zufallsvariablen X wird $E[X] = (E[X^i])_{i=1\ldots n}$ koordinatenweise definiert; für komplexwertige Zufallsvariablen entsprechend für Real- und Imaginärteil.

Lemma 5.43
Sei $P_X = P \circ X^{-1}$ Verteilung von X für

$$X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$$
$$f : (\Omega', \mathcal{F}') \to (\mathbb{R}, \mathcal{B})$$

mit $f \geq 0$ oder $f \circ X \in L^1(P)$ messbar.
Dann gilt:

$$E[f(x)] = \int_{\Omega} f(X(w))P(d\omega) = \int_{\Omega'} f(x)P_X(\mathrm{d}x)$$

Korollar 5.44
Sei X Zufallsvariable mit absolutstetiger Verteilung mit Dichte ρ, f messbare reelle Funktion, so dass $Y := f \circ X \geq 0$ oder in $L^1(P)$ ist, dann gilt:

$$E[f(X)] = \int_{\Omega'} f(x)\rho(x)\mathrm{d}x$$

Satz 5.45 (wichtige Ungleichungen)
Sei X Zufallsvariable auf $(\Omega, \mathcal{F}, P)$. Dann gilt:

a) ***Markovsche Ungleichung***

$$P[|X| \geq \epsilon] \leq \frac{E[|X|^p]}{\epsilon^p}, \quad p \in [1, \infty), \epsilon > 0$$

*b) Spezialfall von (a) mit $p = 2$ nennt man die **Tschebyscheff-Ungleichung** (auch Chebysev-U.): Bei Anwendung auf zentrierte Zufallsvariablen $X := Y - E[Y]$ mit $Y \in L^1(P)$ liefert das mit [10] $V(Y) := E[(Y - E[Y])^2]$*

$$P[|Y - E[Y]| \geq \epsilon] \leq \frac{V(Y)}{\epsilon^2}, \quad \epsilon > 0\,.$$

c) ***Exponentielle Markov-Ungleichung***

$$P[\alpha X \geq \epsilon] \leq \frac{E[\mathrm{e}^{\alpha X}]}{\mathrm{e}^{\epsilon}}, \quad \alpha, \epsilon > 0$$

d) ***Cauchy-Schwarzsche Ungleichung***
für $X, Y \in L^2(P)$ Zufallsvariablen gilt $X \cdot Y \in L^1(P)$ und

$$E[|XY|] \leq \sqrt{E[X^2]E[Y^2]}$$

e) ***Höldersche Ungleichung***
für $X \in L^p(P), Y \in L^q(P)$ mit $p \in (1, \infty)$ und q so dass: $\dfrac{1}{p} + \dfrac{1}{q} = 1$. Dann gilt:

$$E[|XY|] \leq E[|X|^p]^{\frac{1}{p}} E[|Y|^q]^{\frac{1}{q}}$$

f) ***Minkowski Ungleichung***
für $X, Y \in L^p(P), p \in [1, \infty)$, dann gilt:

$$\|X + Y\|_{L_p} \leq \|X\|_{L_p} + \|Y\|_{L_p}$$

[10] vgl. anschließende Definition der Varianz $V(Y)$.

Lemma 5.46 (Jensensche Ungleichung)
Seien X eine reelle Zufallsvariable, $f : \mathbb{R} \to \mathbb{R}$ eine konvexe Funktion und sei X in $\mathrm{L}^1(P)$. Dann gilt: $f(E[X]) \leq E(f(X))$

Satz 5.47
Seien $X, Y \in L^2(P)$ unabhängig. Dann gilt $E[XY] = (EX)(EY)$

Definition 5.48
Seien X, Y reellwertige ZVen.

(a) *$V(X) := E[(X - E[X])^2] = E[X^2] - (E[X])^2$ die **Varianz** von X, und $\sqrt{V(X)}$ heißt die **Standardabweichung** von X.*

(b) *Für $X, Y \in L^2(P)$ heißt $\mathrm{Cov}(X, Y) := E[(X - E[X])(Y - E[Y])] = E[XY] - (E[X])(E[Y])$ die **Kovarianz** von X und Y und ist in $\mathbb{R}$.*

(c) *Falls $\mathrm{Cov}(X, Y) = 0$ gilt, heißen X und Y **unkorreliert**.*

Lemma 5.49
$X, Y, X_1, X_2, \ldots \in L^2(P), \quad a, b, c, d \in \mathbb{R}$

(a) $\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y)$ *insbesondere* $V(aX + b) = a^2 V(X)$

(b) $\mathrm{Cov}(X, Y) \leq \sqrt{V(X)V(Y)}$ *wegen Schwarz'scher Ungleichung*

(c) $\sum_{k=1}^{n} X_k \in L^2$ *und* $V(\sum_{k=1}^{n} X_k) = \sum_{k=1}^{n} V(X_k) + \sum_{j \neq k} \mathrm{Cov}(X_j, X_k)$

Falls X_k unkorreliert sind gilt insbesondere $V(\sum_{k=1}^{n} X_k) = \sum_{k=1}^{n} V(X_K)$

(d) *Falls X, Y unabhängig sind, dann sind sie auch unkorreliert*

Beweis:Übung.

Definition 5.50

1. *Für $\mathbb{R}$-wertige Zufallsvariable $X \in L^2(P)$ mit $V(X) > 0$ heißt*

$$\tilde{X} := \frac{X - E[X]}{\sqrt{V(X)}}$$

standardisierte ZVe: $E[\tilde{X}] = 0$ und $V(X) = 1$.

2. *Für X $\mathbb{R}^n$-wertig, $X \in L^2(P)$, d.h. $X^i \in \mathrm{L}^2\ \forall i$. Dann ist*

$$\Big(\mathrm{Cov}(X^i, X^j)\Big)_{i,j}, \quad 1 \leq i, j \leq n$$

die Kovarianzmatrix von X.

Bemerkung
Unkorreliertheit von X, Y impliziert **<u>nicht</u>**, dass X und Y unabhängig sind. Gegenbeispiele.

Definition 5.51 (Korrelationskoeffizient)
Seien $X, Y \in L^2(P),\ V(X), V(Y) > 0$. Dann heißt

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{V(X)V(Y)}}$$

*die **Korrelation** von X mit Y. Notation: Häufig wird $\mathrm{Corr}(X, Y)$ bezeichnet mit $\rho(X, Y)$.*

Die Korrelation zweier reller Zufallsvariablen ist invariant unter (nicht-trivialen) positiv-affinen Transformationen.

Lemma 5.52
X, Y wie in Definition 5.51. Dann

a) $\mathrm{Corr}(X,Y) \in [-1,1]$

b) Sei X zentriert ($E[X]=0$), dann gilt

$$\min_{a,b\in\mathbb{R}} E[|Y-(aX+b)|^2] = E[|Y-(a^*X+b^*)|^2]$$

für $a^* = \frac{\mathrm{Cov}(X,Y)}{V(X)} = \mathrm{Corr}(X,Y)\sqrt{\frac{V(Y)}{V(X)}}$, $b^* = EY$ *und* $\min\limits_{a,b\in\mathbb{R}} E[...] = V(Y)(1-(\mathrm{Corr}(X,Y))^2)$

6 Die Gesetze der großen Zahlen

(starkes und schwaches Gesetz)

Bemerkung
Klassische Formulierung der Tschebyscheff (Chebyshev) Ungleichung. Für $Y \in L^2(P), \varepsilon > 0$

$$P[|Y-EY| > \varepsilon] \leq \frac{V(Y)}{\varepsilon^2}$$

(Satz 5.45 anwenden für $Y - EY \checkmark$)

Definition 6.52
Seien $Y, (Y_n)_{n\in\mathbb{N}}$ Zufallsvariable mit Werten in $\mathbb{R}$ (oder $\mathbb{R}^d$) auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$.

*$(Y_n)_{n\in\mathbb{N}}$ **konvergiert stochastisch** (auch „in Wahrscheinlichkeit" oder „in P") gegen Y, falls*

$$P[|Y_n - Y| \geq \varepsilon] \overset{n\to\infty}{\longrightarrow} 0 \quad \forall \varepsilon > 0$$

$$\text{(oder, äquivalent: } P\left[|Y_n - Y| \leq \varepsilon\right] \overset{n\to\infty}{\longrightarrow} 1 \quad \forall \varepsilon > 0)$$

Stochastische Konvergenz wird analog für Zufallsvariablen definiert, die Werte in normierten (oder allgemeiner in metrischen) Räumen annehmen.

Satz 6.53 (Schwaches Gesetz der großen Zahlen)
Seien $(X_n)_{n\in\mathbb{N}}$ unkorrelierte Zufallsvariablen in $L^2(P)$ mit $\sup\limits_{n\in\mathbb{N}} V(X_n) \leq c < \infty$ ($c \in \mathbb{R}$). Dann gilt für alle $\varepsilon > 0$:

$$P\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - E[X_i])\right| \geq \varepsilon\right] \leq \frac{\sup_{n\in\mathbb{N}} V(X_n)}{n\varepsilon^2} \overset{n\to\infty}{\longrightarrow} 0, \qquad (*)$$

insbesondere konvergiert $\frac{1}{n}\sum_{i=1}^{n}(X_i - E[X_i])$ stochastisch gegen 0.

Definition 6.54 (fast-sichere Konvergenz)
Seien Y, $(Y_n)_{n\in\mathbb{N}}$ reellwertige Zufallsvariablen auf demselben $(\Omega, \mathcal{F}, P)$. Dann ***konvergiert*** *Y_n* ***P-fast-sicher (P-fs.)*** *gegen Y falls die Folge außerhalb einer Nullmenge konvergiert. Das heißt*

$$P(\lim_{n\to\infty} Y_n = Y) = 1$$

Man sagt auch, die Folge Y_n konvergiert ***P-fast-überall*** *(P-f.ü.).*

Bemerkung
Für ein positives Maß $\mu : \mathcal{F} \to [0, \infty]$, das kein Wahrscheinlichkeitsmaß zu sein braucht, sagt man allgemeiner, dass eine durch eine Teilmenge $A \in \Omega$ beschriebene Eigenschaft $\mu-$fast überall gilt, falls es eine μ-Nullmenge gibt mit $A^c \subset N$, $\mu(N) = 0$.

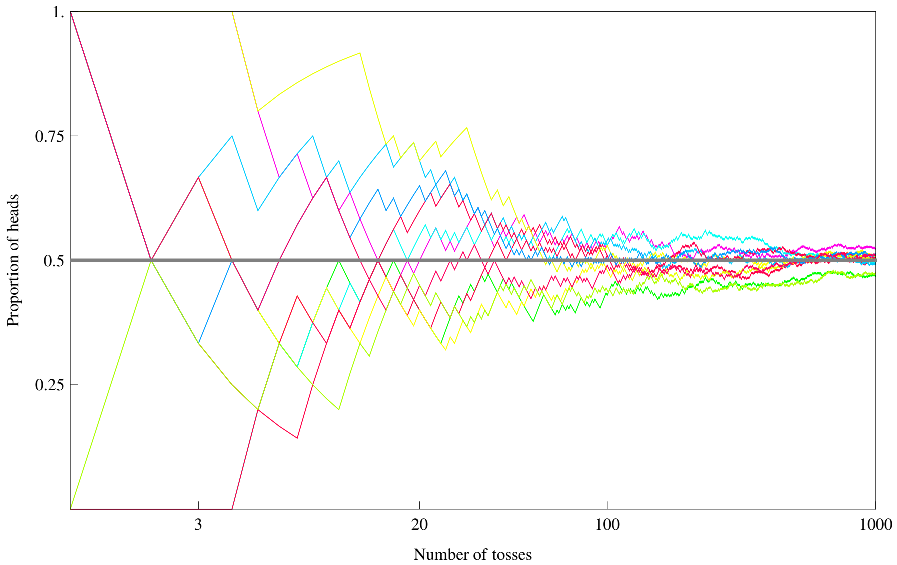
Lemma 6.55
Konvergieren $Y_n, n \in \mathbb{N}$, P-fast-sicher gegen Y, so gilt auch $Y_n \xrightarrow{P} Y$

Bemerkung
Umkehrung gilt nicht: Stochastische Konvergenz impliziert nicht die P-fast-sichere Konvergenz!

Satz 6.56 (Starkes Gesetz der großen Zahlen)
$(X_n)_{n\in\mathbb{N}}$ unkorrelierte Zufallsvariablen, $X_n \in L^2(P)$, auf $(\Omega, \mathcal{F}, P)$, $\sup_{n\in\mathbb{N}} V(X_n) =: c < \infty$. Dann

$$\frac{1}{n}\sum_{k=1}^{n}(X_K - E[X_k]) \to 0 \quad \textit{P-fast sicher}$$

Beispiel (zur Konvergenz der relativen Häufigkeit von Kopf bei Münzwürfen gegen $\frac{1}{2}$)



Bemerkung
Das starke Gesetz der großen Zahlen gilt auch unter schwächeren Voraussetzungen als in dem vorigen Theorem

angegeben. 1981 hat Etemadi gezeigt:
Sind $X_1, X_2, \ldots$ in $L^1(P)$, unkorreliert und identisch verteilt. Dann gilt:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to E[X_1] \qquad \text{(P-fast sicher)}$$

(vgl. Klenke, Thm.5.17, oder Etemadi 1981 `https://doi.org/10.1007/BF01013465`)

Definition 6.57
Seien $X_1, X_2, \ldots$ unabhängig identisch verteilte ("i.i.d." in engl. Abkürzung) Zufallsvariablen in $L^2(P)$ Für jede Realisierung $\omega \in \Omega$ von $X_1(\omega), X_2(\omega), \ldots$ sei

$$x \mapsto F_n(x) := F_n(x,\omega) := \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{(-\infty,x]}(X_k(\omega)), \quad x \in \mathbb{R}.$$

*Wir nennen F_n die **empirische Verteilungsfunktion** von $X_1, \ldots X_n$ $(n \in \mathbb{N})$. F_n ist die Verteilungsfunktion des empirischen Wahrscheinlichkeitsmaßes*

$$P_n := \frac{1}{n}\sum_{k=1}^{n} \delta_{X_k}.$$

Als Funktion der zufälligen Stichprobe is F_n also eine zufällige Funktion $\omega \mapsto F_n(\cdot,\omega)$.

Beispiel
Dann gilt

$$Y_k := \mathbb{1}_{(-\infty,x]}(X_k)$$

sind i.i.d. Zufallsvariablen mit Verteilung Bernoulli$(F(x))$, wobei F Verteilungfunktion von X_k ist.

$$\overset{\text{starkes Gesetz d. gr. Zahlen}}{\Longrightarrow} \qquad \forall x \in \mathbb{R}: \quad F_n(x) = \frac{1}{n}\sum_{k=1}^{n} Y_k \overset{n\to\infty}{\longrightarrow} F(x) \quad P\text{-fast-sicher}$$

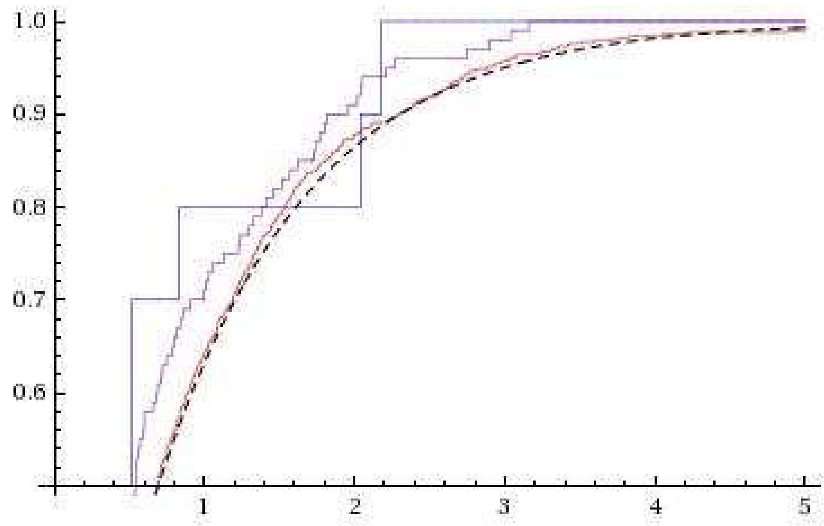
d.h. "die empirischen Verteilungsfunktionen konvergieren punktweise fast-sicher gegen die Verteilungsfunktion F aus der die i.i.d. Ziehungen kommen."

Dies motiviert die stärkere Aussage des "Hauptsatzes der Statistik", welcher sogar f.s. gleichmäßige Konvergenz zeigt:

Satz 6.58 (Glivenko-Cantelli)
Seien $(X_k)_{k\in\mathbb{N}}$ $\mathbb{R}$-wertige Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$ unabhängig und identisch verteilt ("i.i.d."), bezeichne F_n die empirische Verteilungsfunktion der $X_1, \ldots, X_n$ und sei F die Verteilungsfunktion von X_n. Dann gilt:

$$\limsup_{n\to\infty}\left(\sup_{x\in\mathbb{R}} |F_n(x) - F(x)|\right) \equiv \limsup_{n\to\infty} \|F_n - F\|_\infty = 0 \quad \textit{P-fast sicher}$$

Beispiel
Die folgende Grafik zeigt (in einem Ausschnitt) die empirische Verteilungsfunktionen einer Realisierung von 10, 100, bzw. schließlich 1000 i.i.d. Exp(1)-verteilten Zufallsvariablen, und die "theoretische" Verteilungsfunktion der Exp(1)-Verteilung (gestrichelt in Schwarz).



7 Charakteristische Funktionen

Grundidee: Wir charakterisieren Wahrscheinlichkeitsmaße auf $(\mathbb{R}^d, \mathcal{B}^d(\mathbb{R}) = \mathcal{B}(\mathbb{R}^d))$ durch komplexwertige Funktionen und können damit nützliche Aussagen über Maße auf $\mathbb{R}^d$ mittels charakteristischen Funktionen formulieren und "nachrechnen".

komplexe Zahlen: $\mathbb{C} \widehat{=} \mathbb{R}^2$

$z = x + iy, \quad i^2 = -1$

Kartesische Koordinaten und Polarkoordinaten:

$$
\begin{aligned}
&z = x + iy = r \cdot \exp(i\theta), \quad \mathrm{Re}(z) = x, \quad \mathrm{Re}(z) = y,\\
&(r,\theta) \in (0,\infty) \times [0, 2\pi), \quad (x,y) \in \mathbb{R}^2, \quad \mathbb{R}^2 \backslash \{0\} \ni (x,y) \longleftrightarrow (r,\theta) \text{ für } z \neq 0\\
&\text{Exponentialfunktion: } \exp(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!},\\
&r \cdot \exp(i\theta) = r(\cos\theta + i\sin\theta); \quad r \geq 0,\ \theta \in \mathbb{R},\\
&|z| = (x^2 + y^2)^{1/2} = r,
\end{aligned}
$$

Notation (Skalarprodukt in $\mathbb{R}^d$)

$\langle x, y\rangle = \sum_{k=1}^{d} x_i y_i, \qquad x^T y = x \cdot y = xy$

Definition 7.59

a) sei μ ein Wahrscheinlichkeitsmaß aus $(\mathbb{R}^d, \mathcal{B}^d(\mathbb{R}))$. Dann heißt $\hat{\mu}$ definiert durch:

$$
\begin{aligned}
&\hat{\mu} : \mathbb{R}^d \to \mathbb{C},\\
&\hat{\mu}(u) := \int_{\mathbb{R}^d} \mathrm{e}^{i\langle u,x\rangle} \mu(\mathrm{d}x)
\end{aligned}
$$

*die **charakteristische Transformierte** (oder Fouriertransformierte) von μ*

b) Für eine Zufallsvariable X, mit Werten in $(\mathbb{R}^d, \mathcal{B}^d(\mathbb{R}))$ heißt

$$\varphi_X(u) := \widehat{P}_X(u) = \int_{\mathbb{R}^d} \mathrm{e}^{i\langle u,x\rangle} P_X(\mathrm{d}x) = E[\mathrm{e}^{i\langle u,X\rangle}]$$

*die **charakteristische Funktion** von X.*

Bemerkung

- für f $\mathbb{C}$-wertig wirddefiniert $\int f d\mu = \int \mathrm{Re}(f) d\mu + i \int \mathrm{Im}(f) d\mu$
  und somit übertragen sich Resultate für Maßintegrale $\mathbb{R}$-wertiger Integranden auf $\mathbb{C}$-wertige Integranden.
- man kann nachprüfen, dass $|\int f d\mu| \leq \int |f| d\mu$

Bemerkung

$$\hat{\mu}(u) = \int e^{i\langle u,x\rangle}\mu(\mathrm{d}x) = \int_{\mathbb{R}^d} \cos(\langle u,x\rangle)\mu(\mathrm{d}x) + i\int_{\mathbb{R}^d} \sin(\langle u,x\rangle)\mu(\mathrm{d}x) \quad \in \mathbb{C},\ \forall u \in \mathbb{R}^d$$

$$\varphi_X(u) = E[\cos(\langle u,X\rangle)] + iE[\sin(\langle u,X\rangle)] \quad \in \mathbb{C},\ \forall u \in \mathbb{R}^d$$

Lemma 7.60
Sei μ Wahrscheinlichkeitsmaß auf $(\mathbb{R}^d, \mathcal{B}^d(\mathbb{R}^d))$. Dann ist $\hat{\mu}$ eine beschränkte, stetige Funktion auf $\mathbb{R}^d$ mit $\hat{\mu}(0) = 1$

Bemerkung (m-tes Moment)
Sei $X = (X_1, \dots, X_d)$ eine $\mathbb{R}^d$-wertige Zufallsvariable mit $|X|^m = (\sum_{j=1}^d |X_j|^2)^{m/2} \in L^1(P)$ (d.h. $X \in L^m(P)$) mit $m \in \mathbb{N}$.
Dann nennen wir $E[X_{j_1} \cdots X_{j_m}] \in \mathbb{R}$ mit $j_k \in \{1, \dots d\}$ ein (gemischtes) unzentriertes **m-tes Moment** von X.
Die zentrierten Moemente von X sind diejenigen von $X' := X - EX \in L^m(P)$. Für eine univariate Zufallsvariable Y in $L^m(P)$ (zB. $Y := X_k$) ist $E[Y^m] \in R$ das unzentrierte m-te Moment und $E[(Y - E[Y])^m] \in \mathbb{R}$ das zentrierte m-te Moment.

Satz 7.61 (Beziehung zwischen Moment und Ableitung der charakteristischen Funktion)
X eine $\mathbb{R}^d$-wertige Zufallsvariable, mit $E[|X|^m] < \infty$ für ein $m \in \mathbb{N}$. Dann ist die charakteristische Funktion φ_x von X stetig-partiell differenzierbar bis zur m-ten Ordnung und es gilt mit $j_k \in \{1, \dots, d\}$, $k \leq m$, dass

$$\frac{\partial^m}{\partial x_{j_1} \partial x_{j_2} \dots \partial x_{j_m}} \varphi_X(u) = i^m E[X_{j_1} X_{j_2} \cdots X_{j_m} e^{i\langle u,X\rangle}]\,.$$

Insbesondere folgt mit $u = 0$ dass $\frac{\partial^m}{\partial x_{j_1} \partial x_{j_2} \dots \partial x_{j_m}} \varphi_X(0) = i^m E[X_{j_1} X_{j_2} \cdots X_{j_m}]$, und bei Wahl gleicher j_k ($k = 1, \dots, m$):

$$\frac{\partial^m}{\partial^m x_{j_k}} \varphi_X(0) = i^m E[X_{j_k}^m]$$

Beispiel

a) $X \sim \mathrm{Bernoulli}(p)$:
$\varphi_X(u) = E[e^{iuX}] = (1-p)e^0 + pe^{iu} = 1 + p\left(e^{iu} - 1\right)$

b) $X \sim \mathrm{Bin}(n, p)$:
Es gilt: $X = \sum\limits_{k=1}^{n} X_k$, für $X_1, \dots, X_n$ unabhängig $\mathrm{Bernoulli}(p)$ (Gleichheit in Verteilung)
$\varphi_X(u) = E[e^{iuX}] = E[e^{iu\sum_{k=1}^n X_k}] = \prod\limits_{k=1}^{n} E[e^{iuX_k}] = (pe^{iu} + 1 - p)^n$

c) $X \sim \text{Poiss}(\lambda)$:
$\varphi_X(u) = \sum\limits_{k=0}^{\infty} \mathrm{e}^{iuk} \frac{\lambda^k}{k!} e^{-\lambda} = (\text{Übung}) = \mathrm{e}^{\lambda(\mathrm{e}^{iu}-1)}$

d) $X \sim U([-a,a])$ gleichverteilt auf $[-a,a]$:
$\varphi_X(u) = (\text{Übung}) = \frac{\sin(au)}{au}$

e) $X \sim \mathcal{N}(0,1)$:
$\varphi_X(u) = (\text{Übung}) = \exp(\frac{-u^2}{2})$

Lemma 7.62 (charakteristische Funktionen von affinen Transformationen)
X eine $\mathbb{R}^d-$wertige Zufallsvariable, $Y := AX + b$, A eine $(m \times d)-$ Matrix, $b \in \mathbb{R}^m$, dann gilt:

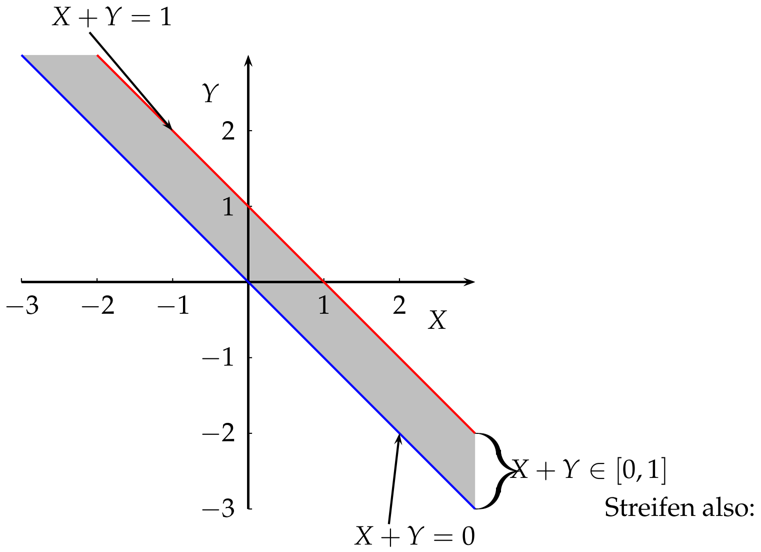
$$\varphi_Y(u) = \mathrm{e}^{i\langle u,b\rangle} \varphi_X(A^T u), u \in \mathbb{R}^m$$

Beispiel
X univariat normalverteilt: $X \sim \mathcal{N}(\mu, \sigma^2)$
$\mu \in \mathbb{R}, \quad \sigma^2 > 0, \quad X = \mu + \sigma Y \quad$ für $Y := \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$
$\Rightarrow \varphi_X(u) = \exp(iu\mu - \sigma^2 \frac{u^2}{2})$

Summen unabhängiger Zufallsvariablen
Wir werden sehen, dass sich hiervon die charakteristischen Funktionen sehr einfach berechnen lassen, und zudem den Begriff der *Faltung* einführen.

Definition 7.63 (Definition und Satz)
Seien X, Y $\mathbb{R}^d$-wertige unabhängige Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$ mit Verteilungen $P_X = P \circ X^{-1}$, $P_Y = P \circ Y^{-1}$ auf $\mathcal{B}(\mathbb{R}^d)$. Dann heißt die Verteilung von $Z := X + Y$ die ***Faltung*** *(Faltungsprodukt) von P_X und P_Y, notiert $P_Z = P_X * P_Y$, und ist gegeben durch:*

$$P_Z(A) = P_X * P_Y(A) := \int\limits_{\mathbb{R}^d} \int\limits_{\mathbb{R}^d} \mathbb{1}_A(x+y) P_X(\mathrm{d}x) P_Y(\mathrm{d}y), \quad \forall A \in \mathcal{B}(\mathbb{R}^d)$$



Korollar 7.64
X, Y unabhängige Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$, $\mathbb{R}^d$-wertig, $Z := X + Y$. Dann gilt:

$$\varphi_Z(u) = \varphi_X(u) \cdot \varphi_Y(u), \quad u \in \mathbb{R}^d$$

Satz 7.65
Seien X, Y unabhängige, $\mathbb{R}^d$-wertige Zufallsvariablen und $Z := X + Y$;

a) Hat zudem X eine Dichte f_X, dann hat Z eine Dichte f_Z und es gilt

$$f_Z(z) = \int f_X(z - y) P_Y(\mathrm{d}y), \quad z \in \mathbb{R}^d.$$

b) Haben sowohl X als auch Y eine Dichte f_X bzw. f_Y, dann hat Z die Dichte

$$f_Z(z) = \int f_X(z - y) f_Y(y) \mathrm{d}y = \int f_Y(z - x) f_X(x) \mathrm{d}x$$

Warum heißen charakteristische Funktionen "charakteristisch"?

Die Antwort gibt:

Satz 7.66
Sei X Zufallsvariable mit Werten in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Dann charakterisiert $\varphi_X(\cdot)$ bzw. $\widehat{P}_X(\cdot)$ die Verteilung $P_X = P \circ X^{-1}$ von X auf $\mathcal{B}(\mathbb{R}^d)$ in folgendem Sinne: Für Wahrscheinlichkeitsmaße μ_1, μ_2 auf $\mathcal{B}^d(\mathbb{R})$ gilt:

$$\widehat{\mu}_1 = \widehat{\mu}_2 \text{ genau dann, wenn } \mu_1 = \mu_2$$

Zum Beweis von "$\mu_1 = \mu_2$" wird gebraucht

$$\begin{aligned}
\int f(\sigma, u - v) \mu_1(\mathrm{d}u) &= \int \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} \left(\int f(\sigma, x) \mathrm{e}^{i\langle (u-v)/\sigma^2, x\rangle} \mathrm{d}x \right) \mu_1(\mathrm{d}u) \\
&\overset{\text{Fubini}}{=} \int \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} f(\sigma, x) e^{-i\langle v, x\rangle/\sigma^2} \underbrace{\left(\int \mathrm{e}^{i\langle u, x/\sigma^2\rangle} \mu_1(\mathrm{d}u) \right)}_{\hat{\mu}_1(x/\sigma^2)} \mathrm{d}x \\
&= \int \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} f(\sigma, x)\, e^{-i\langle v, x\rangle/\sigma^2} \hat{\mu}_1(x/\sigma^2) \mathrm{d}x
\end{aligned}$$

Bemerkung
will man etwas Konstruktives zur *Berechnung* von μ aus $\hat{\mu}$ sagen, braucht man Resultate zur *Inversion* der char. Funktion (bzw. der Fouriertransformierten).
Im Fall $d = 1$, gilt z.B. für eine $\mathbb{R}^1$- wertige Zufallsvariable X und $\mu := P_X$, $\hat{\mu} = \widehat{P}_X \equiv \varphi_X$, dass:

$$\begin{aligned} g(a,b) : &= \mu((a,b)) + \frac{1}{2}\mu(\{a,b\}) \\ &= \lim_{T \nearrow \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{\exp(-iua) - \exp(-iub)}{iu} \underbrace{\varphi_X(u)}_{\hat{\mu}(u)} \mathrm{d}u \qquad \forall a,b \in \mathbb{R}, a < b \end{aligned}$$

[Vgl. Shiryaev, Probability/Wahrscheinlichkeit, II.§12.Thm.3].
Und aus g lässt sich dann leicht die Verteilungsfunktion $F(b) = P(X \leq b)$ berechnen

Satz 7.67
Sei $X = (X^1, \ldots, X^d)$ eine $\mathbb{R}^d$-wertige Zufallsvariable auf $(\Omega, \mathcal{F}, P)$.
Dann sind $X^1, \ldots, X^d$ unabhängig genau dann, wenn

$$\varphi_X(u) = \prod_{k=1}^{d} \varphi_{X^k}(u_k) \quad \forall u \in \mathbb{R}^d$$

Bemerkung (zu charakteristischen Funtionen.
X sei $\mathbb{R}$-wertige Zufallsvariable. Definiere $\varphi(z) := E[\exp(< z, X >)]$ für $z \in \mathbb{C}$, wo Erwartungswert (endlich) definiert ist.
Man kann zeigen, dass diese Funktion auf einem geeigneten Gebiet komplex-analytisch (d.h. holomorph) ist. [Vgl. Strasser, Mathematical Theory of Statistics, I §85]

8 Mehrdimensionale Normalverteilungen

- werden auch *multivariate* Normalverteilungen (oder auch: Gaußverteilungen) genannt; Im eindimensionalen Fall spricht man dagegen von *univariaten* Verteilungen.
- univariate Normalverteilung:
  $X \sim \mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}^1, \sigma > 0,$
  ist im Falle $\sigma > 0$ absolutstetig mit Dichte $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, und hat charakteristischer Funktion $\varphi_X(u) = \exp(iu\mu - \frac{1}{2}\sigma^2 u^2), \quad u \in \mathbb{R}^1, \quad E[X] = \mu$ und $V[X] = \sigma^2$
- wir bezeichnen zudem mit $X \sim \mathcal{N}(\mu, 0)$ die Dirac'sche-Punktverteilung auf $\mu \in \mathbb{R}^1$, welche wir ebenfalls als (degenerierte, univariate) Normalverteilung auffassen, das heißt $P_X = \delta_\mu$ mit $\delta_\mu(A) = \begin{cases} 1 & \mu \in A \in \mathcal{B}(\mathbb{R}^1) \\ 0 & \text{sonst} \end{cases}$

Definition 8.68
*$X = (X^1, \ldots, X^d)$ eine $\mathbb{R}^d$-wertige Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$ heißt **Gauß'sche Zufallsvariable** (oder multivariat normalverteilt), falls jede Linearkombination $\langle a, X \rangle = \sum_{k=1}^{d} a_k X_k$ mit $a \in \mathbb{R}^d$ univariat normalverteilt ist.*

Bemerkung
letzteres ggf auch degeneriert normalverteilt mit $V[Y] = 0$ für $Y := \langle a, X\rangle$, d.h. $Y \sim \mathcal{N}(\mu, 0)$ also Dirac'sches-Punktmaß δ_μ in $\mu \in \mathbb{R}^1$

Satz 8.69
X eine $\mathbb{R}^d$-wertige Zufallsvariable auf $(\Omega, \mathcal{F}, P)$.
Dann ist X multivariat normalverteilt (Gauß'sch) genau dann, wenn die charakteristische Funktion von X die Form $\varphi_X(u) = \exp(i\langle u, \mu\rangle - \frac{1}{2}\langle u, Qu\rangle)$, $u \in \mathbb{R}^d$, hat, mit $\mu \in \mathbb{R}^d$ und Q eine $d \times d$ Matrix symmetrisch, positiv semidefinit (d.h $\langle u, Qu\rangle = u^t Qu \geq 0 \ \forall u \in \mathbb{R}^d$).
Und es gilt dann, dass Q die Kovarianzmatrix $\mathrm{Cov}(X)$ ist und μ der Mittelwertvektor $E[X]$.

Beispiel
Seien $X_1, \dots, X_d$ unabhängige $\mathbb{R}^1$-wertige Zufallsvariablen
$X_j \sim \mathcal{N}(\mu_j, \sigma_j^2), \quad \mu_j \in \mathbb{R}, \ \sigma_j \geq 0$
Dann ist $X = (X_1, \dots, X_d)$ multivariat normalverteilt, denn:

$$\begin{aligned} \varphi_X(u) &= \prod_{k=1}^d \varphi_{X_k}(u_k) = \prod_{k=1}^d \exp(iu_k\mu_k - \frac{1}{2}\sigma_k^2 u_k^2) \\ &= \exp(i\langle u, \mu\rangle - \frac{1}{2}\langle u, Qu\rangle) \end{aligned}$$

$$\text{mit } \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} \qquad \text{und} \quad Q = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \end{pmatrix}$$

Satz 8.70
Sei X $\mathbb{R}^d$-wertig, multivariat normalverteilt.
Dann sind die Koordinaten $X_1, \dots, X_d$ von X unabhängig genau dann, wenn die Kovarianzmatrix $\mathrm{Cov}(X) = Q$ von X Diagonalform hat.

Lemma 8.71
Sei X $\mathbb{R}^d$-wertige multivariate normalverteilte Zufallsvariable, $X \sim \mathcal{N}(\overset{\overset{E[X]}{\downarrow}}{\mu}, \overset{\overset{\mathrm{Cov}(X)}{\downarrow}}{Q})$. Dann existieren unabhängige, univariate $Y_1, \dots, Y_d$, wobei $Y_j \sim \mathcal{N}(0, \lambda_j^2)$ mit $\lambda_j \geq 0$, so dass $X = \mu + AY$ für $\mu = E[X]$ und eine orthogonale Matrix A ($d \times d$ Matrix, $AA^T = I_d$).

Bemerkung
Zur Simulation von $X \sim \mathcal{N}(\mu, Q)$ braucht man nur soviele Y_j zu simulieren, wie der Rang von Q (bzw. Γ) ist.
Genauer: Mit $m := \mathrm{rank}(Q)$ gilt: $\exists$ unabhängige $Y_1, \dots Y_m$ mit $Y_j \sim \mathcal{N}(0, 1)$ und eine $d \times m$-Matrix B, so dass $X = \mu + BY$ die gewünschte $\mathcal{N}(\mu, Q)$ Verteilung hat. [Übung]

Bemerkung Eine $\mathbb{R}^d$-wertige multivariant-normalverteilte ("Gauß'sche") Zufallsvariable X hat eine Dichte g.d.w. $\det(Q) \neq 0$ ist (d.h. falls Q nicht singulär).

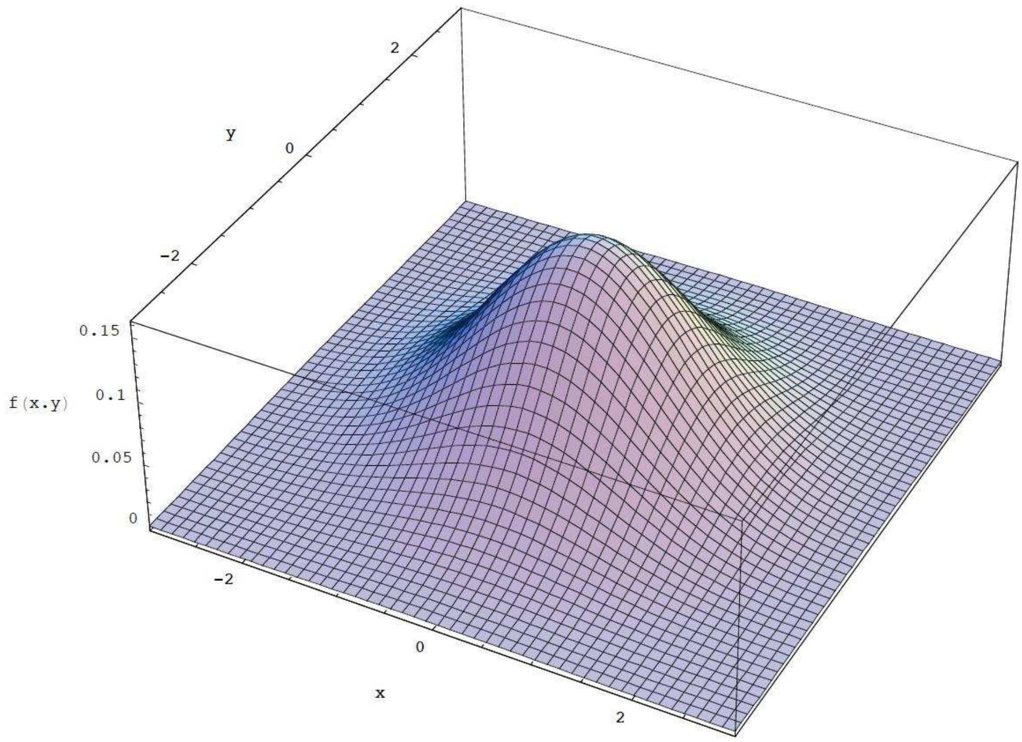
Mittels der Integral-Transformationsformel berechnet sich die Dichte dann als

$$f_X(x) = \frac{1}{(2\pi)^{\frac{d}{2}}} \frac{1}{\sqrt{\det Q}} \exp(-\frac{1}{2} < x - \mu, Q^{-1}(x - \mu) >)$$

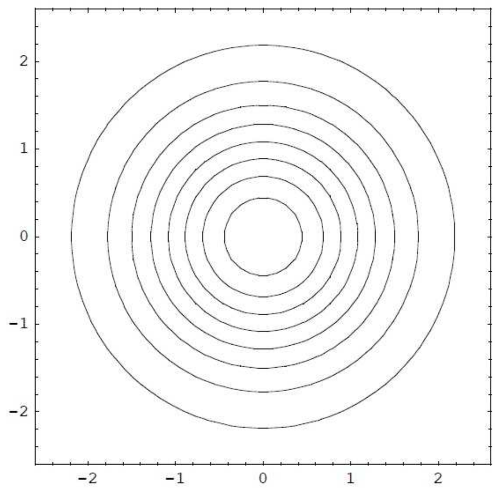
wobei $\det B^{-1} = (\det B)^{-1} = (\det Q)^{-1/2} > 0$, da Q hier pos.definit ist.

Beispiel (für bivariate Normalverteilungen)

1. Standard bivariate Normalverteilung $\mathcal{N}(0, I_2)$, d.h Mittelwerte $\mu_X = \mu_Y = 0$, Einzelstandardabweichungen $\sigma_X = \sigma_Y = 1$, Korrelation $\rho = 0$.

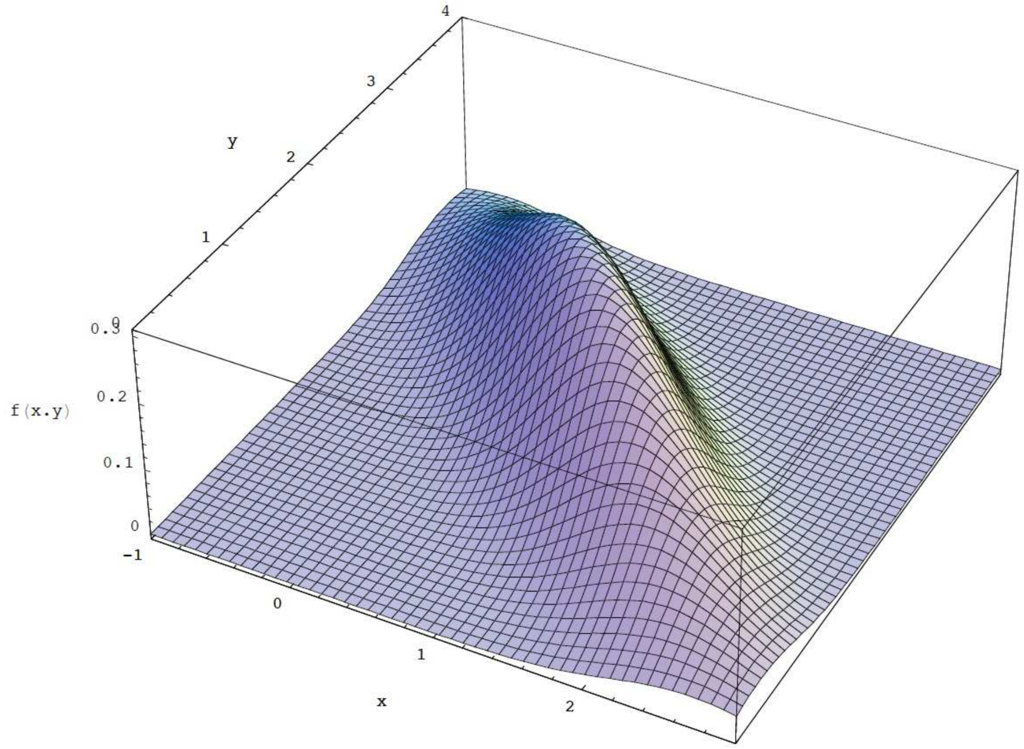


(Dichte der bivariaten Normalverteilung)

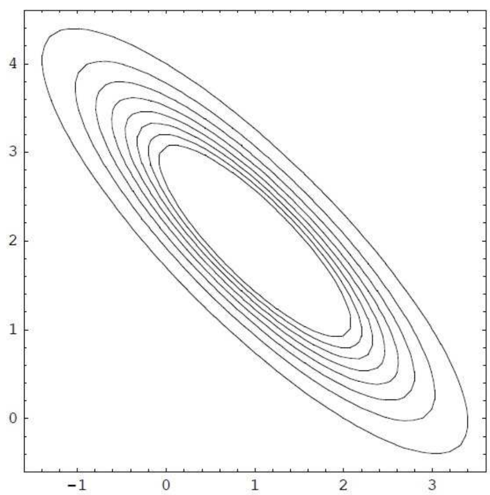


(Kontur- / Niveau- / Höhenlinien)

2. bivariate Normalverteilung mit $\mu_X = 1$, $\mu_Y = 2$, $\sigma_X = \sigma_Y = 1$ und negativer Korrelation $\rho = -0.85$



(Dichte der bivariaten Normalverteilung)



(Kontur- / Niveau- / Höhenlinien)

9 Konvergenz in Verteilung und Zentraler Grenzwertsatz

Ziel

Neuer Konvergenzbergriff für Zufallsvariablen, der zentral für den zentralen Grenzwertsatz sein wird.

Erinnerung

$X, X_n,\ n \in \mathbb{N},\ \mathbb{R}^d$-wertige Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$

a) Stochastische Konvergenz, $X_n \xrightarrow{P} X$: $[|X_n - X| \geq \varepsilon] \xrightarrow{n\to\infty} 0 \quad \forall \varepsilon > 0$

b) P-fast sichere Konvergenz, $X_n \longrightarrow X$ P-fast sicher $P[\lim\limits_{n\to\infty} X_n = X] = 1$

c) L^p-Konvergenz mit $p \in [1, \infty)$, $X_n \xrightarrow{L^p} X:\ ||X - X_n||_p = E[|X_n - X|^p]^{\frac{1}{p}} \xrightarrow{n\to\infty} 0$

Bemerkung
Für a), b), c) müssen die X, X_n auf gleichem Wahrscheinlichkeitsraum leben!

Definition 9.72
Sei (E, d) metrischer Raum, z.B. $E = \mathbb{R}^d$ mit $\mathrm{d}(x, y) = |x - y|$, mit Borelscher σ-Algebra $\mathcal{B}(E)$.

*a) Seien μ, μ_n Wahrscheinlichkeitsmaße auf $(E, \mathcal{B})$, dann **konvergiert** μ_n **schwach** gegen μ, falls*

$$\underbrace{\int f d\mu_n}_{(E_{\mu n}[f])} \longrightarrow \underbrace{\int f d\mu}_{(E_\mu[f])}$$

für alle $f \in \mathcal{C}_b$ (stetig und beschränkte Funktionen)

*b) Seien X, X_n E-wertige Zufallsvariablen auf Wahrscheinlichkeitsräumen $(\Omega, \mathcal{F}, P)$ bzw. $(\Omega_n, \mathcal{F}_n, P_n)$ $n \in \mathbb{N}$, dann **konvergiert** X_n in **Verteilung** gegen X, falls*

$$P_{X_n} = P_n \circ X_n^{-1}$$

schwach konvergieren gegen $P_X = P \circ X^{-1}$ d.h. falls

$$E_n[f(X_n)] \longrightarrow E[f(X)] \quad \forall f \in \mathcal{C}_b$$

Notation & Sprechweisen
$X_n \xrightarrow{\mathcal{D}} X$ "converges in distribution" (auch $X_n \xrightarrow{\mathcal{L}} X$ "in law")
$X_n \xrightarrow{\mathcal{D}} P_X$
$\mu_n \longrightarrow \mu$ (auch $\mu_n \xrightarrow{\mathrm{w}} \mu$ "weakly")

Für Zufallsvariablen speziell auf $E := \mathbb{R}^1$ kann Konvergenz in Verteilung durch eine geeignete Konvergenz der Verteilungsfunktionen $F_{X_n}(x) = P[X_n \leq x]$ beschrieben werden.

Satz 9.73
X, X_n $\mathbb{R}^1$-wertige Zufallsvariablen, mit Verteilungsfunktionen F_X bzw. F_{X_n}, $n \in \mathbb{N}$. Dann sind äquivalent:

a) $X_n \xrightarrow{\mathcal{D}} X$

b) $F_{X_n}(c) \longrightarrow F_X(c) \quad \forall c \in \mathbb{R}$, an welchen $F_X(\cdot)$ stetig ist.

Beziehung zw. den Konvergenzarten für Zufallsvariablen

Satz 9.74
Y_n, Y seien $\mathbb{R}^d$-wertige Zufallsvariablen, mit $Y_n \to Y$ P-fast-sicher, dann $Y_n \xrightarrow{\mathcal{D}} Y$

Satz 9.75
Y, Y_n $\mathbb{R}^d$-wertige Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$. Dann sind äquivalent:

(a) $Y_n \xrightarrow[\text{stoch. Konv.}]{P} Y$

(b) jede Teilfolge $(Y_{n_k})_{k \in N}$ hat eine Unterteilfolge $(Y_{n_{k_l}})_{l \in N}$, so dass $Y_{n_{k_l}} \xrightarrow{l \to \infty} Y$ P-fast-sicher

Satz 9.76 ((verallgemeinert) zu majorisierten Konvergenz mit stoch. Konvergenz statt P-fast-sicherer Konv. der Zufallsvar
seien Y_n, Y $\mathbb{R}^d$-wertige Zufallsvariablen aus $(\Omega, \mathcal{F}, P)$ mit $Y_n \xrightarrow{P} Y$, es gelte: $|Y_n| \leq Z$ ($\forall n \in \mathbb{N}$) für $Z \in L^1(P)$.
Dann gilt:

$$Y_n \xrightarrow{L^1} Y$$

und insbesondere also:

$$E[Y_n] \xrightarrow{n \to \infty} E[Y]$$

Satz 9.77
Y_n, Y seien $\mathbb{R}^d$-wertige Zufallsvariablen, mit $Y_n \xrightarrow{P} Y$, dann gilt $Y_n \xrightarrow[n \to \infty]{\mathcal{D}} Y$.

Bemerkung
<u>Die Umkehrung gilt nicht:</u>

$$Y_n \xrightarrow[n \to \infty]{\mathcal{D}} Y \not\Rightarrow Y_n \xrightarrow{P} Y$$

ein Beispiel dafür: Sei $X \sim \mathcal{N}(0,1)$ und $Y_n := (-1)^n X$ eine $\mathbb{R}^1$-wertige Zufallsvariable, dann:
$P_{Y_n} = P \circ X_n^{-1} = P_X = \mathcal{N}(0,1)$ also:
$Y_n \xrightarrow[n \to \infty]{\mathcal{D}} P_X$, aber Y_n konvergiert nicht stochastisch, denn

$$\begin{aligned} &Y_{2n} \xrightarrow{P} X, \\ &Y_{2n+1} \xrightarrow{P} -X \\ \Longrightarrow\ &Y_n \overset{P}{\nrightarrow} X, \text{ weil } P[X \neq -X] > 0 \end{aligned}$$

Bemerkung

1. Allgemein gilt (Übung):
   Falls $Y_n \xrightarrow{P} Y, \quad Y_n \xrightarrow{P} \bar{Y}$ folgt:
   $$Y = \bar{Y} \quad P-\text{fast-sicher}$$
2. Übung: Sei (X_n) eine Folge univariater (1-dim.) normalverteilter zentrierter Zufallsvariablen mit Varianzen $\sigma_n^2 \to 0$. Folgt daraus, dass eine Konvergenz $X_n \to 0$ in Wahrscheinlichkeit, in Verteilung, oder fast-sicher?

Proposition 9.78
Sei $(\mu_n)_{n\in\mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $F_n(x) = \mu_n((-\infty, x])$, $x \in \mathbb{R}$, Verteilungsfunktion von μ_n. Dann existiert eine Teilfolge $G_k = F_{n_k}$, $k \in \mathbb{N}$ und eine rechtsstetige, monoton wachsende Funktion $F: \mathbb{R} \to [0,1]$ mit $G_k(c) \underset{k\to\infty}{\to} F(c)$ für alle $c \in \mathbb{R}$, an denen $F(\cdot)$ stetig ist.

Bemerkung
F induziert via $\mu((a,b]) = F(b) - F(a)$ ein Maß auf $\mathcal{B}(\mathbb{R})$, welches aber im Allgemeinen kein Wahrscheinlichkeitsmaß zu sein braucht.

Definition 9.79
Sei $(\mu_i)_{i\in I}$ eine beliebige Menge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R})^d)$. Die Menge heißt ***gleichgradig straff*** *(kurz[11]: ẅstraff", falls gilt:$\forall \varepsilon > 0$ $\exists$ ein Kompaktum $K \subset \mathbb{R}^d$ (insbesondere also $K \in \mathcal{B}(\mathbb{R})^d$) so dass:*

$$\mu_i(K^c) < \varepsilon \quad \forall i \in I \tag{$*$}$$

Bemerkung
$(*) \iff$

$$\forall \varepsilon > 0\ \exists M < \infty \text{ so dass } \mu_i((-M, M]^c) < \varepsilon \quad \forall i \in I.$$

Diese Äquivalenz ist leicht zu zeigen (Übung) und liefert uns im $\mathbb{R}^1$ eine elementarere Bedingung, die wir im Weiteren nutzen. Die allg. Bedingung der Definition verallgemeinert sich natürlich auf topologische (z.B. metrische) Räume mit Borel-σ-Algebra.

Beispiele (für gleichgradig straffe Mengen von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$)

a) $|I| < \infty$ (endlich), dann ist $(\mu_i)_{i\in I}$ gleichgradig straff!

b) Oft ist $I = \mathbb{N}$. Eine Folge $(\mu_n)_{n\in\mathbb{N}}$ ist gleichgradig straff g.d.w.

$$\forall N \in \mathbb{N}: \quad (\mu_n)_{n\geq N} \text{ gleichgradig straff ist,}$$

oder (äquivalent)

$$\exists N \in \mathbb{N}: \quad (\mu_n)_{n\geq N} \text{ gleichgradig straff ist.}$$

Das folgt jeweils mit a) aus c).

c) Sind $(\mu_i)_{i\in I_1}$ und $(\mu_i)_{i\in I_2}$ gleichgradig straffe Mengen, dann ist auch die Vereinigung

$$(\mu_i)_{i\in I_1\cup I_2}$$

gleichgradig straff. Dies verallgemeinert sich leicht auf endliche Vereinigungen (wie?).

d) Falls $\mu_n \overset{w}{\to} \mu$ (schwache Konvergenz) gilt, so ist $(\mu_n)_{n\in\mathbb{N}}$ gleichgradig straff. [Übung: Zeigen für den Fall $d = 1$. Hinweis: Theorem 9.73 nutzen.]

[11] in Englisch: tight

Satz 9.80 (Helly'sches Selektionsprinzip)
Sei $(\mu_n)_{n\in\mathbb{N}}$ eine gleichgradig straffe Folge von Wahrscheinlichkeitsmaßen auf $\mathcal{B}(\mathbb{R})$. Dann existiert eine Teilfolge $(\mu_{n_k})_{k\in\mathbb{N}}$ und ein Wahrscheinlichkeitsmaß μ auf $\mathcal{B}(\mathbb{R})$ so dass

$$\mu_{n_k} \xrightarrow{w} \mu$$

Satz 9.81 (Stetigkeitssatz von Paul Lévy)
Seien $\mu, \mu_n, n \in \mathbb{N}$ Wahrscheinlichkeitsmaße auf $(\mathbb{R}^d, \mathcal{B}^d)$ mit charakteristischen Transfomierten $\varphi = \hat{\mu}$ bzw. $\varphi_n = \hat{\mu}_n$. Dann gilt:

(a) $\mu_n \xrightarrow{\omega} \mu \Rightarrow \varphi_n(u) \to \varphi(u), \quad \forall u \in \mathbb{R}^d$

(b) Falls $\varphi_n(u) \xrightarrow[n\to\infty]{} \psi(u), \;\; u \in \mathbb{R}^d$, für eine Funktion $\psi : \mathbb{R}^d \to \mathbb{C}$, die in 0 stetig ist, dann ist ψ die charakteristische Transformierte eines Wahrscheinlichkeitsmaßes ν auf $\mathcal{B}^d$ und $\mu_n \xrightarrow{\omega} \nu$

Bemerkung
In der Tat kann man in (a) sogar die stärkere Aussage zeigen: $\varphi_n \to \varphi$ gleichmäßig auf Kompakta.

Beispiel (zur Anwendung von Satz 9.81)
Betrachte $X_n \sim$Poisson(λ_n) mit $\lambda_n := n$, $Z_n := \dfrac{X_n - EX}{\sqrt{V(X_n)}} = \dfrac{X_n - n}{\sqrt{n}}$
dann gilt: $Z_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$

Beispiele (zur Motivation des zentralen Grenzwertsatzes:)

- Aus vorigem Beispiel:
$$\Rightarrow Z_n = \frac{X_n - n}{\sqrt{n}} := \frac{\sum\limits_{j=1}^{n} Y_j - nE[Y_1]}{\sqrt{nV(Y_1)}} \xrightarrow{n\to\infty} \mathcal{N}(0,1)$$
- Man kann andere Beispiele für (iid) Verteilungen der Summanden Y_j betrachten und durch direkte Berechnungen zum gleichen Grenzwert gelangen (Übung).

Bemerkung
Der zentrale Grenzwertsatz zeigt nun, dass eine entsprechende Aussage wie in den vorigen Beispielen allgemeiner für "beliebige" zu Grunde liegende Verteilungen der unabhängigen Summanden Y_j gilt, etwas genauer gilt, dass Summen vieler unabhängiger gleichgroßer Zufallsgrößen approximativ Gauß-verteilt sind; bei Standardisierung Standard-Gauß-verteilt, das rechtfertigt die Bezeichnung dieser generischen Limesverteilung als Normalverteilung.

Satz 9.82 (Zentraler Grenzwertsatz)
Seien $(X_n)_{n\in\mathbb{N}}$ i.i.d., $\mathbb{R}$-wertig, mit $\mu := E[X_n]$, $\sigma^2 := V(X_n) \in (0,\infty)$ und sei

$$S_n := \sum_{j=1}^{n} X_j, \quad S_n^* := \frac{S_n - n\cdot\mu}{\sqrt{n\cdot\sigma^2}}$$

Dann gilt:

$$S_n^* \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,1)$$

Bemerkung Sei X_n wie in Satz 9.82 (zentraler Grenzwertsatz).
Aus dem starken Gesetz der großen Zahlen wissen wir, dass

$$\bar{X}_n := \frac{1}{n}\sum_{j=1}^{n} X_j = S_n/n \to [n\to\infty]\mu = E[X_1] \qquad P\text{-fast-sicher}$$

Frage: Wie schnell konvergiert das?
Für Konvergenz einer Folge $(y_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$ sagt man, dass (y_n) mit Ordnung $n^{-\alpha}$, $\alpha > 0$, gegen 0 konvergiert, falls

$$\overline{\lim_{n\to\infty}}\, n^{\alpha}|y_n| =: c \in \mathbb{R}$$

gilt. Ein solches α, so dass $n^{\alpha}\,|\bar{X}_n/n - \mu|$ P-fast-sicher gegen ein $c \in \mathbb{R}$ konvergiert, gibt es nicht; Jedoch zeigt der zentrale Grenzwertsatz, dass Konvergenz in Verteilung gegen eine endliche (normalverteilte) Zufallsvariable vorliegt:

$$\underbrace{n^{\frac{1}{2}}\cdot(\bar{X}_n/n-\mu)}_{\frac{\sum_{j=1}^{n} X_j - n\mu}{\sqrt{n}} = \sigma S_n^*} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,\sigma^2)$$

In diesem Sinne gilt: Der ZGWS zeigt, dass die Konvergenzordnung $n^{-1/2}$ ist.

Bemerkung Es gibt Verallgemeinerungen des zentralen Grenzwertsatzes:

→ Für eine mehrdimensionale Version [s. Klenke, Thm.15.56].

→ Die Annahme identischer Verteilungen kann abgeschwächt werden; hilfreich für Anwendungen, wo dieses Kriterium teils nicht erfüllt ist; s. [Klenke] für sog. "Dreieckschemata" von ZV'en.

→ Sehr genaue Bedingungen für zentrale Grenzwertsatzaussage liefern die "Lindeberg-Bedingungen" [vgl. Klenke, Thm.15.43].

→ Der Satz von Berry-Esseem liefert eine Abschätzung dafür, wie schnell die Verteilungsfunktionen der S_n gegen diejenige der Standard-Normalverteilung konvergieren, [s. Klenke, Thm.15.51].

→ Beispiel für Anwendung des zentralen Grenzwertsatzes:
Bestimmung von (approximativen) "Vertrauensbereichen" (Konfidenzbereichen) für unbekannte Verteilungsparameter, die aus Daten geschätzt werden .

10 Grundlagen der mathematischen Statistik

zur Schätz- und Testtheorie.

Definition 10.84 (statistisches Modell)
[12] *Sei $\mathcal{X}$ eine beliebige Menge, $\mathcal{F}$ eine σ-Algebra auf $\mathcal{X}$, und sei $\{P_\vartheta \mid \vartheta \in \Theta\}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{F})$, wobei Θ eine beliebige Menge ist.*

*Dann heißt $(\mathcal{X}, \mathcal{F}, \{P_\vartheta\}_{\vartheta\in\Theta})$ **statistisches Modell**.*

Grundlagen der Schätztheorie

Definition 10.85
*Für eine Funktion[13] $g: \ \Theta \to \mathbb{R}^d$ heißt eine messbare Abbildung $T: \ \mathcal{X} \to g(\Theta) \equiv \mathrm{Im}(g)$ ein **Schätzer** für die Kenngrösse $g(\vartheta)$.*

zur Definition von ML-Schätzer:
Sei μ ein Maß auf $(\mathcal{X}; \mathcal{F})$ derart, dass für jedes $\vartheta \in \Theta$ die Radon-Nikodym Dichten existieren:

$$\frac{\mathrm{d}P_\vartheta}{\mathrm{d}\mu} =: \rho_\vartheta, \quad \text{das heißt } \underbrace{P_\vartheta(A) = \int \rho_\vartheta(x)\mathbb{1}_A(x)\mu(\mathrm{d}x)}_{\text{d.h. } \int \mathbb{1}_A \rho_\vartheta \mathrm{d}\mu = \int \mathbb{1}_A \mathrm{d}P_\vartheta}, \forall\, A \in \mathcal{F} \quad, \vartheta \in \Theta$$

(kompakt notiert also "$\mathrm{d}P_\vartheta = \rho_\vartheta \mathrm{d}\mu$")

Definition 10.86 (Maximum-Likelihood-Schätzer, Likelihood-Funktion)
Für P_ϑ, $\vartheta \in \Theta$, und μ wie oben ist

a) ***Likelihood-Funktion***

$$L_x(\vartheta) = \rho_\vartheta(x), \quad x \in \mathcal{X}, \vartheta \in \Theta$$

b) ***Log-Likelihood-Funktion***

$$\mathcal{L}_x(\vartheta) := \log\left(L_x(\vartheta)\right)$$

*c) $\widehat{\vartheta} \in \Theta$ heißt **Maximum-Likelihood-Schätzer für** ϑ, falls*

$$L_x(\widehat{\vartheta}) = \max_{\vartheta\in\Theta} L_x(\vartheta)$$

*d) $\widehat{g} = g(\widehat{\vartheta})$ mit $\widehat{\vartheta}$ Maximum-Likelihood-Schätzer für ϑ (vgl. c) heißt **Maximum-Likelihood-Schätzer für** die Kenngröße $g(\vartheta)$*

Bemerkung

1. Der Maximum-Likelihood-Schätzer braucht im Allgemeinen nicht eindeutig zu sein.
2. Die mathematische Statistik zeigt: Maximum-Likelihood-Schätzer haben oft (asymptotisch, für wachsenden Stichprobenumfang) gute Eigenschaften.

[12] Frequentistischer Ansatz: alle Wahrscheinlichkeitsmaße werden, grob gesagt, "gleich" behandelt und man macht "gleichmässige Aussagen" (über ϑ) bzgl. jener Familie. Für den bayesschen Ansatz zur Statistik (Bayes, 1701-1761), den wir hier nicht behandeln können, wird bzgl. einer a-priori Verteilung (Wahrscheinlichkeitsmaß) bzgl der ϑ (sic!) zwischen den P_ϑ gemittelt, und eine bedingte a-posteriori Verteilung ggb. die Beobachtungen hergeleitet; Siehe zB. [Georgii, Sekt.7.7]. Für jene Theorie werden bedingen Wahrscheinlichkeitsverteilungen (siehe zB. bedingte Wahrscheinlichkeiten und bedingte Dichten) und "stochastische Kerne" (allgemeiner, vgl. zB. Stochastik-2 VL oder [Klenke]) bzw. reguläre bedingte Wahrscheinlichkeitsmaße wichtig.

[13] Man könnte als Wertebereich hier allgemeiner auch einen messbaren Raum nehmen, wir beschränken uns im Weiteren hier aber auf $\mathbb{R}^d$.

3. Obige Definitionen a) bis d) hängen offenbar zunächst von dem Maß μ ab. Die (log-)Likelihood Funktionen ändern sich auch in der Tat, wenn wir ein anderes μ nehmen; Die M.L. Schätzer hängen jedoch nicht von der Wahl von μ ab. Genauer: Sei $\tilde{p_\vartheta}(x) = dP_\vartheta/d\tilde{\mu}$ für ein zu μ äquivalentes[14] Maß$\tilde{\mu}$. Dann ist die Radon-Nikodym Dichte $d\mu/d\tilde{\mu} > 0$ f.ü (bzgl μ und $\tilde{\mu}$) und $p_\vartheta(x) = \frac{dP_\vartheta}{d\mu}(x) = \frac{dP_\vartheta}{d\tilde{\mu}}\frac{d\tilde{\mu}}{d\mu}(x) = \tilde{p_\vartheta}(x)\frac{d\tilde{\mu}}{d\mu}(x)\,, \quad x \in \mathcal{X}$. Weil der zweite Faktor (f.ü.) strikt positiv ist und nicht von ϑ abhängt, folgt daraus, dass (für fast alle $x \in \mathcal{X}$) $\hat{\vartheta}$ genau dann ein Maximum von $p_\vartheta(x)$ ist, wenn es Maximum von $\tilde{p_\vartheta}(x)$ ist.

4. (Plug-in-Prinzip): Falls $g : \Theta \to \widetilde{\Theta}$ bijektiv ist, so ist $\hat{\vartheta}$ M.L.Schätzer für $\vartheta \in \Theta$ (bzgl. Θ) genau dann, wenn $g(\hat{\vartheta})$ M.L.Schätzer von $\widetilde{\vartheta} := g(\vartheta) \in \widetilde{\Theta}$ ist (bzgl $\widetilde{\Theta}$).

Es gibt zwei typische Fälle für μ:

1) $\mathcal{X}$ abzählbar, $\mathcal{F} = 2^{\mathcal{X}}, P_\vartheta(A) = \sum_{x\in A} \rho_\vartheta(x), \quad A \subset \mathcal{X}$
$\Longrightarrow \ dP_\vartheta = \rho_\vartheta d\mu$ bzgl. Zählmaß $\mu(A) = \sum_{x\in\mathcal{X}} \delta_x(A)$

2) $\mathcal{X} \subset \mathbb{R}^d, \mathcal{F} =$ Spur-σ-algebra von $\mathcal{B}^d(\mathbb{R})$ auf Gebiet $\mathcal{X} \in \mathcal{B}^d(\mathbb{R})$, [z.B.:$\mathcal{X} = \mathbb{R}^d, \mathcal{F} = \mathcal{B}^d$, oder Intervalle etc.]
P_ϑ absolutstetig bzgl. Lebesguemaß, welches wir darum als μ wählen: $\quad dP_\vartheta = \underset{\substack{\uparrow\\ \text{Dichte der absolutstetigen Verteilung } P_\vartheta}}{\rho_\vartheta} d\mu$

Beispiel
Schätzung der Erfolgswahrscheinlichkeit bei i.i.d., Münzwürfen.
Wir beobachten k Erfolge bei n Würfen.
Für $k := x \in \mathcal{X} = \{0, 1, \ldots, n\}$, $p = \vartheta \in \Theta = [0, 1]$, Zähldichte $\rho_\vartheta(k)$ (bzgl. Zählmaß $\mu = \sum_{k=0}^n \delta_k$) ist

$$L_k(\vartheta) = \rho_\vartheta(k) = \binom{n}{k} \vartheta^k (1-\vartheta)^{n-k}, \quad k \in \mathcal{X}\,.$$

Als Maximum-Likelihood-Schätzer für die unbekannte Erfolgswahrscheinlichkeit $\vartheta = p$ errechnet sich [15] hier

$$\hat{\vartheta}(k) = \frac{k}{n}.$$

Präziser formuliert ist der (eindeutige) M.L.Schätzer hier: $\hat{\vartheta} : \mathcal{X} \to \Theta$ mit $x \mapsto \hat{\vartheta}(x) = x/n$.

Erinnerung
Statistisches Modell: $(\mathcal{X}, \mathcal{F}, \{P_\vartheta\}_{\vartheta\in\Theta})$. Um mit Zufallsvariablen X statt mit realisierten beobachteten Daten $x \in \mathcal{X}$ zu argumentieren, sei $X := Id_{\mathcal{X}}$, $X : x \mapsto x$.
Interpretation: $x = X(x)$ ist die (bei bzw als $x \in \mathcal{X}$) realisierte Beobachtung, die Abbildung X steht für die Beobachtung als Zufallsvaiable.

Maximum-Likelihood-Ansatz für Schätzer $\hat{\vartheta}$ für ϑ: $\hat{\vartheta}(x) = \text{argmax}_{\vartheta\in\Theta}\, \rho_\vartheta(x) = \text{argmax}_{\vartheta\in\Theta}\, L_x(\vartheta)$ wobei $dP_\vartheta = \rho_\vartheta d\mu$
Im vorgehenden Münzwurf-Beispiel schreibt sich $\mathcal{X} \ni x \mapsto \hat{\vartheta}(x) = \frac{x}{n} \in \Theta$ damit einfach als $\hat{\vartheta} = \frac{1}{n}X$.

[14] d.h. die Maße haben die gleichen Nullmengen. .
[15] Hinweis: Untersuche getrennt die Fälle $0 < k < n$, $k = 0$, $k = n$.

Einige wünschenswerte Eigenschaften für Schätzer $\hat{g}$ von $g(\vartheta)$ ($g : \Theta \to \mathbb{R}^d$, $\quad \hat{g} : \mathcal{X} \to g(\Theta)$ messbar) sind die Folgenden:

Definition 10.87
*Ein Schätzer $\hat{g}$ für $g(\vartheta)$ heißt **erwartungstreu**, falls*

$$E_\vartheta[\hat{g}(X)] = g(\vartheta) \quad \forall \vartheta \in \Theta\,,$$

wobei E_ϑ den Erwartungswert bzgl.P_ϑ bezeichnet.

Beispiel
Betrachte wieder den Maximum-Likelihood-Schätzer für die Erfolgswahrscheinlichkeit $\rho = \vartheta \in [0,1]$ bei n Münzwürfen mit einer möglicherweise unfairen Münze: Der Schätzer $\hat{g}(x) = \frac{x}{n}$ mit $x =$ Anzahl von Zahl (Erfolg), d.h. $\hat{g} = X/n$, mit $\quad X : x \mapsto X(x)$, $X \overset{P_\vartheta}{\sim}$ Binomial(n, ϑ), ist erwartungstreu.

Definition 10.88
*Eine Folge von Schätzern $\hat{g}_n$ für $g(\vartheta)$, $\vartheta \in \Theta$ heißt (stark)[16]**konsistent**, falls*

$$\hat{g}_n \overset{n\to\infty}{\longrightarrow} g(\vartheta) \quad P_\vartheta\text{-fast sicher} \quad \forall \vartheta \in \Theta\,.$$

Bemerkung
Man kann zeigen, dass Maximum-Likelihood-Schätzer von i.i.d. Stichproben für wachsenden Stichprobenumfang unter gewissen Annahmen typischerweise eine (schwach) konsistente Folge liefern (vgl. Georgii, Satz 7.30).

Beispiel (Wieder der möglicherweise unfaire Münzwurf)
Seien $X_1, X_2, \dots$ i.i.d. Bernoulli(ϑ) (d.h. $X_i \in \{0,1\}$, $i \in \mathbb{N}$), $\vartheta \in [0,1] =: \Theta$, $\mathcal{X} = \{0,1\}^{\mathbb{N}}$, für den Maximum-Likelihood-Schätzer der unbekannten Erfolgswahrscheinlichkeit haben wir

$$\hat{\vartheta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{n\to\infty}{\longrightarrow} E_\vartheta[X_i] = \vartheta \quad P\text{-fast sicher} \quad \forall \vartheta \in \Theta,$$

dank des starken Gesetzes der großen Zahl. Also ist $(\vartheta_n)_{n\in\mathbb{N}}$ eine (stark) konsistente Folge von Schätzern.

Beispiel
Betrachte $\{P_\vartheta \mid \vartheta \in \Theta\}$ abzählbare Familie, so dass $\mathbb{R}$-wertige $X_1, X_2, \dots$ sind unter P_ϑ i.i.d. mit $X_i \in L^2(P_\vartheta)$

a) Betrachte Schätzung von $m_\vartheta := g(\vartheta) := E_\vartheta[X_i]$,
dann ist das arithmetische Mittel $\bar{X}_n := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ ein erwartungstreuer Schätzer .

[16] Gilt die Konvergenz stochastisch, so spricht man von schwacher Konsistenz, welche wir hier aber nicht weiter diskutieren werden.

b) Schätzung von $\sigma_\vartheta^2 := g(\vartheta) := V_\vartheta(X_i) = E_\vartheta[(X_i - m_\vartheta)^2]$, dann ist

$$s_n^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$$

erwartungstreu für σ_ϑ^2.

Bemerkung
$\bar{X}_n$ und s_n^2 sind auch konsistent; für s_n^2: Übung (vgl. H.-O. Georgii)

Beispiel (Maximum-Likelihood-Schätzer im absolutstetigen Fall) (vgl. Georgii)
$X_1, \dots, X_n$ unter P_ϑ i.i.d. mit $X_i \overset{P_\vartheta}{\sim} \mathcal{N}(m, \sigma^2)$, $\sigma > 0$, $m \in \mathbb{R}$, gemeinsame Dichte von $X = (X_1, \dots, X_n)$ bzgl. Lebesguemaß μ

$$L_x(\vartheta) = \rho_\vartheta(x) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - m)^2\right), \quad x \in \mathbb{R}^n =: \mathcal{X}$$

Fall a): $\sigma = \sigma_0$ bekannt, $m \in \mathbb{R}$ unbekannt

$\vartheta \in \Theta := \{(m, \sigma) \mid m \in \mathbb{R}, \sigma = \sigma_0\}$

Als Maximum-Likelihood-Schätzer kann man herleiten: $\widehat{m}(x) = \frac{1}{n}\sum_{i=1}^n x_i$, also $\widehat{m} = \frac{1}{n}\sum_{i=1}^n X_i$.

Fall b): σ und m beide unbekannt

$\Theta := \{(m, \sigma) \mid m \in \mathbb{R}, \sigma \geq 0\}$

Als Maximum-Likelihood-Schätzer errechnen sich hier :

$$\widehat{m} = \frac{1}{n}\sum_{i=1}^n X_i =: \bar{X}_n$$

und

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Man kann nachprüfen, dass zwar $\widehat{m}$, nicht jedoch $\hat{\sigma}^2$, erwartungstreu für m bzw. σ^2 sind. (Letzteres ist jedoch "asymptotisch erwartungstreu" für $n \to \infty$ im Limes, welcher gleich demjenigen für s_n^2 ist.)

Definition 10.89
Der mittlerer quadratische Fehler (engl.: mean squared error, MSE) eines Schätzers $\hat{g}$ für Kenngröße $g(\vartheta)$ unter $\vartheta \in \Theta$ ist

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

Proposition 10.90
(Bias-Varianz-Zerlegung) Für jeden Schätzer $\hat{g}$ von $g(\vartheta)$ mit endlichem MSE $R(\hat{g}, \vartheta) < \infty$ gilt

$$R(\hat{g}, \vartheta) = |E_\vartheta[\hat{g}] - g(\vartheta)|^2 + E_\vartheta\left[|\hat{g} - E_\vartheta[\hat{g}]|^2\right].$$

Statt "Punktschätzer" für unbekannte Parameter betrachtet man auch "Bereichsschätzer" (Vertrauensbereiche, Konfidenzbereiche) :

Definition 10.91
In einem statistischen Model mit $X(x) := x$ und $\alpha \in (0,1)$ heißt eine Abbildung $C : \mathcal{X} \to 2^{\mathbb{R}^d}$, welche jedem möglichen Beobachtungsergebnis $x \in \mathcal{X}$ eine Menge $C(x)$ zuordnet, ein Konfidenzbereich[17] zum Niveau (Irrtumsniveau) α (oder zum Sicherheitsniveau $1-\alpha$) für die zu schätzende Kenngröße $g(\vartheta) \in \mathbb{R}^d$, falls

$$\inf_{\vartheta \in \Theta} P_\vartheta(g(\vartheta) \in C(X)) \geq 1-\alpha \quad \forall \vartheta \in \Theta.$$

Dies erfordert insbesondere, dass $\{g(\vartheta) \in C(X)\} = \{x \in \mathcal{X} \mid g(\vartheta) \in C(x)\}$ für jedes $\vartheta \in \Theta$ eine messbare Menge ist!

Bemerkung 1. Der Niveauparameter muss vor der Datenerhebung so festgelegt werden, wie es für das Anwendungsproblem angemessen ist! Oft findet man für α Werte wie 5% oder 1%, aber auch 0.000001 kann Problemadäquat sein (Wofür, z.B.?).

2. Man will im Allgemeinen für ein vorgegebenes α einen möglichst kleinen Konfidenzbereich finden! Zwar ist etwa $C(x) = g(\Theta)$ ein Konfidenzbereich zu jedem Niveau, aber offenbar nutzlos.

3. Der Konfidenzbereich $C(X) : x \mapsto C(X(x))) = C(x)$ ist zufällig. Für eine realisierte Beobachtung $x \in \mathcal{X}$ ist jedoch $C(x)$ deterministisch, eine Aussage wie "mit Wahrscheinlichkeit von mind. $1-\alpha$ liegt $g(\vartheta)$ in $C(x)$." ist irrig. Für eine realisierte Beobachtung liegt $g(\vartheta)$ entweder in $C(x)$ oder nicht, da ist nichts mehr zufällig. Lediglich vor der Realisierung der Beobachtung, kann man jedoch sinnvoll von einer (Mindest-)Wahrscheinlichkeit sprechen, dass der Zufallsbereich $C(X)$ die unbekannte Kenngröße enthält.

Beispiel Seien $X_1, X_2, \ldots X_n$ unter P_ϑ i.i.d. $N(m, \sigma^2)$ verteilt mit unbekanntem Mittelwert $\vartheta = m \in \mathbb{R}$ und bekannter Varianz $\sigma^2 \in (0,\infty)$. Mit $\frac{1}{n}\sum_{i=1}^n X_i =: \bar{X}_n =: h(X)$ für $X = (X_1, \ldots, X_n)$ und $q > 0$ als $1-\alpha/2$-Quantil der $\mathcal{N}(0,1)$-verteilung (also: $\Phi(-q) = 1 - \Phi(+q) = \alpha/2$ für die Verteilungsfkt. (CDF) Φ der $\mathcal{N}(0,1)$-Verteilung, z.B.[18] $q \approx 1.96$ für $\alpha = 0.05$). Offenbar ist $S_n^* := (\bar{X}_n - m)/(\sqrt{n}\sigma) \sim N(0,1)$ unter jedem P_ϑ, und aus $\{S_n^* \in [-q, +q]\} = \{m \in C(X)\}$ kann man folgern, dass $C(X) := [\bar{X}_n - \sigma q/\sqrt{n}\,,\, \bar{X}_n + \sigma q/\sqrt{n}]$ eine Konfidenzintervall zum (Irrtums-)Niveau α) für $g(\vartheta) = \vartheta = m$ ist. (Sicherheitsniveau ist $1-\alpha$. Man überlegt sich, dass nach Konstruktion dieses Konfindenzintervall unter jedem P_ϑ das Intervall minimaler Länge darstellt, welches das Sicherheitsniveau $1-\alpha$ einhält.

Beispiel Allgemeiner sind oft die X_i zwar vielleicht i.i.d., die Verteilung von X_i jedoch nicht normal, und natürlich ist auch die Varianz $\sigma^2 \in (0,\infty)$, anders als vorigem Beispiel, oft nicht bekannt. Man kann auch in dieser allgemeineren Situation mit Hilfe des Zentralen Grenzwertsatzes ganz ähnlich argumentieren wie in dem Beispiel, um zumindest ein *approximatives* Konfidenzintervall herzuleiten (vgl. Übungsaufgabe): S_n^* ist dann lediglich approximativ $\mathcal{N}(0,1)$-verteilt, wir bestimmen q wie zuvor, die Wahrscheinlichkeitsabschätzungen von zuvor gelten nun also nur approximativ für große n; Falls σ zudem unbekannt ist, kann man in der Festlegung des (approximativen) Konfidenzintervalles $C(X)$ den Parameter σ ersetzen noch durch seinen Schätzer $\sqrt{s_n^2(X)}$. Vgl. zB. [Georgii, Bsp.8.4], Übungsblatt 9, 2019, A.2, und [Georgii, Satz.9.12 und Sekt.12.3, Kor.12.19 u. 12.20].

Grundlagen der Testtheorie

Idee: Teste "Null"-Hypothese(Θ_0) gegen die Alternative(Θ_1)

$$\Theta = \Theta_0 \underset{\substack{\uparrow \\ \text{disjunkt}}}{\dot\cup} \Theta_1$$

Beispiel (einfache Hypothese vs. einfache Alternative)
$\Theta_0 = \{\vartheta_0\}$, $\Theta_1 = \{\vartheta_1\}$, $\Theta = \{\vartheta_0, \vartheta_1\}$ $\quad \vartheta_0 \neq \vartheta_1$
Sprechen Stichproben $\mathcal{X}$ für oder gegen $P_0 := P_{\vartheta_0}$?
(Gegen bzw. für $P_1 := P_{\vartheta_1}$?)

[17] oft auch genannt: Konfidenzintervall (bei $d = 1$, falls Intervall), Bereichsschätzer (Betonung der Unterscheidung von gewöhnlichen Schätzern, welche Punktschätzer sind) oder Vertrauensbereich.
[18] Siehe z.B. Verteilungstabllen in der Literatur, z.B. im Georgii.

Definition 10.92
$(\mathcal{X}, \mathcal{F}, P_\vartheta)$, $\Theta = \Theta_0 \dot\cup \Theta_1$.
Ein Test von Nullhypothese Θ_0 vs. Alternative Θ_1 ist gegeben durch einen Ablehnungsbereich $R \subset \mathcal{X}$ (mit $R \in \mathcal{F}$).

Interpretation ist:

$$x \in R \curvearrowright \Theta_0 \quad \underline{\textit{wird abgelehnt}}$$
$$x \notin R \curvearrowright \Theta_0 \quad \underline{\textit{wird nicht abgelehnt}}$$

Beispiel
Medikamententest:

$\Theta_0 \hat{=}$ Medikament wirkt wie Placebo
$\Theta_1 \hat{=}$ Medikament wirkt besser

Fehlerwahrscheinlichkeiten eines Tests

- Wahrscheinlichkeit für Fehler erster Art (irrtümliches Verwerfen von Θ_0):

$$\sup_{\vartheta \in \Theta_0} P_\vartheta(R) \text{ mit } P_\vartheta(R) \equiv P_\vartheta(X \in R), \quad \vartheta \in \Theta_0$$

- Fehler zweiter Art (irrtümliches Nicht-Verwerfen von Θ_0) unter gegebenem $\vartheta \in \Theta_1$ ist:

$$\mathcal{P}_\vartheta(R^c) = 1 - P_\vartheta(R), \quad \vartheta \in \Theta_1$$

Definition 10.93
*der Test R (von engl. rejection) für Θ_0 vs. Θ_1 hat **(Irrtums-)Niveau** $\alpha \in [0,1]$ (bzw. Sicherheitsniveau $1-\alpha$), falls*

$$\sup_{\vartheta \in \Theta_0} P_{\vartheta_0}(R) \leq \alpha$$

*Der Test hat unter P_{ϑ_1} für $\vartheta_1 \in \Theta_1$ die **Macht** $P_{\vartheta_1}(R)$.*

Bemerkung

- typisch z.B. $\alpha = 1\%$,
(Irrtumsniveau 1% sichert, dass die Wahrscheinlichkeit für Fehler 1. Art kleiner als 1% ist. Da für ein praktisches Testproblem (vor der Auswertung des Tests!) das Niveau α problemadäquat festzulegen ist, ist für einen Test mit Irrtumsniveau α (d.h. Sicherheitsniveau $1-\alpha$) demzufolge die **Ablehnung** der Nullhypothese das auf diesem Niveau signifikante Ereignis. Dieses Niveau sagt aber nichts über die Wahrscheinlichkeit einer irrtümlichen Nicht-Ablehnung der Nullhypothese aus.
- Die Macht gibt an, wie wahrscheinlich sich die Alternative unter einem Maß P_{ϑ_1} aus der Alternative $\vartheta_1 \in \Theta_1$ "durchsetzt". Die Wahrscheinlichkeit, dass unter einem Maß P_{ϑ_1} aus der Alternative $\vartheta_1 \in \Theta_1$ die Nullhypothese nicht abgelehnt wird, ist $P_{\vartheta_1}(R^c) = 1 - P_{\vartheta_1}(R)$.

Im Weiteren:

- **einfache** Hypothese $\Theta_0 = \{\vartheta_0\}$
- **einfache** Alternative $\Theta_1 = \{\vartheta_1\}, \quad \Theta = \{\vartheta_0, \vartheta_1\}$

Satz 10.94 (und Definition, Neyman-Pearson)
Optimale Tests für einfache Hypothese vs. einfache Alternative:
Seien $P_i := P_{\vartheta_i}$, $i = 0, 1$, Wahrscheinlichkeitsmaße aus $(\mathcal{X}, \mathcal{F})$ mit Dichten ρ_i bzgl. eines Maßes μ auf $(\mathcal{X}, \mathcal{F})$ (oft Lebesgue- oder Zählmaß), d.h.

$$\begin{aligned} \mathrm{d}P_i &= \rho_i \mathrm{d}\mu, \quad i = 0,1 \text{ bzw.} \\ P_i(A) &= \int \mathbb{1}_A(x)\rho_i(x)\mu(\mathrm{d}x), \quad A \in \mathcal{F}, i = 0,1 \end{aligned}$$

und mit $\rho_0 > 0$.
Zum Testen von $\Theta_0 = \{\vartheta_0\}$ vs. $\Theta_1 = \{\vartheta_1\}$ auf statistischem Modell $(\mathcal{X}, \mathcal{F}, \{P_\vartheta\}_{\vartheta\in\Theta})$ konstruiert man einen Test mit maximaler Macht bei einem Irrtumsniveau α wie folgt:
*Der **Likelihoodquotient** (Likelihood-Ratio) ist definiert als*

$$q(x) := \frac{\rho_1(x)}{\rho_0(x)}, \quad x \in \mathcal{X}.$$

*Dann ist der Test $R := R_c := \{q \geq c\}$, $\quad c > 0$, **optimal** zum Niveau α mit $\alpha := P_0(R)$ in dem Sinne , dass R maximale Macht hat: D.h. jeder andere Test R' mit $P_0(R') \leq \alpha$ hat weniger Macht als R, d.h. $P_1(R') \leq P_1(R)$*

Beispiel
Betrachte $X_1, \ldots, X_n$ i.i.d. normalverteilt mit

- X_i i.i.d. $\sim \mathcal{N}(0,1)$ unter P_0 und
- X_i i.i.d. $\sim \mathcal{N}(m,1)$ unter P_1

mit $m \in \mathbb{R}, m \neq 0$.
Wir testen:

$$\begin{aligned} \Theta_0 &\mathrel{\widehat{=}} \text{Mittelwert ist } 0 = \{(0,1)\} \\ \Theta_1 &\mathrel{\widehat{=}} \text{Mittelwert ist } m = \{(m,1)\} \\ \Theta &= \{(0,1), (m,1)\} \end{aligned}$$

Der **Likelihoodquotient** $q(x)$ hat die Form

$$\begin{aligned} q(x) &= \frac{\frac{1}{(2\pi\cdot 1)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\sum_{i=1}^n (x_i - m)^2\right)}{\frac{1}{(2\pi\cdot 1)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\sum_{i=1}^n (x_i - 0)^2\right)} \\ &= \exp\left(\langle x, m \cdot \mathbf{1}\rangle - \frac{1}{2}nm^2\right) \quad \text{(nur hier) mit } \mathbf{1} := (1, \ldots, 1)^T \in \mathbb{R}^n \end{aligned}$$

Optimale Tests haben die Form

$$\begin{aligned} R = \{x \mid q(x) \geq c_1\} &\overset{\downarrow c_1 = \exp\left(c_2 - \frac{1}{2}nm^2\right)}{=} \{x \mid \langle x, m \cdot \mathbf{1}\rangle \geq c_2\}, \\ &= \{x \mid m \sum_{i=1}^n x_i \geq c_2\} \end{aligned}$$

Für das **Niveau** von R unter P_0 erhalten wir

$$P_0(R) = 1 - \Phi\left(\frac{c_2}{\sqrt{nm^2}}\right) = \Phi\left(-\frac{c_2}{\sqrt{nm^2}}\right),$$

wobei $\Phi(\cdot)$ die Verteilungsfunktion von $\mathcal{N}(0,1)$ bezeichnet.
Die **Macht** von R ist

$$P_1(R) = 1 - \Phi\left(\frac{c_2 - m^2 n}{\sqrt{nm^2}}\right) = \Phi\left(-\frac{c_2 - m^2 n}{\sqrt{nm^2}}\right)$$

Bemerkung
Für vorgegebenen Niveauparameter $\alpha \in (0,1)$ kann man nun $c_2 = c_2(\alpha, n, m)$ so bestimmen, dass der Test $R = R(\alpha)$ das Niveau $P_0(R) = \alpha$ hat: Für $-\frac{c_2}{\sqrt{nm^2}} = \Phi^{-1}(\alpha)$ ist also die Macht des Tests gleich $\Phi\left(\Phi^{-1}(\alpha) + |m|\sqrt{n}\right)$ und wir sehen, dass die Macht für $m^2 n \to \infty$ gegen Eins geht, z.B. bei fixem m für $n \to \infty$; Bei fixem Stichprobenumfang n sehen wir, dass für $|m| \to \infty$ die Macht ebenfalls asymptotisch gegen Eins geht, während Sie für $|m| \downarrow 0$ gegen α geht.

Übung: Interpretieren Sie diese und ggf weitere ähnliche Aussagen - was bedeuten Sie für das Testproblem?