Stochastik I Skript

Marlene Kretschmer, Moana-Rose Marcello, Steffen Miels, Jacqueline Rosar

15. Juli 2010

Stochastik I
Prof. D. Becherer

Sommersemester 2010
Mitschrift von Marlene Kretschmer, Moana-Rose Marcello, Steffen Miels, Jacqueline Rosar
mit Korrekturen durch Prof. Becherer
Letzte Änderung: 15. Juli 2010

Diese Mitschrift enthält trotz mehrmaliger Durchsicht noch kleinere oder größere Fehler. Die Autoren und der Dozent freuen sich über entsprechende Hinweise. miels@math.hu-berlin.de

Inhaltsverzeichnis

1 Wahrscheinlichkeitstheorie **3**
Einführung: Beispiele und der Satz von Vitali 3

2 Wahrscheinlichkeitsräume **9**
Axiomatische Grundlagen für $(\Omega, \mathcal{F}, P)$ und erste Eigenschaften 9

3 Bedingte Wahrscheinlichkeit und Unabhängigkeit **29**

4 Asymptotische Ereignisse **38**

5 Erwartungswert und Varianz **40**

6 Die Gesetze der großen Zahlen **46**

7 Charakteristische Funktion **54**

8 Mehrdimensionale Normalverteilungen **63**

9 Konvergenz in Verteilung / schwache Konvergenz **69**
Beziehung zw. d. Konvergenzarten 71

10 Grundbegriffe der Schätztheorie **83**
Elementare Testtheorie 88

Index **92**

1 Wahrscheinlichkeitstheorie

Einführung: Beispiele und der Satz von Vitali

Stochastik	
Wahrscheinlichkeit (Modellierung, Schlussfolgerung 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.

Beispiel

Einfaches Würfeln

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

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

n-maliges Würfeln

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

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

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

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 $P[\{w\}] = \dfrac{1}{|\Omega|}$ zuordnet.

einfacher Münzwurf

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

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

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

n-facher Münzwurf

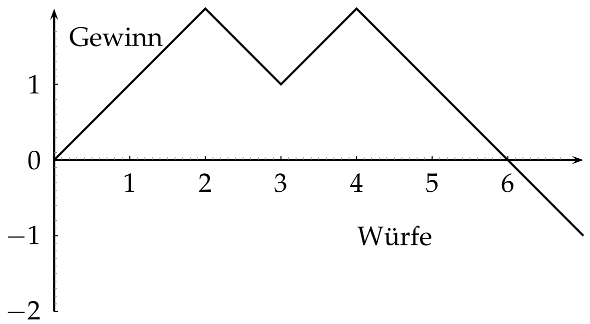
$$\Omega = \{0,1\}^n$$

∞-facher Münzwurf

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

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

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



Aktienindexkurse

$$\Omega = C([0,T], \mathbb{R})$$
$$w(t) \leftrightarrow \text{Aktienkurs zur Zeit } t \leq T$$
$$A = \{w \in \Omega \mid w(t) \geq 5000 \; \forall t \in [0,T]\} \leftrightarrow \text{Dax bleibt über 5000}$$

Bemerkung
Das Ziel ist es, jedem Ereignis A eine Wahrscheinlichkeit $P[A] \in [0,1]$ zuzuordnen, so dass P als Abbildung von

$$P: \quad \mathcal{F} \subset \mathcal{P}(\Omega) \to [0,1]$$

"schöne" Eigenschaften hat.

Definition 1.0 (Potenzmenge)
*Die **Potenzmenge** $\mathcal{P}(\Omega)$ ist die Menge aller Teilmengen von Ω. Häufige Notation: $\mathcal{P}(\Omega) = 2^{\Omega}$*

Beispiel

$$\Omega = \{0,1\}$$
$$\mathcal{P}(\Omega) = \{\emptyset, \Omega, \{1\}, \{0\}\}$$

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

- $P[\Omega] = 1$
- *Sind $A_k \in P(\Omega)$ disjunkte Teilmengen von Ω, $k \in \mathbb{N}$, so gilt, dass*

$$P[\biguplus_{k=1}^{\infty} A_k] = \sum_{k=1}^{\infty} P[A_k],$$

wobei $\uplus$ eine disjunkte Vereinigung bezeichne.

- *$\forall A \in \mathcal{P}(\Omega)$ gilt: $P[A] = P[T_k(A)]$ $\forall k$, wobei T_k einen "Flip" der k-ten Koordinate beschreibt:*

  Definition eines "Flips":
  $T_k : \Omega \to \Omega$
  $(w_i)_{i \in \mathbb{N}} \mapsto (w_1, \dots, w_{k-1}, 1 - w_k, w_{k+1}, \dots)$

Beweis:

Sei auf Ω eine Äquivalenzrelation $\sim$ wie folgt definiert:

$$w \sim w' \Leftrightarrow \exists n_0 \in \mathbb{N} : w_k = w'_k \quad \forall k \geq n_0$$

Ω zerfällt in disjunkte Äquivalenzklassen, nach dem Auswahlaxiom können wir aus jeder Äquivalenzklasse einen Vertreter wählen. Sei A die Menge dieser Vertreter. Sei $\mathfrak{S} := \{S \subset \mathbb{N} : \ |S| < \infty\}$ (endliche Teilmenge). $\mathfrak{S}$ ist abzählbar als abzählbare Vereinigung abzählbarer Mengen. Für

$$S = \{n_1, \cdots, n_k\}$$

und

$$T_s := T_{n_1} \circ T_{n_2} \circ \cdots \circ T_{n_k}$$

gilt dann:

$$\Omega = \biguplus_{S \in \mathfrak{S}} T_s(A)$$

(disjunkte Vereinigung), denn in der Tat:

(i) "$\subset$"

$$\text{Für } w \in \Omega \implies \exists w' \in A \text{ mit } w' \sim w \text{, also } \exists S \in \mathfrak{S} \text{, so dass: } T_s(w') = w \implies w \in T_s(A)$$

(ii) "disjunkt"

$$T_S(A) \cap T_{S'}(A) \neq \emptyset \text{ für } S, S' \in \mathfrak{S} \implies S = S'$$

$$\text{Wähle } w, w' \in A \text{ mit } T_S(w) = T_{S'}(w') \implies w \sim T_S(w) \underset{(\sim)}{=} T_{S'}(w') \sim w' \implies w = w' \implies S = S'$$

Nun folgt für die möglichen Fälle

a)

$$P[A] = 0 \implies P[T_S(A)] = 0 \implies P[\Omega] = \sum_{S \in \mathfrak{S}} P[T_S(A)] = 0 \neq 1$$

b)

$$P[A] =: \epsilon > 0 \implies P[\Omega] = \sum_{S \in \mathfrak{S}} P[T_S(A)] = +\infty \neq 1$$

$\square$

Bemerkung

$\{0,1\}^{\mathbb{N}}$ ist "fast bijektiv" zum $[0,1) = \mathbb{R} \mod 1$:
Definiere $f(\omega)$ für $\omega = (\omega_i)_{i \in \mathbb{N}}$ durch $f(\omega) = \sum_{\mathbb{N}} \omega_i 2^{-i}$
dann gibt es nur abzählbar viele $\omega \in \{0,1\}^{\mathbb{N}}$, für welche ein $\omega' \neq \omega \in \{0,1\}^{\mathbb{N}}$ existiert mit $f(\omega) = f(\omega')$;
$f : \{0,1\}^{\mathbb{N}} \to [0,1]$ ist surjektiv, und mit $\omega \sim \omega' :\Leftrightarrow f(\omega) = f(\omega')$ gibt es in $\{0,1\}^{\mathbb{N}}|_\sim$ nur abzählbar viele Äquivalenzklassen, die mehr als einen Vertreter in $\{0,1\}^{\mathbb{N}}$ haben.

Erläuterung:

$$\Omega := \{0,1\}^{\mathbb{N}} \leftrightarrow [0,1]$$
$$f:\ \omega \mapsto \sum_{k\in\mathbb{N}} \omega_k \cdot 2^{-k}$$

ist surjektiv, aber nicht injektiv. Da z.B.:

$$(0,1,1,\dots) \mapsto \frac{1}{2}$$
$$(1,0,0,\dots) \mapsto \frac{1}{2}$$

jedoch Bijektion $\{0,1\}^{\mathbb{N}}|_{\sim} \leftrightarrow [0.1)$ mit $\omega \sim \omega' \iff f(\omega) = f(\omega')$.
Wenn wir Repräsentanten wählen, dann fallen nur abzählbar viele Elemente aus Ω raus, also bei "Gleichverteilung" eine "Nullmenge".

Bemerkung
Da nun jedes Element aus Ω im Wesentlichen mit einer reellen Zahl zwischen Null und Eins durch die dyadische Zahlendarstellung identifiziert werden kann, bedeutet dies, dass keine P-artige Abbildung auf allen Teilmenge der reellen Zahlen zwischen 0 und 1 definiert werden kann, welche nicht eine der 3 Eigenschaften für P aus Satz 1.1 verletzt. Dies legt folgendes Resultat nahe, das wir direkt beweisen.

Satz 1.2
Sei $\Omega := [0,1) = \mathbb{R} \mod 1$. *Es gibt* *<u>keine</u>* *Abbildung* $P : \mathcal{P}(\Omega) \to [0,1)$ *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) $P[A] = P[T_x(A)] \qquad \forall A \in \mathcal{P}(\Omega)$

wobei $T_x : \Omega \to \Omega,\quad a \mapsto a + x \mod 1$

und $T_x(A) = \{y \in \Omega \mid y - x \mod 1 \in A\}$

Beweis:
Definiere Äquivalenzrelation auf Ω via

$$\omega \sim \omega' \iff \omega - \omega' \in \mathbb{Q} \text{ (rational)}$$

Mittels Auswahlaxiom können wir aus jeder Äquivalenzklasse genau einen Vertreter wählen, und es sei A die Menge dieser Vertreter. Dann ist

$$\Omega = \biguplus_{x\in\Omega\cap\mathbb{Q}} T_x(A)$$

abzählbare Vereinigung disjunkter Mengen und es gilt:

- "$\subset$"
$$\omega \in \Omega \implies \exists\, \omega' \in A \text{ mit } \omega \sim \omega' \implies \omega - \omega' \in \mathbb{Q} \implies \omega \in T_x(A) \text{ für } x = \omega - \omega'$$

- "Disjunktheit"

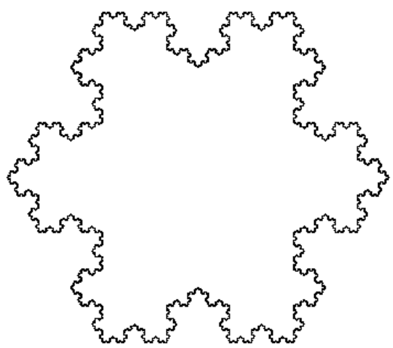
$$
\begin{aligned}
&\text{Sei } x, x' \in [0,1) \cap \mathbb{Q} \text{ sodass } T_x(A) \cap T_{x'}(A) \neq \emptyset \\
&\implies \exists\, \omega, \omega' \in A \text{ sodass } T_x(\omega) = T_{x'}(\omega') \\
&\implies \omega \sim T_x(\omega) \underset{(=)}{\sim} T_{x'}(\omega') \sim \omega' \implies \omega = \omega'
\end{aligned}
$$

also $1 = P[\Omega] \overset{\text{ii)}}{=} \sum_{x \in [0,1) \cap \mathbb{Q}} P[A] \quad \lightning,$
da Gleichverteilung von Wahrscheinlichkeiten bei abzählbarer Aufsummierung $+\infty$ oder 0 ergibt ($P(A) \in [0,1)$). □

Bemerkung
Forderungen i)-iii) in Satz 1 bzw Satz 2 sind "unvernünftig"; was sich als nicht möglich herausstellt, ist diese Forderungen auf der Potenzmenge zu realisieren!

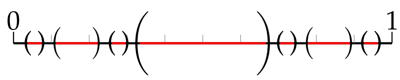
Beispiele dafür, dass abzählbare Additivität wünschenswert ist:

Beispiel Koch'sche Schneeflocke



$$
\text{Fläche } \underbrace{= a\frac{\sqrt{3}}{4}\left(1 + \frac{1}{3} + \frac{4}{3^3} + \frac{4^2}{3^5} + \cdots\right)}_{\text{beschränkt und monoton} \implies \text{konvergent}}
$$

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.
$\mathcal{C}^c$ hat die Länge:

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

Bemerkung: $\mathcal{C}$ ist überabzählbar!

Beispiel
dafür, dass es wünschenswert ist, mit abzählbaren Mengenoperationen in der Menge der "messbaren" Ereignisse ($\mathcal{F}$) zu bleiben:

$$\begin{aligned}
\Omega &= \{0,1\}^{\mathbb{N}} \\
A_k : &= \{\omega \in \Omega \mid \omega_k = 1\} \\
A : &= \text{"Zahl kommt unendlich oft"} \\
&= \underbrace{\bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n}}_{\text{abzählbare Mengenoperationen}} A_k
\end{aligned}$$

2 Wahrscheinlichkeitsräume

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

Definition 2.3
*Sei $\Omega \neq \emptyset$. Ein Mengensystem $\mathcal{F} \subset \mathcal{P}(\Omega)$ heißt **σ-Algebra**, falls*

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

(ii) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$, *wobei* $A^c = \Omega \setminus A$

(iii) $A_k \in \mathcal{F}, k \in \mathbb{N} \implies \bigcup_{k \in \mathbb{N}} A_k \in \mathcal{F}$

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* $\implies P[\dot\bigcup_{k \in \mathbb{N}} A_k] = \sum_{k \in \mathbb{N}} P[A_k]$

*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.

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 Morgan gilt:

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

(vii) $A_k \in \mathcal{F}, k \in \mathbb{N} \implies \bigcap_{k \in \mathbb{N}} A_k \in \mathcal{F}$

Beispiel
Sei $\Omega \neq \emptyset$. Dann ist $\mathcal{P}(\Omega) = 2^\Omega$ eine σ-Algebra.

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

Lemma 2.5 (Definition und Lemma)
Sei $\Omega \neq \emptyset$, $\quad \mathcal{G} \subset \mathcal{P}(\Omega)$
Zudem sei $\mathcal{A}$ eine beliebige σ-Algebra.
Dann existiert eine kleinste σ-Algebra $\mathcal{F}$ auf Ω, so dass $\mathcal{G} \subset \mathcal{F}$.

Beweis:
$\mathcal{F} := \bigcap\limits_{\mathcal{G} \subset \mathcal{A}} \mathcal{A}$ ist σ-Algebra, wobei der Durchschnitt über alle σ-Algebren $\mathcal{A}$ mit $G \subset \mathcal{A}$ genommen wird (Übung, prüfe *(i)-(iii)* aus Definition nach). $\mathcal{F}$ ist natürlich die kleinste! D.h. jede σ-Algebra $\overline{\mathcal{F}}$, die $\mathcal{G}$ enthält, erfüllt $\mathcal{F} \subset \overline{\mathcal{F}}$. Wir schreiben: $\mathcal{F} = \sigma(\mathcal{G})$ für die von $\mathcal{G}$ erzeugte σ-Algebra ist.

$\square$

Sei (Ω, d) ein metrischer Raum.

z.B. $(\Omega, d) = (\mathbb{R}^n, d)$ mit $d(x, y) = |x - y|$

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

Beispiel
$\mathcal{B}^n = \mathcal{B}(\mathbb{R}^n)$ auf $\mathbb{R}^n$

Bemerkung
$\mathcal{B}^n$ ist "sehr groß", aber $\subsetneq \mathcal{P}(\mathbb{R}^n)$

$\mathcal{B}^n$ kann von vielen, verschiedenen Erzeugersystemen $\Phi \subset \mathcal{P}(\mathbb{R}^n)$ erzeugt werden, d.h. $\sigma(\Phi) = \mathcal{B}^n$

Bezeichne

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

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

Beweis:

(i) "$\subset$"
Sei $G \in \mathcal{G}$, dann ist G abzählbare Vereinigung von Elementen aus $\mathcal{K}$
$\implies G \in \sigma(\mathcal{K})$
$\implies \mathcal{G} \subset \sigma(\mathcal{K})$
$\implies \sigma(\mathcal{G}) \subset \sigma(\mathcal{K})$

"$\supset$"

$$\mathcal{K} \ni \textstyle\prod_{i=1}^n [a_i, b_i] = \underbrace{\bigcap_{m \in \mathbb{N}} \underbrace{\prod_{i=1}^{n}]a_i - \frac{1}{m}, b_i + \frac{1}{m}[}_{\in \mathcal{G}}}_{\in \sigma(\mathcal{G})}$$

$\implies \mathcal{K} \subset \sigma(\mathcal{G})$
$\implies \sigma(\mathcal{K}) \subset \sigma(\mathcal{G})$

(ii) $\sigma(\mathcal{G}) = \sigma(\mathcal{A})$, denn für $\mathcal{F} \in \mathcal{A}$

$$\implies \mathcal{F} = (\underbrace{\underbrace{\mathcal{F}^c}_{\in \mathcal{G}})^c}_{\in \sigma(\mathcal{G})} \in \sigma(\mathcal{G})$$

$\implies \sigma(\mathcal{A}) \subset \sigma(\mathcal{G})$

analog für $\sigma(\mathcal{G}) \subset \sigma(\mathcal{A})$

$\square$

Lemma 2.8
Es gilt: $\mathcal{B}^n$ 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 Aussage gilt für $\mathcal{B}^n = \sigma(\Phi_i)$ mit

$$\Phi_1 = \{\prod_{i=1}^{n} (-\infty, b_i) \mid b_i \in \mathbb{R}\}$$

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 weiterer, wichtiger Typ von σ-Algebren betrifft Meßbarkeitsstrukturen auf Produkträumen

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

mit beliebiger Indexmenge $I \neq \emptyset$ und meßbaren Räumen $(E_i, \mathcal{E}_i)$.

Projektionen:

$$\begin{aligned} \pi_i : \qquad\qquad & \Omega \to E_i \\ \omega = (\omega_i)_{i \in I} & \mapsto \omega_i \end{aligned}$$

$$\mathcal{G} := \{\pi_i^{-1}(A_i) \mid A_i \in \mathcal{E}_i\}$$

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

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

<u>Notationen</u>
Falls $(E_i, \mathcal{E}_i) = (E, \mathcal{E})\ \forall i$ schreibt man auch

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

und falls I endlich ist, d.h. $|I| = n,\ n \in \mathbb{N}$, auch

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

Bemerkung
Es gilt: $\mathcal{B}^n \equiv \mathcal{B}(\mathbb{R}^n) = \bigotimes_{i=1}^n \mathcal{B}^1 = (\mathcal{B}^1)^n$ mit Konvention $\mathcal{B} = \mathcal{B}^1$, also $\mathcal{B}^n \equiv \mathcal{B}(\mathbb{R}^n) = (\mathcal{B})^n$, also Notation nicht irreführend.

Erinnerung
Wenn P ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{F}, P)$ ist, dann gilt:

- $P : \mathcal{F} \to [0,1]$
- $P[\Omega] = 1$
- $P[\dot{\bigcup}_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i]$

Satz 2.11 (Elementare Eigenschaften von Wahrscheinlichkeistmaßen *P*)

(i) $P[\emptyset] = 0$

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

(iii) $A, B \in \mathcal{F} \wedge A \subset B \implies P[A] \leq P[B]$ ***(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 $A_n \downarrow A$, d.h. $A_n \supset A_{n+1}\ \forall n$ und $\bigcap_{n \in \mathbb{N}} A_n = A$, dann gilt:*

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

und 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:

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

*(**σ-Stetigkeit** von P)*

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

Beweis: (teilweise, Rest ist Übung)

(i) $\emptyset = \emptyset \cup \emptyset$
$\emptyset \cap \emptyset = \emptyset$ d.h. disjunkt
$\implies P[\emptyset] = \sum_{\mathbb{N}} P[\emptyset] \implies P[\emptyset] = 0$

(ii) $A \cup B = (A \setminus B) \dot{\cup} (A \cap B) \dot{\cup} (B \setminus A)$
$A = (A \setminus B) \dot{\cup} (A \cap B)$ etc.

(iii) Übung , (vgl. Georgii)

(iv) Übung

(v) Übung

(vi)

$$\limsup_{n\to\infty} A_n := \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} A_m$$ "A_n treten ∞ oft auf"

$$\liminf_{n\to\infty} A_n := \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} A_m$$ "schließlich alle" bzw. "alle bis auf endlich viele treten auf"

Aus $A_n \to A$ folgt $\limsup A_n = \liminf A_n = A$. Definiere

$$B_n := \bigcap_{m \geq n} A_m \quad \nearrow A \qquad \text{(wachsend gegen } A\text{)}$$
$$C_n := \bigcup_{m \geq n} A_m \quad \searrow A \qquad \text{(fallend gegen } A\text{)}$$

Dann folgt aus Monotonie und Quetschlemma

$$P[B_n] \leq P[A_n] \leq P[C_n]$$
$$\overset{n\to\infty}{\Longrightarrow} P[A] \leq \lim_{n\to\infty} P[A_n] \leq P[A]$$

□

Bisher $(\Omega, \mathcal{F}, P)$ Wahrscheinlichkeitsraum

Im Weiteren: mehr Beispiele für Wahrscheinlichkeitsmaße ("Verteilungen"), Zufallsvariable

Wichtiger Fall von Wahrscheinlichkeitsmaßen:
Ω abzählbar
$\mathcal{F} = \mathcal{P}(\Omega) \equiv 2^\Omega$

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

Satz 2.13
Sei Ω abzählbar, $\rho(\omega),\ \omega \in \Omega$, Elemente in $\mathbb{R}_+$ mit $\sum_{\omega\in\Omega} \rho(\omega) = 1$
Dann existiert genau ein Wahrscheinlichkeitsmaß P auf $(\Omega, \mathcal{P}(\Omega))$ mit $P(A) = \sum_{\omega\in A} \rho(w) \quad \forall A \in \mathcal{P}(\Omega)$
ρ heißt "Zähldichte", die $\rho(\omega),\ \omega \in \Omega$, nennt man auch "Wahrscheinlichkeitsgewichte" der $\omega \in \Omega$

Beweis:

Existenz: $P(A),\ A \in \mathcal{P}(\Omega)$, definiert wie im Satz 2.13 ist ein Wahrscheinlichkeitsmaß:

$$P(\Omega) \equiv \sum_{\omega\in\Omega} \rho(\omega) = 1 \quad \checkmark \text{ nach Voraussetzung}$$
$$P(\bigcup_{k\in\mathbb{N}}^{\cdot} A_k) = \sum_{k\in\mathbb{N}} \left(\sum_{\omega\in A_k} \rho(\omega) \right) = \sum_{k\in\mathbb{N}} P(A_k) \quad \checkmark$$

Eindeutigkeit:

$$P_1(A) = \sum_{\omega\in A} \rho(\omega) = P_2(A) \quad \forall A \in \mathcal{P}(\Omega)$$

□

Beispiel (Diskrete Verteilungen)

- Ω endlich, $\rho(\omega) = \dfrac{1}{|\Omega|} \implies P(A) = \dfrac{|A|}{|\Omega|} = \dfrac{\text{"günstige"}}{\text{"mögliche"}}$
- Produktverteilungen für mehrstufige Experimente ("Baumdiagramm"): siehe unten

Seien $(E_i, \mathcal{E}_i)$ diskrete, endliche ($|E_i| < \infty,\ \mathcal{E}_i = \mathcal{P}(E_i)$) Maßräume mit "Zähldichten" $\rho_i,\ i \in I = \{1, \cdots, n\}$.
Definiere:

$$\Omega := \bigtimes_{i=1}^{n} E_i \equiv \prod_{i=1}^{n} E_i \qquad \text{(kartesisches Produkt)}$$
$$\mathcal{F} := \bigotimes_{i=1}^{n} \mathcal{E}_i$$

Definition 2.14
Das Wahrscheinlichkeitsmaß P, das durch die Zähldichte

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

*auf $\mathcal{F} = \mathcal{P}(\Omega)$ gegeben ist (Übung), heißt **diskretes Produktmaß**. Wir schreiben:*

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

Falls $(E_i, \mathcal{E}_i, \rho_i) = (E_1, \mathcal{E}_1, \rho_1) \quad \forall i$ gilt, schreiben wir auch:

$$P = \rho_1^{\otimes n}$$

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

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

P bzgl. $\rho = \prod_{i=1}^{n} \rho_1 = \rho_1^{\otimes n}$ liefert:

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

.

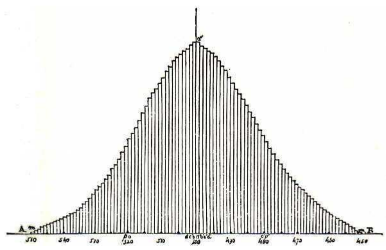
Bemerkung
Man spricht für dieses P auf $\mathcal{P}(\{0, 1\}^n)$ von einem "Bernoulli-Schema".

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

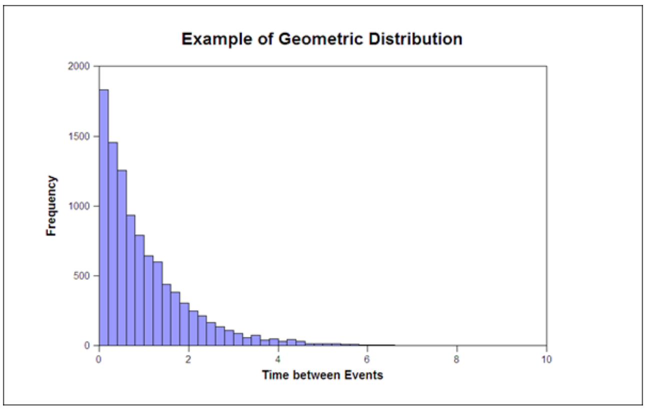
$$
\begin{aligned}
A &= \{\omega \in \Omega \mid \sum_{i=1}^{n} \omega_i = k\} \\
\implies P(A) &= \sum_{\omega \in A} \underbrace{P(\{\omega\})}_{p^k(1-p)^{n-k}} \\
&= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\
&= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} =: \mathrm{Bin}_{n,p}(k)
\end{aligned}
$$

Beispiel für diskrete Verteilungen

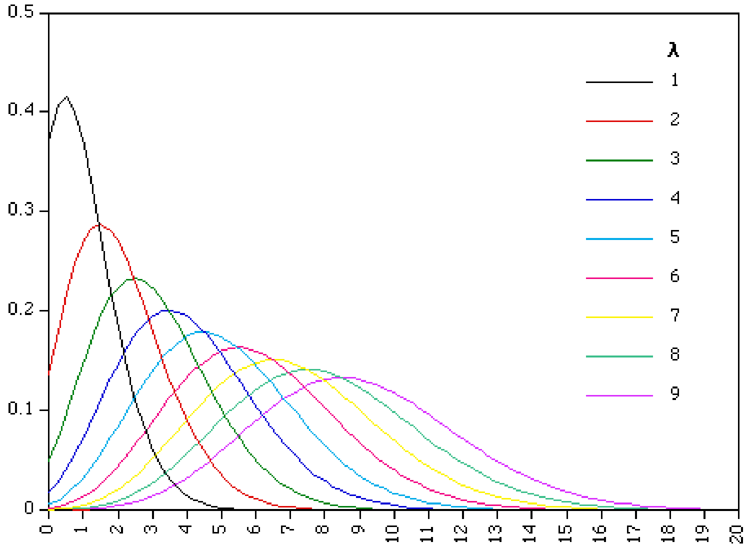
- *Binomialverteilung*
  auf $\Omega = \{0, 1, \ldots, n\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ mit Zähldichte $\rho(k) = \mathrm{Bin}_{n,p}(k)$, $k \in \Omega$, für Parameter $n \in \mathbb{N}$, $p \in [0,1]$



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



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



(Hinweis: Hier wurden diskrete Histogrammdaten stetig interpoliert.)
Die Poissonverteilung ist ein wichtiger Grenzfall der Binomialverteilung, gemäß dem Poisson'schen Grenzwertsatz:

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

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

Beweis:
Schreibe $A_n \sim B_n$ ("asymptotisch gleich"), falls $\lim\limits_{n\to\infty} \frac{A_n}{B_n} = 1, \quad B_n \neq 0 \quad \forall n$
Damit:

$$\binom{n}{k} = \frac{n^k}{k!} \cdot \underbrace{\frac{n(n-1)\cdots(n-k+1)}{n^k}}_{\to 1 \text{ für } n\to\infty} \sim \frac{n^k}{k!}$$

$$\Longrightarrow \mathrm{Bin}_{n,p_n}(k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} \sim \frac{n^k}{k!} p_n^k \frac{(1-\frac{np_n}{n})^n}{(1-p_n)^k}$$

$$\sim \frac{(n\cdot p_n)^k}{k!} \cdot \mathrm{e}^{-n\cdot p_n} \quad (\text{für } p_n \to 0) \sim \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$$

$\square$

Definition 2.16
*Seien $(\Omega, \mathcal{F})$ und $(\Omega', \mathcal{F}')$ meßbare Räume. Eine Abbildung $X : \Omega \to \Omega'$ heißt **Zufallsvariable**, falls gilt:*

$$X^{-1}(A') \in \mathcal{F} \quad \forall A' \in \mathcal{F}' \qquad (\text{"}X \text{ ist } \mathcal{F} - \mathcal{F}'\text{-meßbar"})$$

mit anderen Worten: $\quad X^{-1}(\mathcal{F}') \subset \mathcal{F}$

Warum?

Wollen berechen:

$$P\left[\{X \in A'\}\right] = P\left[\{\omega \in \Omega \mid X(\omega) \in A'\}\right] = P\underbrace{\left[X^{-1}(A')\right]}_{\in \mathcal{F}} = P \circ X^{-1}(A')$$

Beispiel
Sei $(\Omega, \mathcal{F}) = (\Omega, \mathcal{P}(\Omega))$, dann ist jede Abbildung $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ eine Zufallsvariable

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"** von X, bzw. die **"Verteilung"** von X und ist ein Wahrscheinlichkeitsmaß auf dem messbaren Raum $(\Omega', \mathcal{F}')$*

Beweis:
Übung! (Prüfe Normierung & σ-Additivität von $P_X : \mathcal{F}' \to [0, 1]$) □

Definition 2.18
Sei $\Omega \neq \emptyset$.

*$\mathcal{A} \subset \mathcal{P}(\Omega)$ heißt **Algebra**, falls*

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

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

(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* $\bigcup\limits_{n\in\mathbb{N}} A_n \in \mathcal{A} \implies \mu(\dot{\bigcup\limits_{n\in\mathbb{N}}} A_n) = \sum\limits_{n\in\mathbb{N}} \mu(A_n)$

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

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

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

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

Beispiel
auf $\Omega = \mathbb{R}$ ist

$$\Phi = \{ \dot{\bigcup}_{k=1}^{n} (a_k, b_k] \cap \mathbb{R} \mid n \in \mathbb{N},\ -\infty \le a_1 < b_1 < \cdots < b_n \le \infty \}$$

eine Algebra.

Definiere $\lambda(\dot{\bigcup}_{k=1}^{n} (a_k, b_k]) := \sum_{k=1}^{n} (b_k - a_k)$

Dann gilt:

$$\lambda(\dot{\bigcup}_{n=1}^{N} \underbrace{A_n}_{\in \Phi}) = \sum_{n=1}^{N} \lambda(A_n) \qquad \text{“Additivität”}$$

und wir werden zeigen, dass λ sogar σ-additiv ist, mithin ein Prämaß ist.

Wir nennen ein Mengensystem $\mathcal{A} \subset \mathcal{P}(\Omega)$

(i) “abgeschlossen unter wachsenden Limites”, falls $A_k \in \mathcal{A}$ 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. Differenzen”, falls $A, B \in \mathcal{A}$ 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 ($\mathcal{C} \subset \mathcal{P}(\Omega)$) abgeschlossen bzgl. endl. Durchschnitten und mit $\Omega \in \mathcal{C}$.

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

Bemerkung

1. $\mathcal{A}$ ist wohldefiniert!
2. Es gibt auch eine Version für Funktionensysteme, vgl. Jacod/Protter, Theorem 6.3.

Beweis:

- für $B \subset \Omega$:

  Sei $\mathcal{A}_B := \{A \in \mathcal{A} \mid A \cap B \in \mathcal{A}\}$

  $\implies \mathcal{A}_B$ ist abgeschlossen bzgl. wachsender Limites & Differenzen

- für $B \in \mathcal{C}$ gilt:

$$\begin{aligned} & B \cap C \in \mathcal{C} \subset \mathcal{A} \quad \forall C \in \mathcal{C} \\ \implies & C \in \mathcal{A}_B \quad \forall C \in \mathcal{C} \\ \implies & \mathcal{C} \subset \mathcal{A}_B \subset \mathcal{A} \\ \implies & \mathcal{A}_B = \mathcal{A} \end{aligned}$$

- für $B \in \mathcal{A}$ gilt:

$$\begin{aligned} & \text{für } C \in \mathcal{C} \text{ folgt: } B \in \mathcal{A}_C = \mathcal{A} \text{ und } B \cap C \in \mathcal{A} \\ \implies & C \in \mathcal{A}_B \\ \implies & \mathcal{C} \subset \mathcal{A}_B \subset \mathcal{A} \\ \implies & \mathcal{A}_B = \mathcal{A} \end{aligned}$$

$\implies \mathcal{A}$ ist abgeschlossen bzgl. (endl.) Durchschnitte

auch gilt: $\Omega \in \mathcal{C} \subset \mathcal{A} \quad \checkmark$

$\implies \mathcal{A}$ ist abgeschlossen bzgl. Komplemente ($A \in \mathcal{A},\ A^c = \Omega \setminus A \in \mathcal{A}$) $\checkmark$

bleibt zu zeigen: $\mathcal{A}$ ist abgeschlossen bzgl. abzählbarer Vereinigungen

Sei $A_k \in \mathcal{A},\ n \in \mathbb{N}$, dank Abgeschossenheit bzgl. wachsender Limites gilt:

$$\underbrace{\bigcup_{n=1}^{N} A_n}_{\in \mathcal{A}\ \forall N} \underset{N\to\infty}{\nearrow} \underbrace{\bigcup_{n=1}^{\infty} A_n}_{\implies \in \mathcal{A}} \in \mathcal{A}$$

$\implies \mathcal{A}$ ist σ-Algebra, $\mathcal{C} \subset \mathcal{A}$,

$\mathcal{A} \subset \sigma(\mathcal{C})$ klar $\implies \mathcal{A} = \sigma(\mathcal{C})$ □

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

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

Beweis:

O.B.d.A sei $\Omega \in \mathcal{C}$,

Definiere $\mathcal{A} := \{A \in \mathcal{F} \mid P(A) = Q(A)\}$

$\mathcal{A}$ ist abgeschlossen bzgl. wachsender Limites (wegen σ-Stetigkeit von Wahrscheinlichkeitsmaßen) und bzgl. Differenzen,

$\mathcal{C} \subset \mathcal{A}$ nach Voraussetzung

$$\implies \mathcal{A} = \sigma(\mathcal{C}) = \mathcal{F}$$

□

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

Beweis:

siehe Maß- & Integrationstheorie (z.B. Klenke Theorem 1.41 und 1.7) □

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, d.h. $\mu = \nu$ auf Φ ($\mu(A) = \nu(A) \quad \forall A \in \Phi$). 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:

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

Beweis:

Für $n \in \mathbb{N}$, sei $\mathcal{F}_n$ die Spur-σ-Algebra $\mathcal{F}_n := \{A_n \cap A \mid A \in \mathcal{F}\}$ von $\mathcal{F}$ auf A_n.

Betrachte die Einschränkungen von $\mu,\ \nu$ auf $\mathcal{F}_n$.

Analog zu Korollar 2.21 folgert man dann $\mu = \nu$ auf $\mathcal{F}_n$

$$\implies \mu(A) = \lim_{n\to\infty} \mu(A \cap A_n) = \lim_{n\to\infty} \nu(A \cap A_n) = \nu(A) \quad \forall A \in \mathcal{F} \text{ dank } \sigma\text{-Stetigkeit}$$

$$\implies \text{Maße stimmen auf ganz } \mathcal{F} \text{ überein}$$

□

Lemma 2.24

Sei μ ein Maß auf $(\mathbb{R}, \underbrace{\mathcal{B}(\mathbb{R})}_{\text{Borel-}\sigma\text{-Algebra}})$, endlich auf allen 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.

Bemerkung

μ endlich auf Kompakta impliziert wegen $[-n,n] \uparrow \mathbb{R}$, dass μ σ-endlich und $\mu(\emptyset) = 0$, da $\infty > \mu(A) = \mu(A) + \mu(\emptyset) \implies \mu(\emptyset) = 0$; dann ist $G: \mathbb{R} \to \mathbb{R}$ endlich $\forall x: G(x) < \infty$

Beweis:

- wachsend ✓
- rechtsstetig: z.B. für $x > 0,\ x_n \searrow x$ (d.h. $\lim_{n\to\infty} x_n = x$ und $x_n \geq x \quad \forall n \in \mathbb{N}$) ist

$$G(x_n) - G(x) = \mu(\underbrace{(x, x_n]}_{\text{fallend gegen } \emptyset}) \to \mu(\emptyset) = 0$$

  wegen $\bigcap_{n\in\mathbb{N}} (x, x_n] = \emptyset$ und σ-Stetigkeit von Maß μ

  Fälle $x = 0,\ x < 0$ analog ✓

□

Beispiel

z.B. für Lebesgue-Maß ist $G(x) = x$

Definition 2.25
Für ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ist seine (kumulative [engl "CDF" - cumulative distribution function]) ***Verteilungsfunktion*** *F gegeben durch*

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

Bemerkung
$F(x) = G(x) - G(-\infty)$

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

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

Beweis:
aus Lemma 2.24 und mit σ-Stetigkeit von P □

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 für μ eingeschränkt auf $[0, 1]$ die Gleichverteilung U auf $([0, 1], \mathcal{B}([0, 1]))$

Beweis:

Eindeutigkeit: $\{(a, b] \mid a, b \in \mathbb{R},\ a < b\}$ ist $\cap$-stabiler Erzeuger von $\mathcal{B}$ $\Longrightarrow$ Eindeutigkeit nach Lemma 2.23.

Existenz:

$$\Phi := \{\cup_{k=1}^{K} (a_k, b_k] \cap \mathbb{R} \mid K \in \mathbb{N},\ -\infty \leq a_1 < b_1 < a_2 < \cdots < b_k \leq \infty\}$$

Φ ist Algebra (Übung).

Für

$$A := \cup_{k=1}^{K} (a_k, b_k]$$

weisen wir zu:

$$\mu(A) = \sum_{k=1}^{K} (F(b_k) - F(a_k)),$$

dann ist μ additiv auf Φ.

Zu zeigen bleibt:

μ ist σ-additiv (also Prämaß auf Φ, und mit Fortsetzungssatz 2.22 v. Caratheodory folgt dann die Fortsetzung zu Maß $\tilde{\mu}$ auf $\mathcal{B} = \sigma(\Phi)$).

Dazu:

$$\begin{aligned} &\text{Für } A_n = \dot{\bigcup}_{k=1}^{K_n} (a_k^n, b_k^n], \text{ mit disjunkten } (A_n)_{n\in\mathbb{N}}, \\ &\text{und mit } A^\infty := \dot{\bigcup}_{n=1}^{\infty} A_n \overset{!}{\in} \Phi, \end{aligned}$$

also gilt:

$$A^\infty = \dot{\bigcup}_{k=1}^{K_\infty} (a_k^\infty, b_k^\infty] \quad \text{mit } K_\infty < \infty$$

zu zeigen:

$$\begin{aligned} \mu(A_\infty) &\overset{!}{=} \sum_{n=1}^{\infty} \mu(A_n) \\ \iff \sum_{k=1}^{K_\infty} \mu((a_k^\infty, b_k^\infty]) &\overset{!}{=} \sum_{n=1}^{\infty} \sum_{k=1}^{K_n} \underbrace{(F(b_k^n) - F(a_k^n))}_{\geq 0} \\ &= \sum_{k=1}^{K_\infty} \sum_{n=1}^{\infty} \sum_{\left(j\in\{1,\cdots,K_n\}:(a_j^n, b_j^n]\subset(a_k^\infty, b_k^\infty]\right)} \Big(F(b_j^n) - F(a_j^n)\Big) \end{aligned}$$

Es genügt also zu zeigen:
Für $(a^\infty, b^\infty] = \dot{\bigcup}_{n\in\mathbb{N}} (a^n, b^n]$ gilt:

$$\mu\left((a^\infty, b^\infty]\right) \overset{!}{=} \sum_{n\in\mathbb{N}} \mu\left((a^n, b^n]\right) = \sum_{n\in\mathbb{N}} F(b^n) - F(a^n) = \sum_{n\in\mathbb{N}} \mu\left(A_n\right)$$

hierzu:

"$\geq$": wegen $(a^\infty, b^\infty] \supset \dot{\bigcup}_{n=1}^{N}{}^{N}_{n=1} (a^n, b^n]$ gilt "$\geq$"

"$\leq$": Betrache

$$O_n := (a^n, b^n + \delta^n) \supset (a^n, b^n]$$

offen mit $\delta^n = \delta^n(\epsilon)$, so dass

$$\mu\left((a^n, b^n + \delta^n]\right) \leq \mu\left((a^n, b^n]\right) + \epsilon \cdot 2^{-n}$$

für $\epsilon > 0$ beliebig

$\implies$ $(O_n)_{n\in\mathbb{N}}$ offene Überdeckung von $\overset{\text{kompakt}}{[a^\infty + \delta^\infty, b^\infty]} \subset (a^\infty, b^\infty]$ mit $\delta^\infty = \delta^\infty(\epsilon)$, so dass

$$\mu\left((a^\infty, a^\infty + \delta^\infty]\right) < \epsilon$$

$\implies$ (Heine-Borel) Es existiert eine <u>endliche</u> Überdeckung: $\exists N < \infty$, so dass

$$\bigcup_{n=1}^{N} (a^n, b^n + \delta^n) \supset [a^\infty + \delta^\infty, b^\infty] \supset (a^\infty + \delta^\infty, b^\infty]$$

$$
\begin{aligned}
\implies \quad \mu\left((a^\infty, b^\infty]\right) &\overset{\text{additiv}}{=} \underbrace{\mu\left((a^\infty, a^\infty + \delta^\infty]\right)}_{<\epsilon} + \mu\left((a^\infty + \delta^\infty, b^\infty]\right) \\
&\leq \epsilon + \sum_{n=1}^{N} \underbrace{\mu\left((a^n, b^n + \delta^n]\right)}_{\leq \mu((a^n, b^n]) + \epsilon \cdot 2^{-n}} \\
&\leq 2\epsilon + \sum_{n=1}^{\cancel{N}\infty} \mu\left((a^n, b^n]\right)
\end{aligned}
$$

also gilt $\leq$, da $\epsilon > 0$ beliebig war.

$\square$

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

(i)

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

$$
F(x) \text{ entspricht also dem Maß } \mu(A) = \begin{cases} 1, & a \in A \\ 0, & \textit{sonst} \end{cases}
$$

Das Maß $\delta_x := \mu$ heißt das "Dirac'sche Punktmaß" auf $a \in \mathbb{R}$, Notation $\mu = \delta_a$

(ii) $\mu = p \cdot \delta_{a_1} + (q - p) \cdot \delta_{a_2} + (1 - q) \cdot \delta_{a_3}$ 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, hat eine **Dichte** (nämlich die Dichte ρ).*

***Absolutstetig** bedeutet:*

$$
F(x) = \overset{\textit{(Lebesgue-Integral)}}{\int_{-\infty}^{x}} \rho(y) \mathrm{d}y \qquad \forall x \in \mathbb{R}
$$

für eine meßbare Funktion $\rho : \ (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), \ \rho$ nicht negativ

(ii) Umgekehrt:

Gilt

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

für ρ nicht-negativ und messbar, mit

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

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).

Beispiele (für Wahrscheinlichkeitsmaße mit Dichten)

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

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

"Gauß'sche Glockenkurve/Gauß-Verteilung"

$\varphi \geq 0$, messbar, da stetig und $\int_{\mathbb{R}} \varphi(x) \mathrm{d}x = 1$,

denn Normierung ergibt

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

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

und

$$\begin{aligned}\left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \mathrm{d}y\right)^2 &= \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right) \mathrm{d}x\mathrm{d}y \\ &\overset{\text{Polarkoordinaten}}{=} \int_0^\infty \int_0^{2\pi} \frac{1}{2\pi} \cdot r \cdot \exp\left(-\frac{r^2}{2}\right) \mathrm{d}\theta \mathrm{d}r \\ &= \frac{1}{\cancel{2\pi}} \cdot \cancel{2\pi} \left[-\mathrm{e}^{-\frac{r^2}{2}}\right]_0^\infty \\ &= 1 \quad \checkmark\end{aligned}$$

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

mit Parameter $\alpha > 0$ hat Dichte

$$\rho_\alpha(x) = \gamma_{\alpha,1}(x) = \mathbb{1}_{(o,\infty)}(x) \cdot \alpha \cdot \mathrm{e}^{-\alpha x}$$

(iii) *Gammaverteilung*

Parameter $\alpha, r > 0$

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

mit

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

die sogenannte Gamma-Funktion ($\rightarrow$ Analysis),

Man zeigt: $\Gamma(r) = (r-1)!$ via $\Gamma(r) = (r-1) \cdot \Gamma(r-1)$ mit partieller Integration für $r \in \mathbb{N}$.

Motivation der Gammaverteilung:

Betrachte $r \in \mathbb{N}$.

Wie wahrscheinlich ist es, im Zeitintervall $[0, t]$ eine Anzahl von mindestens r Versicherungsschäden zu beobachten?

Standardmodell für Schadensanzahl: Poissonverteilung mit Parameter $\lambda = \alpha t$ (siehe z.B. Georgii).

$$\mathrm{Poiss}_{\alpha t}(k) = \mathrm{e}^{-\alpha t} \cdot \frac{(\alpha t)^k}{k!} \quad \alpha > 0$$

Ansatz also:

$$\begin{aligned} P[\text{“mindestens } r\text{-Schäden in } [0,t]\text{”}] &= 1 - \sum_{k=0}^{r-1} \mathrm{Poiss}_\lambda(k) \\ &= 1 - \sum_{k=0}^{r-1} \mathrm{e}^{-\alpha t} \cdot \frac{(\alpha t)^k}{k!} \\ &= \int_0^t \underbrace{\frac{\alpha^r}{(r-1)!} \cdot x^{r-1} \cdot \mathrm{e}^{-\alpha x}}_{=\gamma_{\alpha,r}(x)} \mathrm{d}x, \quad t \geq 0 \end{aligned}$$

(Begründung folgt)

denn Differenzieren ergibt:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left[1 - \sum_{k=0}^{r-1} \mathrm{e}^{-\alpha t} \cdot \frac{(\alpha t)^k}{k!}\right] &= \alpha \mathrm{e}^{-\alpha t} - \sum_{k=0}^{r-1} \left(-\alpha \frac{(\alpha t)^k}{k!} + \frac{\alpha^k t^{k-1}}{(k-1)!}\right) \cdot \mathrm{e}^{-\alpha t} \\ &= \alpha \frac{(\alpha t)^{r-1}}{(r-1)!} \cdot \mathrm{e}^{-\alpha t} \end{aligned}$$

Erinnerung an die Hauptresultate zum Maßintegral $\int_\Omega f(\omega)\mu(\mathrm{d}\omega)$

(lediglich Ausblick, genaueres siehe Maß- & Integrationstheorie, vgl. Klenke)

Betrachte messbaren Raum $(\Omega, \mathcal{F})$ mit Maß $\mu : \mathcal{F} \to [0,1]$ und erkläre “$\int_\Omega f \mathrm{d}\mu = \int_\Omega f(\omega)\mu(\mathrm{d}\omega)$” für $f : \Omega \to \mathbb{R}$ “messbar“ in 3 Schritten:

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

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

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

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

erkläre:

$$\int f \mathrm{d}\mu := \sup_{\substack{g \text{ elementar} \\ 0 \leq g \leq f}} \int g \mathrm{d}\mu$$

(iii) für f $\mathbb{R}$-wertig und messbar mit $\int |f| \mathrm{d}\mu < \infty$ erkläre:

$$\int f \mathrm{d}\mu := \underbrace{\int f^+ \mathrm{d}\mu}_{<\infty} - \underbrace{\int f^- \mathrm{d}\mu}_{<\infty} \quad \in \mathbb{R}$$

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

Solche f heißen "integrierbar“ (bzgl. μ);

Notation: $f \in \mathfrak{L}^1(\mu) \equiv \mathfrak{L}^1(\Omega, \mathcal{F}, \mu)$

Bemerkung

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

(i) **Lemma von Fatou**

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

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

(ii) **Satz über monotone Konvergenz**

Falls $f_n \geq 0 \quad \forall n \in \mathbb{N}$ und $f_1 \leq f_2 \leq \cdots$ (punktweise monoton)

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

(iii) **Satz von der majorisierten Konvergenz**

$(f_n)_{n\in\mathbb{N}}$ $\mathbb{R}$-wertig, und $\exists g \in \mathfrak{L}^1(\mu)$ (d.h. $\int |g| \mathrm{d}\mu < \infty$), so dass $|f_n| \leq g$ μ-fast überall $\forall n$

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

und falls $f_n \underset{n\to\infty}{\longrightarrow} f_\infty$ μ-fast überall gilt

$$\left(\text{d.h. } \mu(\{\omega \in \Omega \mid \lim_{n\to\infty} f_n(\omega) \neq f^\infty(\omega)\}) = 0\right)$$

$$\implies \int f_n \mathrm{d}\mu \underset{n\to\infty}{\longrightarrow} \int f_\infty \mathrm{d}\mu \text{ und } \int |f_n - f_\infty| \to 0$$

b) für $f : \mathbb{R} \to [0, \infty)$ Riemann-integrierbar auf $\mathbb{R}$ gilt, dass f Lebesgue-integrierbar ist und das Riemann-Integral gleich dem $\int_\mathbb{R} f \mathrm{d}\lambda$ ($\leftarrow$ Lebesgue-Integral bzgl. des Lebesgue-Maßes $\mu = \lambda$) ist.

Korollar 2.29

Sei $(\Omega, \mathcal{F}, \mu)$ Maßraum, $\mu : \mathcal{F} \to [0, \infty]$ Maß $\rho : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ messbar, $\rho \geq 0$ mit $\int_\Omega \rho \mathrm{d}\mu = 1$. Dann ist

$$P(A) := \int \mathbb{1}_A \cdot \rho \mathrm{d}\mu = \int_A \rho \mathrm{d}\mu \quad A \in \mathcal{F}$$

*ein Wahrscheinlichkeitsmaß (auf $\mathcal{F}$). Man nennt $\rho =$"$\frac{\mathrm{d}P}{\mathrm{d}\mu}$" die **Radon-Nykodym-Dichte** von P bzgl. μ und nennt P absolut-stetig bzgl. μ (mit Dichte ρ).*

$$\begin{aligned} \text{„}\frac{\mathrm{d}P}{\mathrm{d}\mu}\text{"}: \quad & \int_A \rho \mathrm{d}\mu && = \int_A \mathrm{d}P, && \forall A \in \mathcal{F} \\ \iff \quad & \int f \rho \mathrm{d}\mu && = \int f \mathrm{d}P && \forall f \geq 0,\ \Omega \to \mathbb{R} \textit{ messbar} \end{aligned}$$

Beweis:

1. $P(\Omega) = 1$,
2. $P(A^c) \overset{\text{Linearität von } \int f \mathrm{d}\mu}{=} 1 - P(A)$ klar
3. σ-additiv:
   Betrachte $(A_k)_{k\in\mathbb{N}}$ eine disjunkte Folge in $\mathcal{F}$.

$$
\begin{aligned}
P\left(\bigcup_{k\in\mathbb{N}} A_k\right) &= \int_\Omega \underbrace{\left(\sum_{k\in\mathbb{N}} \mathbb{1}_{A_k}\right)}_{\mathbb{1}_{\{\bigcup_{k\in\mathbb{N}} A_k\}}} \rho \mathrm{d}\mu \\
&\overset{\text{monotone Konvergenz}}{=} \lim_{N\to\infty} \underbrace{\int \sum_{k=1}^{N} (\mathbb{1}_{A_k}) \rho \mathrm{d}\mu}_{\sum_{k=1}^{N} \underbrace{\int \mathbb{1}_{A_k} \rho \mathrm{d}\mu}_{P(A_k)} \text{ nach Linearität}} \\
&= \lim_{N\to\infty} \sum_{k=1}^{N} P(A_k) \\
&= \sum_{k\in\mathbb{N}} P(A_k)
\end{aligned}
$$

□

Beispiel

(i) Ω abzählbar, $\mathcal{F} = \mathcal{P}(\Omega) \equiv 2^\Omega$

$$\mu = \sum_{\omega\in\Omega} \delta_\omega \implies \mu(A) = |A|, \quad \mathcal{A} \subset \Omega$$

mit δ_ω das Dirac'sche Punktmaß auf ω.

Für $\rho : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ messbar, $\rho \geq 0$, $\int \rho \mathrm{d}\mu = \sum_{\omega\in\Omega} \rho(\omega) = 1 \implies$ Zähldichte ρ definiert Wahrscheinlichkeitsmaß $P(A) = \int_A \rho \mathrm{d}\mu$

(ii) $(\Omega, \mathcal{F}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $\mu = \lambda$ Lebesguemaß, ρ messbar, $\rho \geq 0$ und $\int_{\mathbb{R}^d} \rho(x)\lambda(\mathrm{d}x) = 1$

$$\implies P(A) := \int_A \rho \mathrm{d}\lambda = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \mathbb{1}_A(x)\rho(x)\mathrm{d}x_1 \cdots \mathrm{d}x_d, \quad A \in \mathcal{B}(\mathbb{R}^d), x = (x_1, \cdots, x_n) \in \mathbb{R}^d$$

ist Wahrscheinlichkeitsmaß, d.h. P ist absolutstetig bzgl. λ.

Beispiel konkrete Beispiele

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

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

b) auch stetige Verteilungsfunktionen F auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ brauchen keine Dichte haben. Betrachte F ist stetige Verteilungsfunktion, die nur "gleichverteilt" auf Cantormenge (Lebesgue-Nullmenge!) wächst, d.h. auf 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
$(\Omega, \mathcal{F}, P)$ Maßraum, $A, B \in \mathcal{F}, P[B] > 0$,
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*

Beispiel

1. (vgl. oben)
   $\Omega = \{g, k\} \times \{p, n\}, \mathcal{F} = P(\Omega)$ und P gegeben durch Zähldichte $\rho : \Omega \to [0, 1]$

   mit $\rho((g, p)) = \frac{20}{1000}, \rho(g, n)) = \frac{20}{1000}, \ldots$

   $A =$ "Patient krank "$= \{\omega = (\omega_1, \omega_2) \in \Omega \mid \omega_1 = k\}$
   $B =$ "Test positiv "$= \{\omega = (\omega_1, \omega_2) \in \Omega \mid \omega_2 = p\}$

   $\Rightarrow P[A|B] = \frac{9}{29}$
2. $(\Omega, \mathcal{F}, P) = ([0, 1], \mathcal{B}[0, 1], \mathcal{U}_{[0,1]})$ mit $\mathcal{U}$ die Gleichverteilung, d.h. Lebegues Maß
   $A = [0, \frac{2}{3}], B = [0, \frac{1}{2}]$

   $\Rightarrow P[A|B] = 1$

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

(i) $Q(A) := P[A|B]$, $\quad A \in \mathcal{F}$, ist Wahrscheinlichkeitsmaß auf $\mathcal{F}$

*(ii) **Formel totaler Wahrscheinlichkeit**:*
Sei $\mathcal{B} = \dot{\bigcup}_{i \in I} B_i$ $\quad$ für $B_i \in \mathcal{F}$ disjunkt, dann gilt:

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

(iii) $A, (B_i)_{i \in I} \in \mathcal{F}$ mit strikt positiven Wahrscheinlichkeiten, I abzählbar, $B_i \cap B_j = \emptyset$, $\quad \forall i \neq j \in I$, und $\Omega = \dot{\bigcup}_{i \in I} B_i$ ("Partition von Ω"), dann folgt:

$$P[B_j|A] = \frac{P[B_j] \cdot P[A|B_j]}{\sum_{i \in I} P[B_i] \cdot P[A|B_i]} \qquad \textbf{(BAYES-FORMEL)}$$

Bemerkung (zu bedingten Wahrscheinlichkeiten:)

(i) Interpretation: frequentisch oder subjektiv

(ii) Achtung: Definition ist statistisch, aber sagt nichts über Kausalitäten!

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

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

Beweis:
per Induktion über n (Übung!) $\square$

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

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

b) ***eine Familie*** *$(A_i)_{i\in I} \in \mathcal{F}$ (I beliebige Indexmenge)* ***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*** *$\mathcal{A}_i \subset \mathcal{F}, \quad i \in I$ (I beliebige Indexmenge), 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*** *$Y_i : (\Omega, \mathcal{F}) \to (\Omega_i, \mathcal{F}_i), \quad i \in I$ (I beliebige Indexmenge), heißt unabhängig,*
falls die σ-Algebren $\sigma(Y_i) = Y_i^{-1}(\mathcal{F}_i), \quad i \in I$ unabhängig sind.

Bemerkung (zur Definition)

(a) sind $(A_i)_{i\in I}$ unabhängig, dann $\Rightarrow$ A_i und A_j sind paarweise unabhängig $\forall i \neq j \in I$. Die Umkehrung 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

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

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

Beweis:
zu zeigen ist: (1) gilt $\forall B_i \in \mathcal{F}_i = \sigma(\mathcal{E}_i)$
Induktion über $\mid \{i \in J | B_i \notin \mathcal{E}_i\} \mid = n$
IA: $(n = 0)$
IS: $(n) \to (n+1) \quad$ OBdA: $J = \{1, \ldots, n+1\}, |J| = n+1, \quad B_{n+1} \in \mathcal{F}_{n+1} \setminus \mathcal{E}_{n+1}$
definiere: $A := \bigcap_{i=1}^n Y_i^{-1}(B_i)$
OBdA gelte: $P[A] > 0$ (sonst Behauptung klar)
definiere: (Wahrscheinlichkeitsmaße auf $\mathcal{F}_{n+1}$)

$$P_{n+1} = P \circ Y_{n+1}^{-1}$$
$$Q_{n+1} = P[\cdot|A] \circ Y_{n+1}^{-1}$$

$\Rightarrow P_{n+1}(B_{n+1}) \equiv P(Y_{n+1} \in B_{n+1}) = Q_{n+1}(B_{n+1}) \quad \forall B_{n+1} \in \mathcal{E}_{n+1}$
denn

$$\begin{aligned} Q_{n+1}(B_{n+1}) &= P[Y_{n+1} \in B_{n+1} | Y_i \in B_i, i = 1, \dots, n] \\ &= \frac{\prod_{i=1}^{n+1} P[Y_i \in B_i]}{\prod_{i=1}^{n} P[Y_i \in B_i]} \\ &= P[Y_{n+1} \in B_{n+1}] \end{aligned}$$

$\Rightarrow P_{n+1} = Q_{n+1}$ auf $\mathcal{F}_{n+1}$, weil "= " auf $\cap$-stabilem Erzeuger
$\Rightarrow$

$$\begin{aligned} P(\bigcap_{i\in J} Y_i^{-1}(B_i)) &= Q_{n+1}(B_{n+1}) P(A) \\ &= P_{n+1}(B_{n+1}) \cdot P(A) \\ &= \prod_{i=1}^{n+1} P(Y_i^{-1}(B_i)) \\ &= \prod_{i=1}^{n+1} P(Y_i \in B_i), \quad \forall B_i \in \mathcal{F}_i, i \in J \end{aligned}$$

□

Korollar 3.35
$(Y_i)_{i \in \{1,\dots,n\}}$ 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, $\mathcal{F}_i = \mathcal{P}(\Omega_i)\ \forall i$, so sind $(Y_i)_{i=1,\dots,n}$ unabhängig genau dann, wenn $\underbrace{P[Y_i = \omega_i, i = 1, \dots, n]}_{P \circ Y^{-1}(\{(\omega_1,\dots,\omega_n)\})} = \prod_{i=1}^{n} \underbrace{P[Y_i = \omega_i]}_{P \circ Y_i^{-1}(\{w_i\})} \quad \forall w_i \in \Omega_i$

(ii) reellwertiger Fall:

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

(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 $\rho_i(\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ sind $\forall i = 1, \dots, n$. Dann gilt: $(Y_i)_{i=1,\dots,n}$ unabhängig genau dann, wenn $Y = (Y^1, \dots Y^n) : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}^n(\mathbb{R}^n))$ eine absolutstetige Verteilung $P \circ Y^{-1}$ auf $(\mathbb{R}^n, \mathcal{B}^n(\mathbb{R}^n))$ mit Dichte $\rho_Y(y) = \rho_Y(y_1, \dots, y_n) = \prod_{i=1}^{n} \rho_i(y_i)$ hat.

Beweis:

(i) „ $\Rightarrow$ “ aus Definition von Unabhängigkeit ✓

„ $\Leftarrow$ “$\mathcal{E}_i = \{\{\omega_i\} : \omega_i \in \Omega_i\}$ bildet (mit $\emptyset$) einen $\cap$-stabilen Erzeuger von $\mathcal{F}_i$, also liefert Satz 3.34 die Behauptung.

(ii) „ $\Rightarrow$ “ ✓

„ $\Leftarrow$ “$\mathcal{E}_i = \{(\infty, c_i] : c_i \in \mathbb{R}\}$ ist $\cap$-stabiler Erzeuger von $\mathcal{F}_i = \mathcal{B}(\mathbb{R})$
$\Rightarrow$ Satz 3.34 liefert die Behauptung.

(iii) „ $\Rightarrow$ “ Für $A_i \in \mathcal{F}_i$ oder $A_i = (a^i, b^i]$ gilt:

$$\begin{aligned}
P[Y_i \in \mathcal{A}_i,\ \ i = 1, \dots, n] &= \prod_{i=1}^n P[Y_i \in \mathcal{A}_i] \\
&= \prod_{i=1}^n \left(\int_{\Omega_i} \mathbb{1}_{A_i} \rho_i(y_i) \mathrm{d}y_i \right) \\
&= \int_{\Omega_1} \dots \int_{\Omega_n} \underbrace{\prod_{i=1}^n \mathbb{1}_{A_i}(y_i)}_{\mathbb{1}_{A_1 \times \dots \times A_n}(y_1, \dots, y_n)} \prod_{i=1}^n \rho_i(y_i) \mathrm{d}y_n \dots dy_1 \\
&= \int_{\Omega_1 \times \dots \times \Omega_n} \mathbb{1}_{A_1 \times \dots \times A_n}(y) \rho(y) \mathrm{d}y
\end{aligned}$$

Da Mengen des Typs $A_1 \times \dots \times A_n$ einen $\cap$-stabilen Erzeuger bilden, folgt, dass ρ in der Tat gemeinsame Dichte von $Y = (Y^1, \dots, Y^n)$ ist.

„ $\Leftarrow$ “ $\mathcal{E}_i = \{(\infty, c_i] : c_i \in \mathbb{R}\}$ ist $\cap$-stabiler Erzeuger. Ähnliche Rechnung wie oben (Übung) zeigt zusammen mit Satz 3.34 die Unabhängigkeit der $Y^1, \dots, Y^n$.

□

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

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

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 $Y_i = (\Omega, \mathcal{F}) \to (\Omega_i, \mathcal{F}_i)$ von Zufallsvariablen ist die Verteilung $P \circ Y^{-1}$ der Zufallsvariablen $Y = (Y_i)_{i \in I} : (\Omega, \mathcal{F}) \to (\times_{i \in I} \Omega_i, \otimes_{i \in I} \mathcal{F}_i)$
2. Umgekehrt: eine mehrdimensionale Zufallsvariable Y induziert eine Verteilung $P \circ Y^{-1}$ auf dem Produktraum $(\times_{i \in I} \Omega_i, \otimes_{i \in I} \mathcal{F}_i)$
3. Die gemeinsame Verteilung induziert die Randverteilung (Einzelverteilung der Y_i) via

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

Satz 3.36
$(\Omega_i, \mathcal{F}_i, P_i), i \in I$ 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:

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

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

Beweis:
$\rightarrow$ Maßtheorie (allgemein), siehe auch im Georgii Satz 3.26 für konstruktiven Beweis für wichtige Fälle abzählbarer Produkte. □

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

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

d.h. die P_i sind die Randverteilung der Koordinatenprojektion π_i, welche Zufallsvariablen auf dem Raum $(\Omega, \mathcal{F}, P)$ sind!

Korollar 3.37
zu gegebenen Wahrscheinlichkeitsmaßen P_i auf $(\Omega_i, \mathcal{F}_i)_{i\in I}$ existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit unabhängigen Zufallsvariablen $(X_i)_{i\in I}$, so dass $P \circ X_i^{-1} = P_i$ gilt.

Beweis:
Wähle "kanonischen Produktraum"
$\Omega = \times_{i\in I}\Omega_i, \quad \mathcal{F} = \otimes_{i\in I}\mathcal{F}_i, \quad P = \otimes_{i\in I}P_i$ und $X_i(\omega) = \pi_i(\omega) = \omega_i$ □

Beispiel

a) $|I| = 2, (\Omega_i, \mathcal{F}_i), \quad i = 1,2$ diskret mit Zähldichte ρ_i auf Ω_i, dann ist $\rho(y_1, y_2) := \rho_1(y_1) \cdot \rho_2(y_2)$ Zähldichte auf $\mathcal{P}(\Omega_1) \otimes \mathcal{P}(\Omega_2) = \mathcal{P}(\Omega_1 \times \Omega_2)$

(in der Tat: $\rho \geq 0, \sum\limits_{y_1}\sum\limits_{y_2}\rho(y_1, y_2) = \sum\limits_{y_1}(\rho_1(y_1)\sum\limits_{y_2}\rho_2(y_2)) = \sum\limits_{y_1}\rho_1(y_1) = 1$)
und das zugehörige Maß P ist das Produktmaß $P_1 \otimes P_2$, denn

$$\begin{aligned} P(y_1, y_2) &= \rho(y_1, y_2) \\ &= \rho_1(y_1)\rho_2(y_2) \\ &= \rho_1(\{y_1\})\rho_2(\{y_2\}) \end{aligned}$$

b) $|I| = 2, (\Omega_i, \mathcal{F}_i) = (\mathbb{R}, \mathcal{B}(\mathbb{R})), i = 1,2$ und P_i mit Dichte ρ_i, dann ist $\rho(y) \equiv \rho(y_1, y_2)$ die Dichte des Produktmaßes $P = P_1 \otimes P_2$

Bemerkung

- Wegen Korollar 37 gilt: Zufallsvariable sind unabhängig genau dann, wenn ihre gemeinsame Verteilung durch das Produktmaß der Einzelverteilungen gegeben ist. Das impliziert, dass die Randverteilungen gerade die Verteilung der einzelnen Zufallsvariablen sind.
- Grundidee der bedingten Verteilungen: Auffassen einer gemeinsamen Verteilung als mehrstufiges Experiment, d.h. zunächst eine Koordinate ziehen, dann die zweite, etc. ...

Bedingte Verteilungen aus gemeinsamen Verteilungen

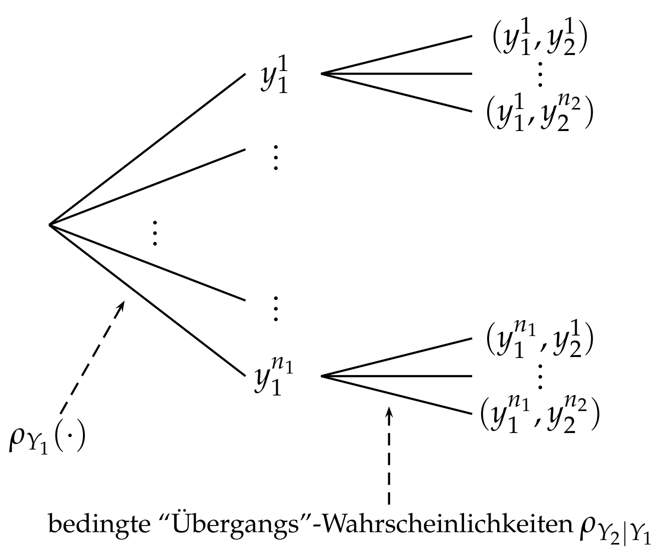
a) Diskreter Fall:
$\Omega = \Omega_1 \times \Omega_2, \mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2, Y = (Y_1, Y_2)$ diskrete Zufallsvariable mit Zähldichte $\rho(y_1, y_2)$. Die bedingte Verteilung von Y_2 gegeben $Y_1 = y_1$ ist beschrieben durch die bedingte Zähldichte:

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

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

Visualisierungen

i) "mehrstufiges Experiment":



ii) Kontingenztafel, z.B.

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

bedingte Verteilungen durch Re-normalisierung (?) auf gegebenen Spalten wie im Beispiel zu bedingten Wahrscheinlichkeiten

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

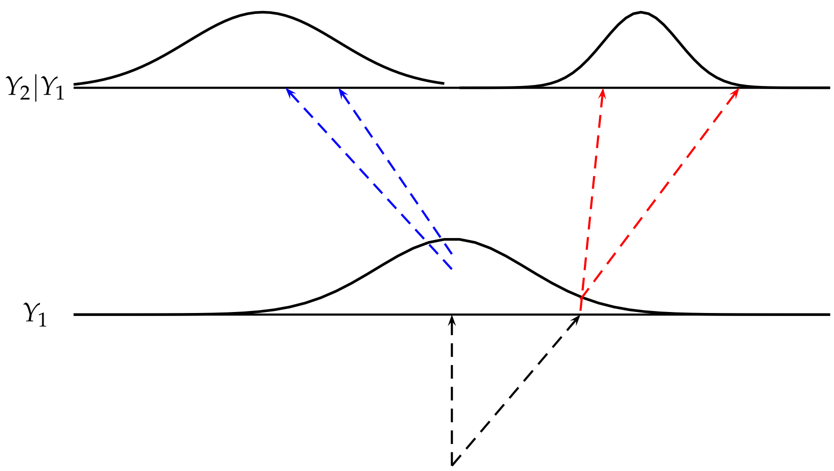
$$\rho_{Y_2|Y_1}(y_2|y_1) := \frac{\rho(y_1,y_2)}{\int_{\Omega_2} \rho(y_1,y_2)\mathrm{d}y_2} \equiv \frac{\rho(y_1,y_2)}{\rho_{Y_1}(y_1)}, \quad y_1 \in \Omega_1,\ y_2 \in \Omega_2,\ \rho_{Y_1}(y_1) > 0\ (:= 0 \text{ sonst})$$

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

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

Visualisierung/Veranschaulichung:

i) mehrstufiges Experiment



c) allgemeiner Fall (Ausblick auf Stochastik 2 und Maßtheorie)
Sei Y_2 Zufallsvariable, die Werte im "polnischem Raum"(separabler, vollständiger metrischer Raum mit Borel'scher σ-Algebra) annimmt, z.B. $(\mathbb{R}^n, \mathcal{B}^n)$, dann existiert ein stochastischer Kern (oder Markov- Übergangskern) K,
d.h. $K : \Omega_1 \times \mathcal{F}_2 \to [0,1]$ mit

(i) $y_1 \mapsto (y_1, A_2)$ ist $\mathcal{F}$-messbar $\forall A_2 \in \mathcal{F}_2$

(ii) $A_2 \mapsto K(y_1, A_2)$ ist Wahrscheinlichkeitsmaß auf $\mathcal{F}_2$, $\forall y_1 \in \Omega_1$

so dass gilt $P_{Y_1,Y_2} = “P_{Y_1} \otimes K''$, d.h. $\forall A \in \mathcal{F}_1 \otimes \mathcal{F}_2$ gilt

$$P[(Y_1,Y_2) \in A] = \int_{\Omega_1 \times \Omega_2} \mathbb{1}_A P_{Y_1,Y_2}(\mathrm{d}y_1, \mathrm{d}y_2) = \int_{\Omega_1} \left(\int_{\Omega_2} \mathbb{1}_{A_{y_1}}(y_2) K(y_1, \mathrm{d}y_2) \right) P_{Y_1}(\mathrm{d}y_1)$$

mit Sektion $A_{y_1} := \{y_2 \in \Omega_2 | (y_1, y_2) \in A\}$, $A_{y_1} \in \mathcal{F}$ falls $A \in \otimes_{i=1}^2 \mathcal{F}_i$, z.B. für $A = A_1 \times A_2$ gilt

$$A_{y_1} = \begin{cases} A_2 & \text{für } y_1 \in A_1 \\ \emptyset & \text{sonst} \end{cases}$$

$$P[Y_1 \in A_1, Y_2 \in A_2] = \int_{\Omega_1} \mathbb{1}_{A_1}(y_1) \left(\int_{\Omega_2} \mathbb{1}_{A_2}(y_2) K(y_1, \mathrm{d}y_2) \right) P_{Y_1}(\mathrm{d}y_1)$$

Bemerkung
Analoge Aussagen für 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 $I = \dot{\bigcup}_{k\in K} I_k$, 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 $\tilde{Y}_k := \varphi((Y_i)_{i\in I_k})$, $k \in K$ unabhängig.

Beweis:

- $\hat{Y}_k = (Y_i)_{i\in I_k}$ Zufallsvariablen, $\hat{Y}_k : (\Omega, \mathcal{F}) \to (\bigtimes_{I_k} \Omega_i, \bigotimes_{I_k} \mathcal{F}_i) =: (\hat{\Omega}_k, \hat{\mathcal{F}}_k)$
- definiere Projektionen $X_{k,i} : \hat{\Omega}_k \to \Omega_i$, $(\omega_j)_{j\in I_k} \mapsto \omega_i$ für $i \in I_k$,
  $\hat{\mathcal{E}}_k := \{\cap_{j\in J} X_{k,j}^{-1}(B_j) \mid \text{endliche } J \subset I_K, B_j \in \mathcal{F}_j, j \in J\}$ sind $\cap$-stabile Erzeuger von $\bigotimes_{i\in I_k} \mathcal{F}_i = \sigma(X_{k,j} \mid j \in I_k)$
- $(\hat{Y}_k)_{k\in K}$ sind unabhängige Zufallsvariablen. Genügt zu zeigen: ist für endliche $L \subset K$, $\hat{B}_k = \bigcap_{j\in J_k} X_{k,j}^{-1}(B_j)$, $(B_j \in \mathcal{E})$ für $k \in L$, so gilt für endliches $J_k \subset I_k$:

$$\begin{aligned}
P\left[\bigcap_{k\in L} \hat{Y}_k^{-1}(\hat{B}_k)\right] &= P\left[\bigcap_{k\in L} \hat{Y}_k^{-1}\left(\bigcap_{j\in J_k} X_{k,j}^{-1}(B_j)\right)\right] \\
&= P\left[\bigcap_{k\in L}\bigcap_{j\in J_k} Y_k^{-1} \circ X_{k,j}^{-1}(b_j)\right] \\
&\underset{(X_{k,j}\circ\hat{Y}_k = Y_j)}{=} P\left[\bigcap_{k\in L}\bigcap_{j\in J_k} \underbrace{(X_{k,j}\circ\hat{Y}_k)^{-1}}_{=(Y_j)^{-1}}(B_j)\right] \\
&\underset{(Y_j \text{ unabhängig})}{=} \prod_{k\in L}\prod_{j\in J_k} P\left[Y_j^{-1}(B_j)\right] \\
&= \prod_{k\in L} P \underbrace{\left[\bigcap_{j\in J_k} Y_j^{-1}(B_j)\right]}_{=\hat{Y}_k^{-1}\underbrace{\left(\bigcap_{j\in J_k} X_{k,j}^{-1}(B_j)\right)}_{=B_k}}
\end{aligned}$$

- also sind auch $(\tilde{Y}_k)_{k\in K}$ unabhängig, denn für $L \subset K$ endlich, $\tilde{B}_k \in \tilde{\mathcal{F}}_k \;\; \forall k \in L$ gilt:

$$\begin{aligned}
P[\bigcap_{k\in L} \underbrace{\tilde{Y}_k^{-1}}_{=(\varphi_k\circ\hat{Y}_k)^{-1}}(\tilde{B}_k)] &= P[\bigcap_{k\in L} \hat{Y}_k^{-1}(\underbrace{\varphi_k^{-1}(\tilde{B}_k)}_{\hat{\mathcal{F}}_k})] \\
&= \prod_{k\in L} P[\hat{Y}_k^{-1}(\varphi_k^{-1}(\tilde{B}_k))] \\
&= \prod_{k\in L} P[\tilde{Y}_k^{-1}(\tilde{B}_k)],
\end{aligned}$$

  weil $\tilde{Y}_k = \varphi_k \circ \hat{Y}_k$ ist.

□

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*

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

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
$\mathcal{A}((Y_k)_{k\in\mathbb{N}})$ ist eine σ-Algebra (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:
Fixiere $n \in \mathbb{N}$, definiere $\pi_i : X_{k\geq n} \to \mathbb{R},\ (\omega)_{k\geq n} \mapsto \omega_i,\ i \geq n,$
$B_n = \left\{ (y_{n+i})_{i\in\mathbb{N}} \in \bigtimes_{k\geq n} \Omega_k \mid \lim\limits_{N\to\infty} \frac{1}{N} \sum\limits_{i=1}^{N} Y_{n+i} \text{ existiert in } [a,b] \right\}$ ist messbar in $\bigotimes_{k\geq n} \mathcal{F}_k$,
$A = ((Y_k)_{k\geq n})^{-1}(B_n)$

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.

Beweis:
Projektion $\pi_i : \bigtimes_{k\in\mathbb{N}} \Omega_k \to \Omega_i,\ (\omega_k)_{k\in\mathbb{N}} \mapsto \omega_i,$
$\mathcal{E} := \{\bigcap_{i=1}^{n} \pi^{-1}(A_i) | A_i \in \mathcal{F},\ i = 1, ..., n;\ n \in \mathbb{N}\}$ ist $\cap$-stabiler Erzeuger, erzeugt $\bigotimes_{\mathbb{N}} \mathcal{F} = \sigma(\pi_i | i \in \mathbb{N})$,
für jedes n $\exists B_n \in \bigotimes_{k\geq n} \mathcal{F}_k$, so dass $A = ((Y_k)_{k\geq n})^{-1}(B_n) = \{(Y_k)_{k\geq n} \in B_n\}$.
Dann ist A unabhängig von $((Y_k)_{k\in\mathbb{N}})^{-1}(E)$ für $E \in \mathcal{E}$, denn

$$(Y_k)_{k\in\mathbb{N}}(E) = \{Y_i \in A_i | i = 1, ..., n\} = \bigcap_{i=1}^{n} Y_i^{-1}(A_i)$$

und

$$A = \{(Y_k)_{k\geq n+1} \in B_{n+1}\}$$

sind unabhängig nach Satz 3.38.

$\Longrightarrow$ A auch unabhängig von $(Y_k)_{k\in\mathbb{N}}^{-1}(E)$ für $E \in \bigotimes_{k\in\mathbb{N}} \mathcal{F}_k$

$\Longrightarrow$ A ist unabhängig von A selbst $\Rightarrow P(A) = P(A \cap A) = P(A)P(A)$

$\Longrightarrow$ $P(A) = 0$ oder $P(A) = 1$

$\square$

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

(i)

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

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

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

Beweis:

(i) $A \subset \bigcup_{k\geq n} A_k \ \forall n \implies P(A) \leq P(\bigcup_{k\geq n} A_k) \leq \sum_{k\geq n} P(A_k) \stackrel{n\to\infty}{\longrightarrow} 0$

(ii) $A^c = \bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n} A_k^c$ nach de Morgan

$$\begin{aligned} P(A^c) &\leq \sum_{n\in\mathbb{N}} P(\bigcap_{k\geq n} A_k^c) = \sum_{k\in\mathbb{N}} \lim_{m\to\infty} P(\bigcap_{k=n}^{m} A_k^c) \\ &= \sum_{k\in\mathbb{N}} \lim_{m\to\infty} \prod_{k=n}^{m} \underbrace{P(A_k^c)}_{1-P(A_k)\leq e^{-P(A_k)}} \\ &\leq \sum_{n\in\mathbb{N}} \lim_{m\to\infty} \exp(-\sum_{k=n}^{m} P(A_k)) = 0 \end{aligned}$$

$\square$

Beispiel (Monkey Typing Typewriter)
Seien $X_1, X_2 \ldots$ unabhängige Laplace Zufallsvariablen auf $\{A \ldots Z\} \cup \{\text{Satzzeichen}\}$ dann ist die Wahrscheinlickeit, dass irgendwann ein beliebiges Wort oder auch Goethes Faust kommt gleich eins.

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

- $L^p(P) \equiv L^p(\Omega, \mathcal{F}, P) = \{X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}) \mid \int |X|^p dP < \infty\}$ für $p \in [1, \infty)$ ist Banachraum mit Norm $\|X\|_p = (\int |X|^p dP)^{1/p}$, für $p = 2$ sogar Hilbertraum mit $\langle X_1, X_2 \rangle = \int X_1 X_2 \, dP$
(bei Identifikation von P-fast-überall gleichen Funktionen bzw. bei Betrachtung entsprechender Äquivalenzklassen)
- Man sagt oft auch "Mittelwert"von X bezüglich P statt Erwartungswert
- Wichtige Eigenschaften des Erwartungswertes folgen aus Eigenschaften des Maßintegrals $\int \cdot \, dP$, etwa Linearität, Monotonie und die Konvergenzsätze

Beispiel
X Zufallsvariable auf diskretem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F})$ mit Ω endlich (oder Ω abzählbar und $X \in L^1$):

$$E[X] = \sum_{\omega \in \Omega} X(\omega) * P(\omega)$$

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:

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

Beweis:
für $f = \mathbb{1}_A, A \in \mathcal{F}'$ gilt:

$$\int_\Omega \mathbb{1}_A(X) dP = P[X \in A] = P_X(A)$$

$\Rightarrow$ Behauptung gilt für elementare Funktionen

$$f = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}, \quad A_i \in \mathcal{F}', \alpha_k \in \mathbb{R}$$

$\Rightarrow$ weil jede meßbare Funktion $f \geq 0$ monoton approximierbar ist durch elementare Funktionen f_k, d.h. $\exists f_n$ elementare Funktionen mit : $0 \leq f_n \leq f$ und $f_n \nearrow f$ (punktweise, monotone Konvergenz), also folgt Behauptung für $f \geq 0$ mittels monotoner Konvergenz, für $f \in L^1(P_X)$ folgt Behauptung, dann via $f = f^+ - f^-$, beziehungsweise alternativ durch Approximation von f durch elementare Funktionen f_n mit $0 \leq |f_n| \leq |f|$ und $f_n \to f$ punktweise mit majorisierter Konvergenz. $\square$

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:

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

Beweis:
klar für $f = \mathbb{1}_A, \quad A \in \mathcal{F}'$, folgt also analog wie vorher für alle f mit entsprechenden Voraussetzungen

$$\int_{\Omega'} \mathbb{1}_A P_X = \int \mathbb{1}_A \cdot \rho(x)\mathrm{d}x \tag{1}$$

□

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

a) ***Markov'sche 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 auch die ***Tschebyschev-Ungleichung*** *(auch Chebyshev-Ungleichung)*

c) ***Exponentielle Markov Ungleichung***

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

d) ***Cauchy-Schwarz'sche 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) ***Hölder'sche 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:

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

Lemma 5.46 (Jensen'sche Ungleichung)
X reelle Zufallsvariable, $f : \mathbb{R} \to \mathbb{R}$ konvexe Funktion. $X,\ f(x)$ in $\mathrm{L}^1(P)$. Dann gilt: $f(E[X]) \leq E(f(x))$

Beweis:
f konvex $\implies f(x) = \sup_y(\alpha_y x + \beta_y),\quad x \in \mathbb{R}$, ist (punktweises) Supremum über affine Funktionen mit $\alpha_y, \beta_y \in \mathbb{R}$
$\implies E[f(X)] \geq \sup_y(\alpha_y E[X] + \beta_y) = f(E(X))$ □

Satz 5.47
Seien $X, Y \in L^2(P)$ unabhängig. Dann gilt:

$$E[XY] = (EX)(EY)$$

Beweis:

$$\begin{aligned}
E[XY] &= \int_\Omega XYdP = \int_{\mathbb{R}^2} XYP_{(X,Y)}(\mathrm{d}x, \mathrm{d}y) \\
&= \int_{\mathbb{R}^2} XYP_X \otimes P_Y \overset{Fubini}{=} \int_{\mathbb{R}} \left(X \int_{\mathbb{R}} YP_Y(\mathrm{d}y)\right) P_X(\mathrm{d}x) \\
&= \int XE(Y)P_X(\mathrm{d}x) = E(X)E(Y)
\end{aligned}$$

□

Bemerkung
Für eine $\mathbb{R}^n$-wertige Zufallsvariable X ist $E[X] = (E[X^i])_{i=1\ldots n}$ koordinatenweise definiert.

Definition 5.48
Für $X, Y \in L^2(P)$ heißt

*(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) $\operatorname{Cov}(X, Y) := E[(X - E[X])(Y - E[Y])] = E[XY] - (E[X])(E[Y])$ heißt die **Kovarianz** von X und Y.*

*(c) Falls $\operatorname{Cov}(X, Y) = 0$ 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) $\operatorname{Cov}(aX + b, cY + d) = ac \operatorname{Cov}(X, Y)$ insbesondere $V(aX + b) = a^2 V(X)$

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

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

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

Bemerkung

1. X $\mathbb{R}$-wertige Zufallsvariable, $X \in L^2(P)$, mit $V(X) > 0$, dann heißt
$$\tilde{X} := \frac{X - E[X]}{\sqrt{V(X)}}$$
standardisiert ($E[\tilde{X}] = 0, V(X) = 1$)
2. Für X $\mathbb{R}^n$-wertig, $X \in L^2(P)$, d.h. $X^i \in \mathrm{L}^2$ $\forall i$ Dann ist
$$\left(\mathrm{Cov}(X^i, X^j)\right)_{i,j}, \quad 1 \leq i, j \leq n$$
die Varianz-/ Kovarianzmatrix von X, kurz "Kovarianzmatrix"

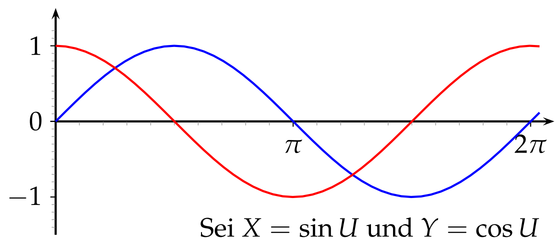
Beispiel (für Varianzberechnung)

1. $X_1, \cdots, X_n$ i.i.d. (independent, identically distributed) Bernoulli(p) Zufallsvariablen
$$\implies X = \sum_{i=1}^{n} X_i \sim \mathrm{Bin}_{(n,p)}$$
$$\implies E[X] = \sum_{i=1}^{n} E[X_i] = n \cdot p$$
$$V(X) = \sum_{i=1}^{n} V(X_i) + 0 = n \cdot V(X_1) = np(1-p)$$
2. Für $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$. Dann ist $E[X] = \mu$ und $V(X) = \sigma^2$. (Übung!)

Bemerkung
im Allgemeinen impliziert Unkorreliertheit von X, Y **nicht**, dass X und Y unabhängig sind!

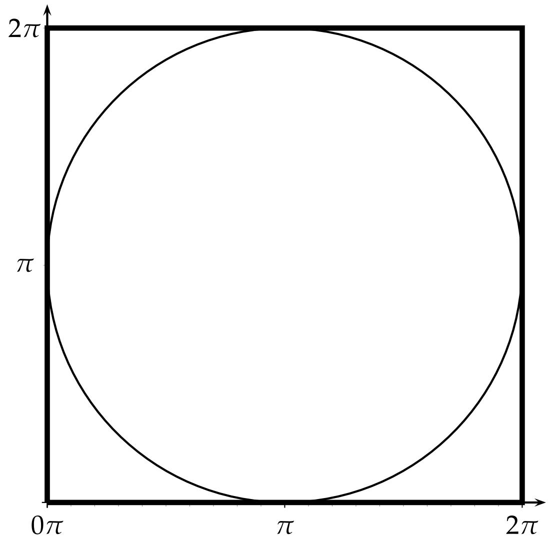
Beispiel (Gegenbeispiele)

1. $U \sim \mathcal{U}((0, 2\pi))$, d.h. gleichverteilt auf $(0, 2\pi)$



Sei $X = \sin U$ und $Y = \cos U$
$$\implies \left.\begin{array}{c} E[X] = E[Y] = 0 \\ \mathrm{Cov}(X, Y) = 0 \end{array}\right\} \implies X \text{ und } Y \text{ unkorreliert,}$$

aber sicher nicht unabhängig, weil $X^2 + Y^2 = 1$



2. $X \sim \mathcal{N}(0,1), \quad Y = X^2 - 1$

$$\implies E[Y] = 1 - 1 = 0$$

Übung: ein $X \sim \mathcal{N}(\mu, \sigma^2)$ hat alle Momente

$$E[|X|^p] < \infty \quad \forall p \in [1, \infty)$$

und

$$E[X^{2k+1}] = 0, \text{ falls } \mu = 0 \ \ (X \text{ zentriert})$$

$\implies$ $\mathrm{Cov}(X,Y) = E[X \cdot Y] = E[X^3 - X] = 0 - 0 = 0 \to$ unkorreliert, aber **nicht** unabhängig

Bemerkung
Falls X $\mathbb{R}^n$-wertig, $X \in \mathrm{L}^2(P)$, $\Sigma := \mathrm{Cov}(X,X)$, $Y := AX + b$ mit A eine $m \times n$-Matrix, $b \in \mathbb{R}^m$
Dann gilt:

$$\mathrm{Cov}(Y,Y) = A\Sigma A^T,$$

denn $\mathrm{Cov}(Y^i, Y^j)) = \mathrm{Cov}\left((AX)^i, (AX)^j\right) = \cdots$ (Übung!)

Definition 5.50 (Korrelationskoeffizient)
$X, Y \in L^2(P)$, $V(X), V(Y) > 0$. 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 durch $\rho(X,Y)$.

Lemma 5.51
X, Y wie in Definition 5.50. 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 $a^* = \frac{\mathrm{Cov}(X,Y)}{V(X)} = \mathrm{Corr}(X, Y)\sqrt{\frac{V(X)}{V(Y)}}$, $b^* = EY$ *und* $\min_{a,b\in\mathbb{R}} E[...] = V(Y)(1 - (\mathrm{Corr}(X, Y))^2)$

Beweis:
Übung!
Für b) Teil 1), a^*, b^* bestimmen, dann einsetzen und Minimum ausrechnen.
Für a) aus Cauchy-Schwarz Ungleichung. □

6 Die Gesetze der großen Zahlen

(starkes und schwaches Gesetz)

Bemerkung
Klassische Formulierung der Tschebyschev (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
Sei $(\Omega, \mathcal{F}, P)$ Wahrscheinlichkeitsraum, $Y, (Y_n)_{n\in\mathbb{N}}$ $\mathbb{R}$-wertige Zufallsvariablen.

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

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

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 $\varepsilon > 0$

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

Bemerkung

1. $(*) \Leftrightarrow \frac{1}{n}\sum_{i=1}^{n}(X_i - E[X_i])$ konvergiert stochastisch gegen 0.
2. Falls $E[X_k] = E[X_1]$, $\forall k$, so gilt:

$$\left(\frac{1}{n}\sum_{i=1}^{n} X_n\right) - E[X] \underset{\text{„Konv. in P"}}{\overset{P}{\longrightarrow}} 0 \quad (n \to \infty)$$

Beweis:
$Y_n := \frac{1}{n}\sum_{i=1}^{n}(X_i - EX_i) \;\Rightarrow Y_n \in L^2$ und $E[Y_n] = 0$

$$V(Y_n) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) \leq \frac{1}{n^2} nc \overset{n\to\infty}{\longrightarrow} 0$$

$\Rightarrow Y_n \overset{P}{\to} 0,$ denn dank Tschebyschev gilt:

$$P[|Y_n - 0| \geq \varepsilon] \leq \frac{V(Y_n)}{\varepsilon^2} \leq \frac{c}{n\varepsilon^2} \searrow 0$$

$\square$

Wir werden sehen, dass unter gleichen Voraussetzungen sogar P-fast-sichere Konvergenz anstelle von "nur" stochastischer Konvergenz gilt.

Konvergenzbegriffe für Folgen
$(Y_n)_{n\in\mathbb{N}}, Y$ Zufallsvariablen, $\mathbb{R}-$wertig.

- $Y_n \to Y$ **stochastisch in** P heißt:

$$P[\underbrace{|Y_n - Y| \geq \epsilon}_{\{\omega\in\Omega:\, |Y_n(\omega)-Y(\omega)|\geq\epsilon\}}] \xrightarrow[n\to\infty]{} 0 \quad \forall\epsilon > 0$$

- $Y_n \to Y$ **P- (P-f.s.)** heißt:

$$P[\underbrace{\lim_{n\to\infty} Y_n = Y}_{\{\omega\in\Omega:\, Y_n(\omega)\to Y(\omega)\}}] = 1$$

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** gegen Y falls die Folge außerhalb einer Nullmenge konvergiert. Das heißt*

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

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

Bemerkung (zu Definition 6.54)
Der Begriff der P-fast-sicheren Konvergenz ist wohldefiniert, denn $\{Y_n \xrightarrow{n\to\infty} Y\}$ ist messbar:

$$\{Y_n \xrightarrow{n\to\infty} Y\} = \{\omega \in \Omega : \forall n \in \mathbb{N}, \exists k \in \mathbb{N} : \forall l \geq k : |Y_l(\omega) - Y(\omega)| \leq \frac{1}{n}\} = \bigcap_{n\in\mathbb{N}} \bigcup_{k\in\mathbb{N}} \bigcap_{l=k}^{\infty} \left\{|Y_l - Y| \leq \frac{1}{n}\right\} \in \mathcal{F}$$

Bemerkung
Für ein Maß $\mu : \mathcal{F} \to [0, \infty]$ sagt man allgemeiner, dass ein $A \in \mathcal{F}$ $\quad \mu-$fast überall gilt, falls $\mu(A^C) = 0$ gilt.

Beispiel
$f, g : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ meßbare Funktionen, μ = Lebesguemaß, $f, g \in L^p(\mu)$, d.h. zu Beispiel:

$$\underbrace{(\int_{\mathbb{R}} |f|^p d\mu)^{\frac{1}{p}}}_{=\|f\|_{L^p(\mu)}} < \infty$$

<u>Schwaches Gesetz der großen Zahl (vgl. Satz 6.53)</u>
$(X_n)_{n\in\mathbb{N}}, \mathbb{R}-$ wertige Zufallsvariable, paarweise unkorreliert, mit $\sup_{n\in\mathbb{N}}(V[X_n]) < \infty$. Dann gilt, dass

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - E[X_i]) \xrightarrow{n\to\infty} 0 \quad P-stochastisch$$

Anwendungsbeispiele

a) *Monte-Carlo Integration*
Wir betrachten eine messbare Funktion $f : [0,1]^d \to [0,c]$ mit $c \in \mathbb{R}^+$ (z.B. f stetig und positiv) und suchen eine numerische Approximation von $\int_{[0,1]^d} f(x)\mathrm{d}x$ wobei die Dimension d groß ist. Dazu simulieren wir unabhängige Zufallsvariablen X_i, welche gleichverteilt auf $[0,1]^d$ sind. Dann gilt

$$P\left[\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \int_{[0,1]^d} f\mathrm{d}x\right| \geq \epsilon\right] = P\left[\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right| - E[f(X_1)] \geq \epsilon\right] \leq \frac{V(X_1)}{n\epsilon^2} \xrightarrow{n\to\infty} 0$$

Das heißt für genügend großes n können wir $\int f\mathrm{d}x$ durch Monte-Carlo Simulation approximativ berechnen.

b) Wir untersuchen die gleichmäßige Approximation einer stetigen Funktion $f : [0,1] \to \mathbb{R}$ durch Polynome. Seien $X_1, X_2, \cdots$ unabhängige Bernoulli(p) verteilte Zufallsvariablen. Dann gilt

$$E\left[f\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)\right] = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\mathrm{Bin}_{n,p}(k) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\cdot\binom{n}{k} p^k(1-p)^{n-k} =: \underbrace{f_n(p)}_{\text{BERNSTEIN Polynom } n\text{-ten Grades}}$$

Dann gilt:

$$\|f_n - f\|_\infty = \sup_{p\in[0,1]} |f_n(p) - f(p)| \xrightarrow{n\to\infty} 0$$

Denn: f ist stetig, also gleichmäßig stetig auf dem Kompaktum $[0,1]$:

$$\forall\epsilon > 0\ \exists\delta > 0 \text{ so dass } \forall x,y : |x-y| < \delta \implies |f(x) - f(y)| \leq \epsilon$$

Damit folgt für beliebiges $\epsilon > 0$

$$\begin{aligned}
|f_n(p) - f(p)| &= \left|E\left[f\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) - f(p)\right]\right| \\
&\leq E\left[\left|f\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) - f(p)\right| \left(\mathbb{1}_{\{|\frac{1}{n}\sum_{k=1}^{n} X_k - p| < \delta\}} + \mathbb{1}_{\{|\frac{1}{n}\sum_{k=1}^{n} X_k - p| \geq \delta\}}\right)\right] \\
&\leq \epsilon + 2\|f\|_\infty \frac{1}{n\delta^2} p(1-p) \searrow_{n\to\infty} \epsilon
\end{aligned}$$

$$\implies \lim_{n\to\infty} \|f_n - f\|_\infty = 0$$

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

Beweis:

$$\begin{aligned}
1 &= P(Y_n \to Y) \\
&= P(\bigcap_{n=1}^{\infty}\bigcup_{k=1}^{\infty}\bigcap_{l=k}^{\infty}\{|Y_l - Y| \leq \frac{1}{n}\}) \\
&= \lim_{n\to\infty} P(\bigcup_{k=1}^{\infty}\bigcap_{l=k}^{\infty}\{|Y_l - Y| \leq \frac{1}{n}\})
\end{aligned}$$

$\implies$

$$\begin{aligned}
1 &= P(\bigcup_{k=1}^{\infty}\bigcap_{l=k}^{\infty}\{|Y_l - Y| \leq \frac{1}{n}\}) \\
&= \lim_{k\to\infty} P(\bigcap_{l=k}^{\infty}\{|Y_l - Y| \leq \frac{1}{n}\})
\end{aligned}$$

$\Longrightarrow$

$$P(\{|Y_k - Y| > \frac{1}{n}\}) \leq P(\bigcup_{l=k}^{\infty}\{|Y_l - Y| > \frac{1}{n}\})$$
$$= 1 - P(\bigcap_{l=k}^{\infty}\{|Y_l - Y| \leq \frac{1}{n}\}) \xrightarrow{k\to\infty} 0$$

Das heißt gerade $Y_k \xrightarrow{P} Y$. □

Bemerkung
Im Allgemeinen gilt die Umkehrung nicht, stochastische Konvergenz impliziert nicht die P-fast-sichere Konvergenz.

Gegenbeispiel: Wir wählen $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$ und P als Gleichverteilung (mit Lebesguemaß). Dann sei

$$Y_k := \mathbb{1}_{[m2^{-n},(m+1)2^{-n}]}, \quad k = 2^n + m, \quad 0 \leq m < 2^n, \quad n \in \mathbb{N}$$

Dann gilt $P(|Y_k - 0| > \epsilon) \leq \frac{1}{2^n}$ für $2^n \leq k < 2^{n+1}$. Also konvergiert Y_k P-stochastisch gegen 0 jedoch nicht P-fast-überall. Es gilt sogar $\limsup\limits_{k\to\infty} Y_k(\omega) = 1$ und $\liminf\limits_{k\to\infty} Y_k(\omega) = 0 \quad \forall\omega \in \Omega$. Das heißt, Y_k konvergiert nirgends punktweise.

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

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

Beweis:

- O.B.d.A gelte $E[X_n] = 0$ (sonst betrachte $X_n' = X_n - E[X_n]$)
- $Y_n := \frac{1}{n}\sum\limits_{k=1}^{n} X_k$, $n \in \mathbb{N}$. Dann gilt:

$$Y_{n^2} \overset{n\to\infty}{\longrightarrow} 0 \quad P\text{-fast sicher}$$

Denn:

$$P[\underbrace{\{|Y_{n^2}| > \varepsilon\}}_{=:A_n(\varepsilon)}] \overset{\text{Tschebyschev}}{\leq} \frac{c}{n^2\varepsilon^2}, \quad \varepsilon < 0$$

$\Rightarrow \sum\limits_{n=1}^{\infty} P(A_n(\varepsilon)) < \infty$

$\overset{\text{Borel-Cantelli}}{\Longrightarrow}$ P(nur endlich viele der $A_n(\varepsilon), n \in \mathbb{N}$, treten auf) $= 1$ d.h.

$$P\left(\bigcup_{m\in\mathbb{N}}\bigcap_{n\geq m} A_n(\varepsilon)^c\right) = 1$$

Also gilt für fast alle $\omega \in \Omega$ (d.h. $\forall \omega \in \Omega \setminus N$, mit $N \in \mathcal{F}, P(N) = 0$), dass $\exists n \in \mathbb{N}$, so dass $\forall m \geq n$: $|Y_{m^2}(\omega)| \leq \varepsilon$
$\Rightarrow$ für fast alle $\omega \in \Omega$ also

$$\limsup_{n\to\infty} |Y_{n^2}(\omega)| \leq \varepsilon \quad \forall \varepsilon > 0$$

$\Longrightarrow$

$$\bigcap_{k\in\mathbb{N}} \underbrace{\bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} A_n\left(\frac{1}{k}\right)}_{P(\ldots)=1}$$

ist Ereignis mit P-Wahrscheinlichkeit =1 und für jedes ω daraus gilt:

$$Y_{n^2}(\omega) \overset{n\to\infty}{\longrightarrow} 0$$

$\Rightarrow$ also $Y_{n^2} \to 0$ P-fast sicher

- Zeigen nun "$Y_n \to 0$ P-fast sicher„:
  für $m \in \mathbb{N}$ gibt es $n = n(m)$, so dass $n^2 \leq m < (n+1)^2$ dann

$$P\left[|mY_m - n^2 Y_{n^2}| \geq n^2\varepsilon\right] \overset{\text{Tschebischev}}{\leq} \frac{1}{\varepsilon^2 n^4} V(\sum_{k=n^2+1}^{m} X_k) \leq \frac{1}{\varepsilon^2 n^4} c(m-n^2) < \infty$$

$\Longrightarrow$

$$\sum_{m\in\mathbb{N}} P\left[|mY_m - n(m)^2 Y_{n(m)^2}| \geq n(m)^2\varepsilon\right] \leq \sum_{m\in\mathbb{N}} \frac{1}{\varepsilon^2 n(m)^4} c(m - n(m)^2) = \frac{c}{\varepsilon^2} \sum_{n\in\mathbb{N}} \sum_{m=n^2}^{(n+1)^2-1} \frac{m-n^2}{n^4}$$
$$= \frac{c}{\varepsilon^2} \sum_{n\in\mathbb{N}} \sum_{k=1}^{2n} \frac{k}{n^4} = \frac{c}{\varepsilon^2} \sum_{n\in\mathbb{N}} \frac{2n(2n+1)}{2n^4} < \infty$$

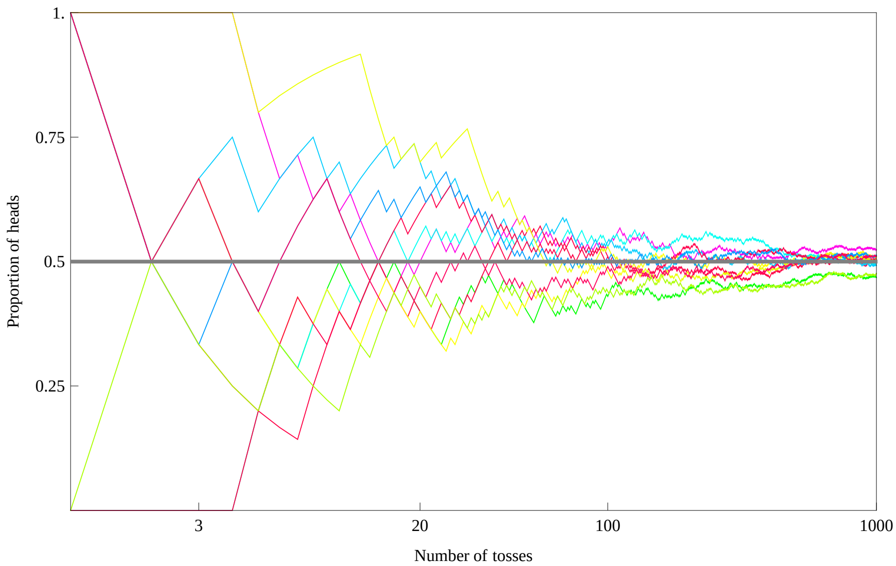
mit Borel-Cantelli folgt analog zum vorigem Schritt, dass

$$P\left[|\frac{m}{n(m)^2} Y_m - Y_{n(m)^2}| \overset{m\to\infty}{\longrightarrow} 0\right] = 1$$

d.h.$|\frac{m}{n(m)^2} Y_m - Y_{n(m)^2}| \overset{m\to\infty}{\longrightarrow} 0$ P-fast sicher zusammen mit $Y_{n^2} \to 0$ P-fast sicher folgt $\frac{m}{n(m)^2} Y_m \overset{m\to\infty}{\longrightarrow} 0$ P-fast sicher also, weil $\frac{m}{n(m)^2} \overset{m\to\infty}{\longrightarrow} 1$, folgt $Y_m \to 0$ P-fast sicher.

$\square$

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 $X_1, X_2, \dots$ 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, S.112)

Definition 6.57
Seien $X_1, X_2, \dots$ i.i.d. Zufallsvariablen in $L^2(P)$, für jede Realisierung $\omega \in \Omega$ von $X_1(\omega), X_2(\omega), \dots$ heißt

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

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

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

Bemerkung

1. Für jedes $x \in \mathbb{R}$ gilt:
$$\omega \mapsto F_n(x,\omega)$$
ist Zufallsvariable mit Werten in $[0,1]$!
2. Für jedes $\omega \in \Omega$ gilt:
$$x \mapsto F_n(x,\omega)$$
ist Realisierung einer (zufälligen) Verteilungsfunktion.

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

Dies ist eine <u>Motivation</u> der stärkeren Aussage von:

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 $X_1, \dots, X_n$ und sei F die Verteilungsfunktion von X_n. Dann gilt:

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

Beweis:
$Y_n(x) := \mathbb{1}_{(-\infty,x]}(X_n)\,,\ \ n \in \mathbb{N},\ \ x \in \mathbb{R}$
$Z_n(x) := \mathbb{1}_{(-\infty,x)}(X_n)$
$\Longrightarrow$ $(Y_n(x))_{n\in\mathbb{N}}$ und $(Z_n(x))_{n\in\mathbb{N}}$ sind jeweils i.i.d. Folgen von Zufallsvariablen und sind jeweils Bernoulli-Folgen mit Erfolgswahrscheinlichkeit $F(x)$ bzw. $F(x-)$, wobei

$$F(x-) := \lim_{y\uparrow x} F(y)$$

Analog definiert sei $F_n(x-)$.
$E[Y_n(x)] = F(x), \quad E[Z_n(x)] = F(x-)$
Nach dem Satz 6.56 (starkes Gesetz der großen Zahlen) gilt:

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} Y_k(x) \overset{n\to\infty}{\longrightarrow} F(x) \quad P\text{-fast sicher}$$
$$F_n(x-) = \frac{1}{n}\sum_{k=1}^{n} Z_k(x) \overset{n\to\infty}{\longrightarrow} F(x-) \quad P\text{-fast sicher}$$

Sei $F(-\infty) = 0,\ \ F(+\infty) = 1$
Fixiere $N \in \mathbb{N}$, wir definieren

$$x_j := \inf\{x \in \overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\} \cup \{-\infty\} \mid F(x) \geq \frac{j}{N}, \ j = 0, \dots, N\}$$
$$R_N := \max_{j=0,\dots,N}\{\left|F_n(x_j) - F(x_j)\right| + \left|F_n(x_j-) - F(x_j-)\right|,\ \ n \in \mathbb{N}\}$$

$$\Longrightarrow \ R_n \underset{n\to\infty}{\longrightarrow} 0 \quad P\text{-fast sicher}$$

Für jedes $x \in \mathbb{R}$ mit $x \in \left(x_{j-1}, x_j\right)$ gilt, dass

$$F_n(x) \leq F_n(x_j-) \leq F_n(x_j-) + R_n \leq F(x) + R_n + \frac{1}{n}$$

und analog

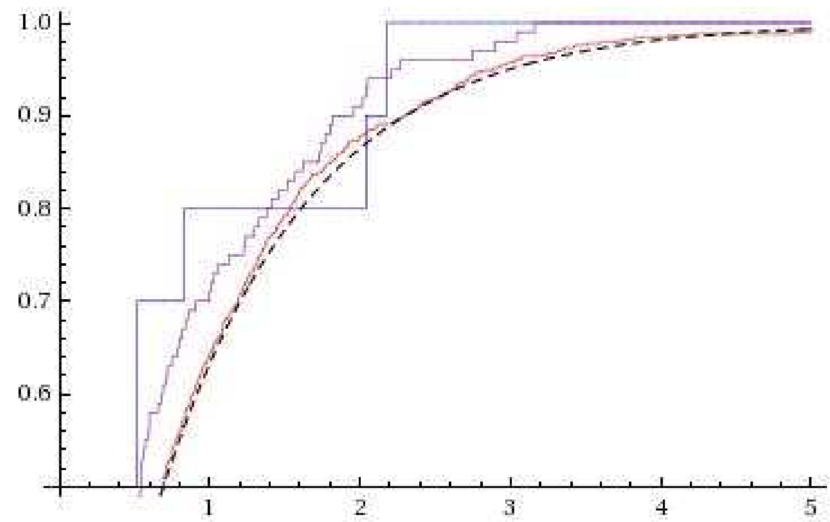
$$F_n(x) \geq F_n(x_{j-1}) \geq F(x_{j-1}) - R_n \geq F(x) - R_n - \frac{1}{n}$$

$$\Longrightarrow \qquad \limsup_{n\to\infty} \left(\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \right) \leq \frac{1}{n} + \underbrace{\limsup_{n\to\infty} R_n}_{=0} \leq \frac{1}{n} \qquad \forall n \in \mathbb{N}$$

$$\Longrightarrow \qquad \limsup_{n\to\infty} \left(\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \right) = 0$$

□

Beispiel
Die folgende Grafik zeigt 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 Funktion

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 $\mathbb{R}^n$ 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:

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

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:

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

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

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

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

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

Bemerkung

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

Bemerkung

$$\begin{aligned}
\hat{\mu}(u) &= \int e^{i\langle u, x \rangle} \mu(dx) = \int_{\mathbb{R}^d} \cos(\langle u, x \rangle) \mu(dx) + i \int_{\mathbb{R}^d} \sin(\langle u, x \rangle) \mu(dx) \quad \in \mathbb{C}, \ \forall u \in \mathbb{R}^d\\
\varphi_X(u) &= E[\cos(\langle u, X \rangle)] + iE[\sin(\langle u, X \rangle)] \quad \in \mathbb{C}, \ \forall u \in \mathbb{R}^d
\end{aligned}$$

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 $\hat{\mu}(0) = 1$

Beweis:

- $\hat{\mu}(0) = \cdots = 1$ ✓
- $\hat{\mu}$ beschränkt, weil $|\hat{\mu}(u)| \leq \int\limits_{\mathbb{R}^d} |e^{i\langle u,x\rangle}|\mu(dx) \leq 1$ ✓
- Stetigkeit von $\hat{\mu}$:
  Es gelte $u_n \to u$, zu zeigen ist, dass $\hat{\mu}(u_n) \to \hat{\mu}(u)$:
  $e^{i\langle u_n,x\rangle} = \cos(\langle u_n, x\rangle) + i\sin(\langle u_n, x\rangle) \longrightarrow \cos(\langle u, x\rangle) + i\sin(\langle u, x\rangle) = e^{i\langle u,x\rangle} \quad \forall x \in \mathbb{R}^d$
  da $|e^{i\langle u_n,x\rangle}| \leq 1 \in L^1(\mu)$ gilt mit Theorem der majorisierten Konvergenz, dass:
  $E[e^{i\langle u_n,X\rangle}] \xrightarrow{n\to\infty} E[e^{i\langle u,X\rangle}]$ ✓

□

Bemerkung (das m-te Moment)
X eine $\mathbb{R}^d$-wertige Zufallsvariable.
$E[|X|^m]$ heißt **m-tes Moment** von $X, m \in \mathbb{N}$.
$E[|X - E[X]|^m]$ heißt **zentriertes m-tes Moment** von $X, m \in \mathbb{N}$.

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:

$$\frac{\partial^m}{\partial x_{j1}\partial x_{j2}\ldots\partial x_{jm}}\varphi_X(u) = i^m E[x_{j1}x_{j2}\ldots x_{jm}e^{\langle u,X\rangle}]$$

Beweis:
Sei $\mu := P_X$ Verteilung auf $\mathcal{B}^d(\mathbb{R})$, $\int |X|^m\mu(dx) < \infty$, d.h. $|X|^m \in L^1(\mu)$

- Zeigen Behauptung für $m = 1$. Durch wiederholte Anwendung des Arguments zeigt man Behauptung für $m > 1$
- Existenz von $\dfrac{\partial}{\partial x_j}\hat{\mu}$ an jedem $u \in \mathbb{R}^d$:
  Zu zeigen:
  $\dfrac{\hat{\mu}(u + t_n e_j) - \hat{\mu}(u)}{t_n}$ konvergiert für $t_n \to 0$ für jeden Einheitsvektor e_j gegen den behaupteten Grenzwert, welcher stetig in u ist:

$$\frac{\hat{\mu}(u + t_n e_j) - \hat{\mu}(u)}{t_n} = \int\limits_{\mathbb{R}^d} e^{i\langle u,x\rangle}\frac{e^{i\langle t_n e_j,x\rangle} - 1}{t_n}\mu(dx) \qquad (*)$$

Für den Bruch im Integranden gilt:

$$\frac{\mathrm{e}^{i\langle t_n e_j, x\rangle} - 1}{t_n} \xrightarrow{n\to\infty, t_n\to 0} \frac{\cos(\langle t_n e_j, x\rangle) - 1 + i\sin(\langle t_n e_j, x\rangle)}{t_n}$$
$$\xrightarrow[\text{l'Hospital}]{n\to\infty} -x_j \sin(0) + i x_j \cos(0) = i x_j$$

mit punktweiser Konvergenz $\forall x \in \mathbb{R}$; weiter gilt:

$$\left| \frac{\mathrm{e}^{i\langle t_n e_j, x\rangle} - 1}{t_n} \right| \leq 2|x| \in \mathrm{L}^1(\mu) \qquad \text{für } n \text{ groß genug}$$

Mit majorisierter Konvergenz folgt also:

$$(*) \xrightarrow{n\to\infty} \int\limits_{\mathbb{R}^d} \mathrm{e}^{i\langle u,x\rangle}(i x_j)\mu(\mathrm{d}x) = iE[X_j \mathrm{e}^{i\langle u,X\rangle}] = \frac{\partial}{\partial x_j}\varphi_X(u)$$

Stetigkeit von $\frac{\partial}{\partial x_j}\varphi_X(u)$ in $u \in \mathbb{R}^d$ zeigt man wieder mittels majorisierter Konvergenz.

$\square$

Beispiel

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

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

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

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

e) $X \sim \mathcal{N}(0,1)$:
$$\varphi_X(u) = E[\mathrm{e}^{iux}] = \int\limits_{\mathbb{R}} \cos(ux)\frac{1}{\sqrt{2\pi}}\exp(\frac{-x^2}{2})\mathrm{d}x + i \underbrace{\int\limits_{\mathbb{R}} \sin(ux)\frac{1}{\sqrt{2\pi}}\exp(\frac{-x^2}{2})\mathrm{d}x}_{=0,\text{ da Integrand in } L^1(\mathrm{d}x) \text{ und ungerade}} > 0 \quad \forall u \in \mathbb{R}$$

$$\implies \varphi_X'(u) = (\text{mit vorherigem Theorem}) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} -x\sin(ux)\exp(\frac{-x^2}{2})\mathrm{d}x$$
$$\overset{\text{part.Int.}}{=} -\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} u\cos(ux)\exp(\frac{-x^2}{2})\mathrm{d}x = -u\varphi_X(u)$$
$$\implies \frac{\varphi_X'(u)}{\varphi_X(u)} = -u \xrightarrow{\text{integrieren}} \ln\varphi_X(u) = -\frac{u^2}{2} + c \Rightarrow \varphi_X(u) = \exp(\frac{-u^2}{2} + c) = \exp(\frac{-u^2}{2})$$

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:

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

Beweis:
Übung! □

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 $Y := \dfrac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$
$\Rightarrow \varphi_X(u) = \exp(iu\mu + \sigma^2 \frac{u^2}{2})$

Beispiel
$X_1, .., X_d$ unabhängig identisch normalverteilt: X_k i.i.d. $\sim \mathcal{N}(0,1)$, dann heißt $X = (X_1, \ldots, X_d)$ standardnormalverteilt in $\mathbb{R}^d$ und

$$\begin{aligned}\varphi_X(a) &= E[\exp(i\langle a, X\rangle)] \\ &= \prod_{k=1}^{n} \exp(\frac{-a_k^2}{2}) \\ &= \exp(\frac{-|a|^2}{2})\end{aligned}$$

Für X wie oben, A, b wie aus Lemma 7.62 heißt $Y := AX + b$ **multivariat normalverteilt** und $\varphi_Y(\mu)$ berechnet sich entsprechend Lemma 7.62.

charakteristische Funktionen: X Zufallsvariable, $\mathbb{R}^d$-wertig

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

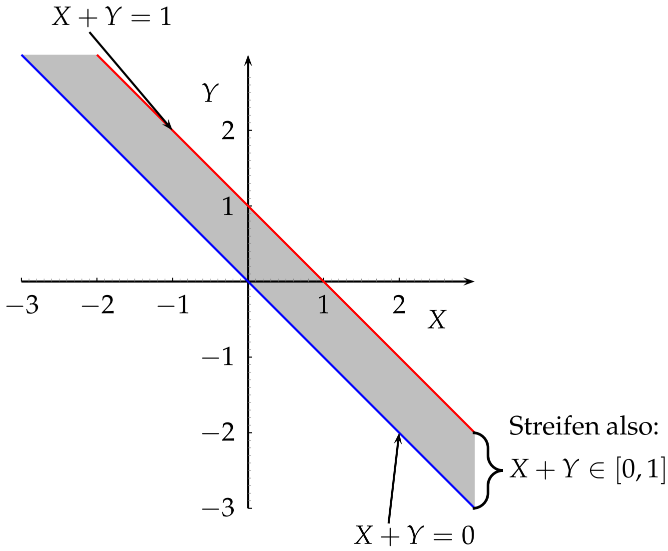
Beispiel $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}^1$, $\sigma > 0$

$$\varphi_X(u) = \exp(iu\mu - \frac{\sigma^2 u^2}{2})$$

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 $\mathbb{R}$-wertige unabhängige Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$ mit Verteilungen $P_X = P \circ X^{-1}$, $P_Y = P \circ Y^{-1}$ auf $\mathcal{B}^1(\mathbb{R})$. 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:*

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



Beweis:

$$\begin{aligned}
P_{(X,Y)} = P_X \otimes P_Y, &\quad \text{da unabhängig} \\
\implies E[g(X,Y)] &= \int\int g(x,y)P_X(\mathrm{d}x)P_Y(\mathrm{d}y) \\
\text{für } g : \mathbb{R}^2 \mapsto \mathbb{R} \text{ messbar}, &\ g(X,Y) \in \mathrm{L}^1(P_X \otimes P_Y) \text{ oder } g \geq 0 \\
\implies (\text{für } g(x,y)) = f(x+y) \text{ also}) &= \int\int f(x+y)P_X(\mathrm{d}x)P_Y(\mathrm{d}y) \\
\implies (\text{für } f = \mathbb{1}_A \text{ also}) &= \int\int \mathbb{1}_A(x+y)P_X(\mathrm{d}x)P_Y(\mathrm{d}y) \\
&= P\left[X + Y \in A\right] \equiv P\left[Z \in A\right] = P_Z(A)
\end{aligned}$$

□

Bemerkung
Faltung kann analog in Dimension $n \geq 1$ definiert werden.

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

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

Beweis:

Wähle für f wie oben (Real- und Imaginärteil von) $e^{i\langle u,X\rangle}$. Dann ist

$$\begin{aligned} E\left[e^{i\langle u,X+Y\rangle}\right] &= E\Big[\underbrace{e^{i\langle u,X\rangle}\cdot e^{i\langle u,Y\rangle}}_{\text{unabhängige Faktoren}}\Big] \\ &= E\left[e^{i\langle u,X\rangle}\right]\cdot E\left[e^{i\langle u,Y\rangle}\right] \end{aligned}$$

□

Bemerkung

Falls $Z = X + Y$ gilt, genügt $\varphi_Z(u) = \varphi_X(u)\cdot\varphi_Y(u) \;\; \forall u \in \mathbb{R}$ **NICHT**, um zu schließen, dass X und Y unabhängig sind!

Wir werden sehen: hinreichend und notwendig für "X, Y unabhängig" wäre, dass

$$\varphi_{(X,Y)}(u_1, u_2) = \varphi_X(u_1)\cdot\varphi_Y(u_2) \quad \forall (u_1,u_2) \in \mathbb{R}^2$$

Satz 7.65

Seien X, Y unabhängige, $\mathbb{R}^1$-wertige Zufallsvariablen und $Z := X + Y$

a) hat zudem X eine Dichte f_X, dann hat Z eine Dichte f_Z auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und es gilt

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

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

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

Beweis:

a)

$$P_Z(A) = \int \underbrace{\underbrace{\int \mathbb{1}_A(x+y)P_X(\mathrm{d}x)}_{\int \mathbb{1}_A(x+y)f_X(x)\mathrm{d}x}}_{\overset{z=x+y}{\Longrightarrow}\int \mathbb{1}_A(z)f_X(z-y)\mathrm{d}z} P_Y(\mathrm{d}y) = \int \mathbb{1}_A(z)\underbrace{\int f_X(z-y)P_Y(\mathrm{d}y)}_{\overset{!}{=}f_Z(z)}\mathrm{d}z \quad \forall A \in \mathcal{B}(\mathbb{R})$$

$$\Longrightarrow f_Z(z) = \int_{\mathbb{R}} f_X(z-y)\underbrace{P_Y(\mathrm{d}y)}_{\text{ggf. } f_Y(y)\mathrm{d}y} \qquad (a \;\checkmark)$$

und weiter

$$f_Z(z) = \int f_X(z-y)f_Y(y)\mathrm{d}y, \qquad (b \;\checkmark)$$

falls Y Dichte f_Y hat.
Die zweite Gleichung in $b)$ folgt per Symmetrie zwischen X und Y.

$\square$

Warum heißen charakteristische Funktionen "charakteristisch"?

Die Antwort gibt:

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

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

Beweis:

Der Beweis nutzt die lokal-kompakte (erweiterte) Version des "Stone-Weierstraß-Theorems" aus der Funktionalanalysis, siehe z.B. Simmons "Introduction to Topology and Modern Analysis" Seite 166, Theorem A.

Betrachte die Funktionen

$$f(\sigma, x) = \frac{1}{(\sqrt{2\pi\sigma^2})^d} \exp\left(-\frac{\|x\|_2^2}{2\sigma^2}\right), \quad x \in \mathbb{R}^d, \sigma > 0$$

$$\hat{f}(\sigma, u) := \exp\left(-\frac{\|u\|_2^2 \cdot \sigma^2}{2}\right), \quad u \in \mathbb{R}^d$$

$$\text{mit } \|x\|_p = \left(\sum_{k=1}^n |x_k|^p\right)^{\frac{1}{p}}$$

Also ist $f(\sigma, \cdot)$ die gemeinsame Dichte von $X = (X_1, \ldots, X_d)$ für $X_1, \ldots, X_d$ i.i.d. mit $X_j \sim \mathcal{N}(0, \sigma^2)$

$$\begin{aligned}
\implies \varphi_X(u) = E\left[\exp(i\langle u, X\rangle)\right] &= \prod_{j=1}^d \varphi_{X_j}(u_j) \\
&= \prod_{j=1}^d \exp\left(-\frac{u_j^2 \cdot \sigma^2}{2}\right) \\
&= \exp\left(-\frac{\|u\|_2^2 \cdot \sigma^2}{2}\right) \overset{\text{(Def.)}}{=} \hat{f}(\sigma, u)
\end{aligned}$$

$$\implies f(\sigma, u - v) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} \hat{f}\left(\sigma, \frac{u - v}{\sigma^2}\right) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} E\left[e^{i\langle \frac{u-v}{\sigma^2}, X\rangle}\right]$$

Seien μ_1, μ_2 Wahrscheinlichkeitsmaße auf $(\mathbb{R}^d, \mathcal{B}^d(\mathbb{R}))$ mit $\hat{\mu}_1 = \hat{\mu}_2 =: \hat{\mu}$.

Zu zeigen ist "$\mu_1 = \mu_2$". Dazu:

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

Eine analoge Formel gilt mit $\mu_2, \hat{\mu}_2$, also folgt:

$$\int f(\sigma, u-v)\mu_1(\mathrm{d}u) = \int f(\sigma, u-v)\mu_2(\mathrm{d}u), \quad \forall \sigma > 0, v \in \mathbb{R}^d$$

$\Longrightarrow$ $\int g \mathrm{d}\mu_1 = \int g \mathrm{d}\mu_2$ gilt für alle Funktionen g aus dem Vektorraum $\mathcal{H}$, der aufgespannt wird durch $\{f(\sigma, \cdot - v) \mid \sigma > 0, v \in \mathbb{R}^d\}$. $\mathcal{H}$ trennt einzelne Punkte in $\mathbb{R}^d$, durch Anwendung von Stone-Weierstraß folgt, dass $\mathcal{H}$ dicht bzgl. gleichmäßiger Konvergenz (Konvergenz bzgl. der Supremumsnorm) in der Banachalgebra $\mathcal{C}^0(\mathbb{R}^d)$ liegt, wobei $\mathcal{C}^0(\mathbb{R}^d)$ den Banachraum der stetigen Funktionen $\mathbb{R}^d \mapsto \mathbb{R}$ bezeichnet (mit Konvergenz bzgl. der Supremumsnorm), welche "gegen ∞ verschwinden" (d.h. für $g \in \mathcal{C}^0(\mathbb{R}^d)$ gilt:
$\forall \varepsilon > 0 \; \exists$ Kompaktum $K \subset \mathbb{R}^d$ s.d. $|g| \leq \varepsilon$ auf $\mathbb{R}^d \setminus K$).

Weil $\int g \mathrm{d}\mu_1 = \int g \mathrm{d}\mu_2 \quad \forall g \in \mathcal{H}$ gilt, und weil jede Indikatorfunktion auf Rechtecken monoton approximiert werden kann durch Funktionen aus $\mathcal{H}$, d.h.

$$\mathbb{1}_{\times_{j=1}^d [a_j, b_j]} = \uparrow - \lim_{n\to\infty} g_n, \quad \text{für } g_n \in \mathcal{H}$$

$\Longrightarrow$ mit monotonem Konvergenztheorem folgt also

$$\mu_1\left(\times_{j=1}^d [a_j, b_j]\right) = \mu_2\left(\times_{j=1}^d [a_j, b_j]\right) \quad \forall a_j, b_j \in \mathbb{R},\ a_j < b_j$$

$\square$

Bemerkung
will man etwas Konstruktives zur *Berechnung* von μ aus $\hat{\mu}$ sagen, braucht man Resultate aus der Fourieranalysis zur *Inversion 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. Schirjajew, Probability/Wahrscheinlichkeit II, §12, Th. 3.].
Damit folgt:

$$\begin{aligned} G(b) :&= \lim_{a\downarrow-\infty} g(a,b) \\ &= \mu((-\infty, b)) + \frac{1}{2}\mu(\{b\}) \quad \text{wegen } \sigma\text{-Stegigkeit und } \mu(\{a\}) \leq \mu((-\infty, a]) \searrow 0 \text{ für } a \searrow -\infty \\ \Rightarrow P(X \leq b) &=: F(b) \\ &= G(b) + \frac{1}{2}(G(b) - G(b-)) \quad \forall b \in \mathbb{R} \end{aligned}$$

Aus $\hat{\mu}$ gewinnt man also konstruktiv g, G, F und somit $P_X = P \circ X^{-1}$ zurück.

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

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

Beweis:

"$\Rightarrow$" $X^1, \dots, X^d$ unabhängig:

$$E\left[\exp(iu\sum_{k=1}^{d} X_k)\right] = E\left[\prod_{k=1}^{d} \exp(iuX_k)\right]$$
$$= \prod_{k=1}^{d} E[\exp(iuX_k)] \quad \text{wegen Unabhängigkeit}$$

"$\Leftarrow$" es gilt $\widehat{P}_X = \big(P_{X_1} \otimes \widehat{P_{X_2} \otimes \cdots} \otimes P_{X_d}\big)$ auf $(u \in \mathbb{R}^d)$,

aus Satz 7.66 folgt, dass $P_X = \bigotimes_{k=1}^{d} P_{X_k}$

$\square$

8 Mehrdimensionale Normalverteilungen

- auch *multivariate* Normalverteilungen sind im Unterschied zu *univariaten* Normalverteilungen normalverteilt
- univariate Normalverteilung:
  $X \sim \mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}^1, \sigma > 0,$
  falls X absolutstetig mit Dichte $f(x) = \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, dann
  $\varphi_X(u) = \exp(iu\mu - \frac{1}{2}\sigma^2 u^2), \quad u \in \mathbb{R}^1, \quad E[X] = \mu$ und $V[X] = \sigma^2$
- wir bezeichnen nun mit $X \sim \mathcal{N}(\mu, 0)$ die Dirac-Punktmasse auf μ,
  das heißt $P_X = \delta_\mu$ mit $\delta_\mu(A) = \begin{cases} 1 & \mu \in A \\ 0 & \text{sonst} \end{cases}$

Motivation zum Studium und der Bezeichnung "normal" für solche Verteilungen

- Treten auf als generischer Verteilungsgrenzwert von Summen vieler unabhängiger Zufallsvariablen im *zentralen Grenzwertsatz*.
- daher **Normal**verteilung
- Praktische Relevanz, denn in Anwendungen treten solche Approximationen oft auf

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

Bemerkung
unter Umständen auch **degeneriert** normalverteilt mit $V[X] = 0$, d.h. $X \sim \mathcal{N}(\mu, 0)$ also Punktmasse 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 $\langle u, Qu\rangle = u^t Q u \geq 0 \;\; \forall u \in \mathbb{R}^d$).
Und es gilt dann, dass Q die Kovarianzmatrix $\mathrm{Cov}(X)$ *ist und μ der Mittelwertvektor* $E[X]$.

Beweis:

„$\Leftarrow$"

charakteristische Funktion φ_X habe Form aus Satz, $a \in \mathbb{R}^d$, $Y := \langle a, X\rangle$

$\Rightarrow \varphi_Y(v) = E[\exp(iv \sum_{k=1}^{d} a_k X_k)] = \varphi_X(va) = \exp(iv\langle a, \mu\rangle - \frac{1}{2}v^2\langle a, Qa\rangle), \quad (v \in \mathbb{R}^1,\ a \in \mathbb{R}^d)$

$\xRightarrow{\text{Satz 7.66}} Y \sim \mathcal{N}(\langle a, \mu\rangle, \langle a, Qa\rangle)$

und es folgt mit Satz 7.61 aus der Form von φ_X, dass Q die Kovarianzmatrix ist:

denn

$$E[X_j] = \frac{1}{i}\frac{\partial}{\partial x_j}\varphi_X(u)|_{u=0} = \frac{1}{i}(i\mu_j) = \mu_j,$$

und

$$\begin{aligned} E[X_iX_j] &= \frac{1}{(i^2)}\frac{\partial}{\partial x_i \partial x_j}\varphi_X(u)|_{u=0} \\ &= (-1)(-\mu_i\mu_j - Q_{ij}) \\ &= \mu_i\mu_j + Q_{ij} \\ \Rightarrow \operatorname{Cov}(X_i, X_j) &= E[X_iX_j] - E[X_i]E[X_j] = Q_{ij} \end{aligned}$$

„ $\Rightarrow$ “

Sei X Gauß'sch, $\quad Y := \langle a, X\rangle = \sum_{k=1}^d a_k X_k, \quad Q := \operatorname{Cov}(X)$
$\Rightarrow E[Y] = E[\langle a, X\rangle] = \langle a, E[X]\rangle = \langle a, \mu\rangle$
mit $\mu := E[X]$ und $V(Y) = a^t Q a = \langle a, Qa\rangle$
wegen $Y \sim \mathcal{N}(\langle a, \mu\rangle, \langle a, Qa\rangle)$ folgt $\varphi_Y(v) = \exp(iv\langle a, \mu\rangle - \frac{1}{2}v^2\langle a, Qa\rangle)$ für $v \in \mathbb{R}$
$\Rightarrow \varphi_Y(1) = \varphi_{\langle a, X\rangle}(1) = E[\exp(i \cdot 1\langle a, X\rangle)] = \varphi_X(a)$
also hat φ_X die behauptete Form!

$\square$

Beispiel
Seien $X_1, \ldots, X_d$ 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 $X = (X_1, \ldots, X_d)$ 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 $X_1, \ldots, X_d$ von X unabhängig genau dann, wenn die Kovarianzmatrix $\operatorname{Cov}(X) = Q$ *von X Diagonalform hat.*

Beweis:

„ $\Rightarrow$ “

siehe vorheriges Beispiel

„ $\Leftarrow$ “

Sei P_X, wir haben mit Satz 8.69, dass

$$\widehat{P}_X = (P_1 \widehat{\otimes \cdots \otimes} P_d)$$

für $P_j := \mathcal{N}(E[X_j], V(X_j))$
Mit Satz 7.66 folgt, dass $P_X = \bigotimes_{j=1}^d P_{X_j}$, also sind $X_1, \ldots, X_d$ unabhängig

□

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

Lemma 8.71
Sei X $\mathbb{R}^d$-wertige multivariate normalverteilte Zufallsvariable, $X \sim \mathcal{N}(\underset{\downarrow}{\overset{E[X]}{\mu}}, \underset{\downarrow}{\overset{\mathrm{Cov}(X)}{Q}})$. 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, $AA^T = I_d$).

Beweis:
$Q := \mathrm{Cov}(X)$, symmetrisch, nichtnegativ semidefinit

$$\Longrightarrow Q = A\Lambda A^T \text{ mit } A \text{ orthogonale Matrix, } \Lambda \text{ Diagnalmatrix } \begin{pmatrix} \lambda_1^2 & & 0 \\ & \ddots & \\ 0 & & \lambda_d^2 \end{pmatrix} \text{ mit } \lambda_j \geq 0$$

$$\Longrightarrow Y := A^T(X - \mu) \text{ ist } \mathcal{N}(0,1) \text{ verteilt, also } AY + \mu = (X - \mu) + \mu = X, \text{ also } \mathcal{N}(\mu, Q) \text{ verteilt.} \quad \checkmark$$

□

Bemerkung
Zur Simulation von $X \sim \mathcal{N}(\mu, Q)$ braucht man nur soviele Y_j zu simulieren, wie der Rang von Q (bzw. Λ) 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 $\bar{A}$, so dass $X = \mu + \bar{A}Y$ die gewünschte $\mathcal{N}(\mu, Q)$ Verteilung hat. [$\rightarrow$ lineare Algebra, Übung]

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

Beweis: (Beweisskizze)

"$\Leftarrow$" nach Lemma 8.71 gilt dann $X = \mu + \overline{A}Y$ mit $Y \sim \mathcal{N}(0, Id_d)$

$$\text{Wähle } \overline{A} = A \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix}$$

$$\Longrightarrow Y \text{ hat Dichte } \frac{1}{(2\pi)^{\frac{d}{2}}} \exp(-\frac{||y||_2^2}{2}) =: f_Y(y)$$

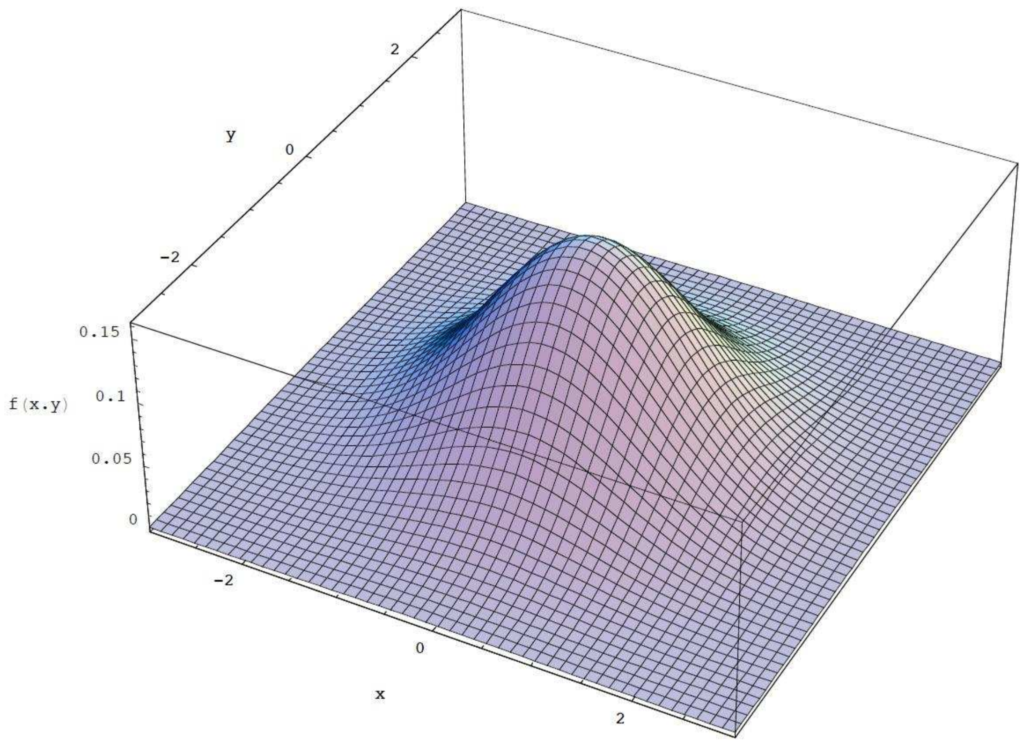
$\Longrightarrow$ mit Transformationsformel (vgl. Analysis 3, Maßtheorie) folgt:

$$\begin{aligned} f_X(x) &= |\det \overline{A}^{-1}| f_Y(\overline{A}^{-1}(x-\mu)) \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \frac{1}{\sqrt{\det Q}} \exp(-\frac{1}{2} < x-\mu, Q^{-1}(x-\mu) >) \end{aligned}$$

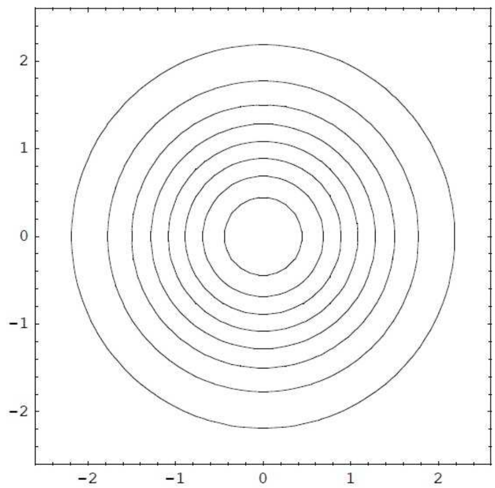
□

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 $\rho = 0$.

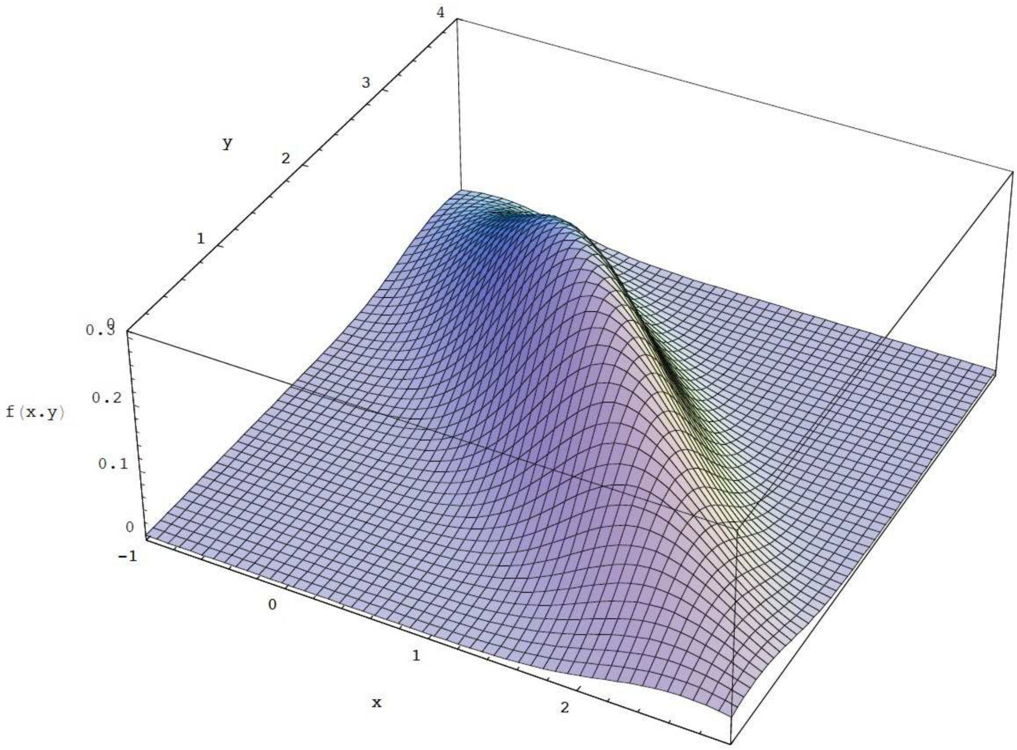


(Dichte der bivariaten Normalverteilung)

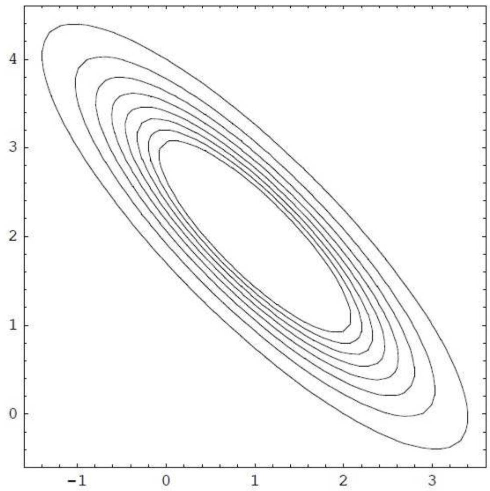


(Kontur- / Niveau- / Höhenlinien)

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 / schwache Konvergenz

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

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 $d(x, y) = |x - y|$, 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 $\mu_n \xrightarrow{w} \mu$ "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:

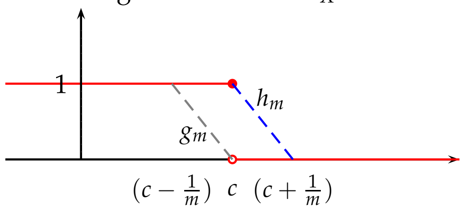
a) $X_n \xrightarrow{\mathcal{D}} X$

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

Beweis:

"a)⇒b)"

Sei c Stetigkeitsstelle von F_X



Wähle Funktionen g_m, h_m, $n \in \mathbb{N}$, so dass

$$\mathbb{1}_{(-\infty, c-\frac{1}{m}]} \leq g_m \leq \mathbb{1}_{(-\infty, c]} \leq h_m \leq \mathbb{1}_{(-\infty, c+\frac{1}{m}]}$$

$\Rightarrow$

$$\begin{array}{ccccc}
E[\mathbb{1}_{(-\infty,c-\frac{1}{m}]}(X_n)] \leq & E[g_m(X_n)] & \leq F_{X_n}(c) \leq & E[h_m(X_n)] \leq & E[\mathbb{1}_{(-\infty,c+\frac{1}{m}]}(X_n)] \\
 & \big\downarrow{\scriptstyle n\to\infty} & & \big\downarrow{\scriptstyle n\to\infty} & \\
\forall m \in \mathbb{N}: F_X(c-\frac{1}{m}) \leq & E[g_m(X)] & & E[h_m(X)] \leq & F_X(c+\frac{1}{m}) \\
 & \searrow{\scriptstyle m\to\infty} & & \swarrow{\scriptstyle m\to\infty} & \\
 & & F_X(c) & &
\end{array}$$

$\overset{\text{"Sandwich-Lemma"}}{\Rightarrow}$ dass $F_{X_n}(c) \to F_X(c)$

"b)⇒a)"

für $f \in \mathcal{C}_b(\mathbb{R}^1, \mathbb{R})$ eine beschränkte, messbare Funktion, $\varepsilon > 0$,
$F_Y(\cdot)$ höchstens abzählbar viele Sprungstellen.
Wähle $c_1, \ldots, c_m$ so dass:

- F_Y stetig bei c_i
- $F_Y(c_1) < \varepsilon$ und $F_Y(c_m) > 1 - \varepsilon$
- $\sup\limits_{x \in [c_i, c_{i+1}]} |f(x) - f(c_i)| \leq \varepsilon$

Dies ist möglich, denn f ist auf einem Kompaktum $[c_1, c_m]$ gleichmäßig stetig

$$\Rightarrow E[f(Y_n)] = E\left[f(Y_n)\mathbb{1}_{\{Y_n \in (-\infty, c_1] \cup (c_m \infty]\}}\right] + \sum_{i=2}^{m} E\left[f(Y_n)\mathbb{1}_{\{Y_n \in (c_{i-1}, c_i]\}}\right]$$

$$\leq \|f\|_\infty \cdot 2\varepsilon + \sum_{i=2}^{m} (f(c_i) + \varepsilon) \Big(\underbrace{F_{Y_n}(c_i)}_{\xrightarrow{n\to\infty} F_Y(c_i)} - \underbrace{F_{Y_n}(c_{i-1})}_{\xrightarrow{n\to\infty} F_Y(c_{i-1})} \Big)$$

$$\Rightarrow \overline{\lim_{n\to\infty}} \text{ linke Seite } \leq 2 \cdot \varepsilon \|f\|_\infty \quad \underbrace{+E[f(Y)] + 2\varepsilon + \|f\|_\infty \cdot 2\varepsilon}_{\geq E\left[\sum_{i=2}^{m} f(c_i)\mathbb{1}_{\{Y \in (c_{i-1}, c_i]\}}\right]}$$

$$= E[f(Y)] + 2\varepsilon(2\|f\|_\infty + 1) \qquad \forall \varepsilon > 0$$

$\Rightarrow \overline{\lim_{n\to\infty}} E[f(Y_n)] = \limsup_{n\to\infty} E[f(Y_n)] \leq E[f(Y)]$
Analoges Argument mit $-f$ anstelle von f liefert $\underline{\lim}_{n\to\infty} E[f(Y_n)] \geq E[f(Y)]$

$\square$

Beziehung zw. d. Konvergenzarten

[von Zuvallsvariablen]

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

Beweis:
$f \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}) \Rightarrow f(Y_n) \to f(Y)$ P-fast-sicher, da f stetig.
$|f(Y_n)| \leq \|f\|_\infty \in L^1(P)$, mit Satz der majorisierten Konvergenz (Maßtheorie) folgt:
$E[f(Y_n)] \xrightarrow{n\to\infty} E[f(Y)]$ $\square$

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

Beweis:
"$b) \Rightarrow a)$"
Annahme: $Y_n \xrightarrow{P} Y$ gilt nicht
$\Rightarrow \exists \varepsilon, \delta > 0$ und Teilfolge $n_k : P[|Y_{n_k} - Y| \geq \varepsilon] \geq \delta > 0 \quad \forall k \in \mathbb{N}$
nach $b)$ können wir annehmen, dass Y_{n_k} fast sicher gegen Y konvergiert (durch Übergang zu Unterfolge)
$\Rightarrow Y_{n_k} \xrightarrow{k\to\infty} Y$ P-fast-sicher $\lightning$ Widerspruch zur Annahme
"$a) \Rightarrow b)$"
Y_{n_k} Teilfolge von Y_n
$\Rightarrow P[|Y_{n_k} - Y| \geq \varepsilon] \xrightarrow{k\to\infty} 0 \quad \forall \varepsilon > 0$
$\Rightarrow P[|Y_{n_k} - Y_{n_j}| \geq \varepsilon] \xrightarrow{k,j\to\infty} 0,$
denn $P[|Y_{n_k} - Y_{n_j}| \geq \varepsilon] \leq P[\{|Y_{n_k} - Y \geq \frac{\varepsilon}{2}\} \cup \{|Y_{n_j} - Y| \geq \frac{\varepsilon}{2}\}] \leq P[\{|Y_{n_k} - Y \geq \frac{\varepsilon}{2}\}] + P[\{|Y_{n_j} - Y| \geq \frac{\varepsilon}{2}\}]$

$\Rightarrow \exists k_1 \in \mathbb{N}$ so dass: $P[|Y_{n_k} - Y_{n_{k_1}}| \geq \frac{1}{2}] \leq \frac{1}{2} \qquad \forall k \geq k_1$
$\exists k_2 > k_1$ so dass: $P[|Y_{n_k} - Y_{n_{k_2}}| \geq \frac{1}{2^2}] \leq \frac{1}{2^2} \qquad \forall k \geq k_2$
$\vdots$
$\exists k_j > k_{j-1}$ so dass: $P[|Y_{n_k} - Y_{n_{k_j}}| \geq \frac{1}{2^j}] \leq \frac{1}{2^j} \qquad \forall k \geq k_j \quad \forall j \in \mathbb{N}$
Definiere:

$Z_j := Y_{n_{k_j}}$ Unterfolge von Y_{n_k} $\quad A_j := \{|Y_{n_{k_{j+1}}} - Y_{n_{k_j}}| \geq \frac{1}{2^j}\}$
$\Rightarrow \sum\limits_{j=1}^{\infty} P[A_j] \leq \sum \frac{1}{2^j} < \infty$
$\Rightarrow$ mit Borel-Cantelli gilt, das P-fast-sicher nur endlich viele $A_j,\ j \in \mathbb{N}$, eintreten,
d.h. P["nur endlich viele A_j $j \in \mathbb{N}$ treten ein"] $= 1$
$\Rightarrow P$["$Y_{n_{k_j}}$ ist Cauchyfolge"] $= 1$,
denn für alle $j \leq l,\ j, l \in \mathbb{N}$ gilt:
$|Y_{n_{k_j}} - Y_{n_{k_l}}| \leq \sum\limits_{i=j}^{l-1} |Y_{n_{k_{i+1}}} - Y_{n_{k_i}}| \leq \sum\limits_{i=j}^{l-1} 2^{-i} = 2^{-j+1} \xrightarrow{j\to\infty} 0$
$\Rightarrow \overline{Y} :=$ P-fast-sicher-$\lim\limits_{j\to\infty} Y_{n_{k_j}}$ existiert.
$\Rightarrow Y_{n_{k_j}} \xrightarrow[j\to\infty]{P} \overline{Y}$, $\quad$ ebenso $Y_{n_{k_j}} \xrightarrow[j\to\infty]{P} Y$
$\overset{\text{Übung}}{\Longrightarrow} Y = \overline{Y}$ $\quad$ P-fast-sicher

$\square$

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

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

und insbesondere also:

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

Beweis:
Angenommen $Y_n \xrightarrow{L^1} Y$ gilt nicht, das heisst:
$E[|Y_n - Y|] \nrightarrow 0$

$$\Rightarrow \quad \exists \varepsilon > 0 \text{ und Teilfolge } Y_{n_k}, \text{ so dass } E[|Y_{n_k} - Y|] \geq \varepsilon, \qquad \forall k \qquad (\bigstar)$$

nach Satz 9.75 können wir annehmen, dass $Y_{n_k} \to Y$ P-fast-sicher (durch Übergang zu weiteren Teilfolgen), daher gilt, nach dem "klassischen" Satz der majorisierten Konvergenz, dass $Y_{n_k} \xrightarrow[k\to\infty]{L^1} Y$ $\quad$ Widerspruch zu ($\bigstar$) $\square$

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

Beweis:
$f \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R})$
$\Rightarrow f(Y_n) \xrightarrow{P} f(Y)$, da f stetig, dank Satz 9.75: $|f(Y_n)| < \|f\|_\infty \in L^1(P)$
$\overset{\text{Satz 9.76}}{\Longrightarrow} f(Y_n) \xrightarrow{L^1} f(Y) \Rightarrow E[f(Y_n)] \xrightarrow{n\to\infty} E[f(Y)]$ $\square$

Bemerkung
Die Umkehrung gilt nicht:

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

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
Allgemein gilt:
Falls $Y_n \xrightarrow{P} Y, \quad Y_n \xrightarrow{P} \bar{Y}$ folgt:

$$Y = \bar{Y} \quad P-\text{fast-sicher}$$

denn

$$P\left[|\bar{Y} - Y| > \varepsilon\right] \leq P\left[|Y - Y_n| > \frac{\varepsilon}{2} \text{ und } |\bar{Y} - Y_n| > \frac{\varepsilon}{2}\right] \leq P\left[|Y - Y_n| > \frac{\varepsilon}{2}\right] + P\left[\bar{Y} - Y_n| > \frac{\varepsilon}{2}\right] \underset{n\to\infty}{\searrow} 0$$

$\Longrightarrow \{Y \neq \bar{Y}\} = \bigcup_{n\in\mathbb{N}} \{|Y - \bar{Y}| > \frac{1}{n}\}$ hat P-Wahrscheinlichkeit 0

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.

Beweis:

- Behauptung: Es existiert Teilfolge $(n_k)_{k\in\mathbb{N}}$ so dass

  $$H(q) := \lim_{k\to\infty} F_{n_k}(q) \quad \forall q \in \mathbb{Q}$$

  existiert.

$Q = \{q_m \mid m \in \mathbb{N}\}$ ist abzählbare Menge; da $F_n(q) \in [0,1] \quad \forall q$ gilt, können wir mit dem Satz von Bolzano-Weierstraß eine konvergente Teilfolge wählen, so dass

- $F_{n_k^1}(q_1)$ konvergiert , danach für eine Unterteilfolge $F_{n_k^2}$ von $F_{n_k^1}$
- $F_{n_k^2}(q_1),\ F_{n_k^2}(q_2)$ konvergieren, ... (etc.)
- $F_{n_k^j}$, welche bei $q_1, \dots, q_j$ konvergieren
- wähle also gemäß dem Diagonalargument

$$G_k := F_{n_k^k}$$

$\implies G_k(q)$ konvergiert $\forall q \in \mathbb{Q}$! ✓

- Definiere $H(q) := \lim\limits_{k\to\infty} G_k(q), \quad q \in \mathbb{Q}$

  H ist auch wachsend auf $\mathbb{Q}$, $H(q) \in [0,1]$;

$$F(y) := \lim_{\substack{q \searrow y \\ q \in \mathbb{Q}}} H(q) = \inf_{\substack{q > y \\ q \in \mathbb{Q}}} H(q), \quad y \in \mathbb{R},\ F:\ \mathbb{R} \to [0,1]$$

  $\implies F$ wachsend und rechtsstetig

- Behauptung: $G_k(c) \to F(c) \quad \forall$ Stetigkeitsstellen $c \in \mathbb{R}$ von F

  für c Stetigkeitsstelle, $\varepsilon > 0$ existieren $r, s, r' \in \mathbb{Q}$ mit $r < r' < c < s$ so dass

$$F(c) - \varepsilon \le F(r) \le F(c) \le F(s) \le F(c) + \varepsilon$$

(weil F stetig ist)

$$\implies \overline{\lim_{k\to\infty}}\, G_k(c) \le \overline{\lim_{k\to\infty}}\, G_k(s) \overset{s\in\mathbb{Q}}{=} H(s) \le F(s) \le F(c) + \varepsilon$$

$$\text{und } \underline{\lim_{k\to\infty}}\, G_k(c) \ge \underline{\lim_{k\to\infty}}\, G_k(r') = H(r') \ge F(r) \ge F(c) - \varepsilon, \quad \forall \varepsilon > 0$$

$$\implies G_k(c) \xrightarrow[k\to\infty]{} F(c) \quad ✓$$

□

Definition 9.79
*Sei $(\mu_i)_{i\in I}$ eine beliebige Menge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Die Menge heißt **gleichgradig straff**, falls gilt:*

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

Bemerkung
$(*) \iff \forall \varepsilon > 0\ \exists$ ein Kompaktum $K \subset \mathbb{R}$ (insbesondere gilt also $K \in \mathcal{B}(\mathbb{R})$) so dass:

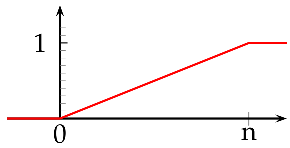
$$\mu_i(K^c) < \varepsilon \quad \forall i \in I$$

Letzteres kann allgemeiner zur Definition von Straffheit für Maße auf topologischen Räumen dienen.

Beispiel (für nicht gleichgradig straffe Folge)

$$\mu_n = \mathcal{U}([0,n]) \quad \text{Gleichverteilung auf } [0,n]$$

Wir bemerken, dass die Verteilungsfunktion der μ_n gegen diejenige des Nullmaßes konvergieren ($F \equiv 0$) (kein Wahrscheinlichkeitsmaß!)



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.

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

Das folgt 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

d) Falls $\mu_n \xrightarrow{\text{w}} \mu$ (schwache Konvergenz) gilt, so ist $(\mu_n)_{n\in\mathbb{N}}$ gleichgradig straff. [→ Übung, nutze Theorem 9.73]

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

$$\mu_n \xrightarrow{w} \mu$$

Beweis:

Seien F_n Verteilungsfunktionen der μ_n

- G_k und F wie aus voriger Proposition 9.78
- ν_k seien die durch G_k festgelegten Maße (via $\nu_k((a, b]) = G_k(b) - G_k(a)$) auf $\mathcal{B}(\mathbb{R})$

Zu zeigen:
F ist Verteilungsfunktion eines Wahrscheinlichkeitsmaßes, also zu zeigen: $F(x) \underset{x\to\infty}{\longrightarrow} 1, \quad F(x) \underset{x\to-\infty}{\longrightarrow} 0$

Für $\varepsilon > 0$ $\exists M > 0$ so dass

$$\mu_n((-M, M]^c) < \varepsilon \quad \forall n$$

Wähle $y > M$, so dass F stetig bei $y, -y$ ist.

$$\implies (1 - F(y)) + F(-y) = \lim_{k\to\infty} \underbrace{(1 - G_k(y) + G_k(-y))}_{\nu_k((-y,y]^c)} \leq \limsup_{k\to\infty} \underbrace{\nu_k((-M, M]^c)}_{<\varepsilon} \leq \varepsilon \quad \forall \varepsilon > 0$$

Wähle μ zu F gehöriges Wahrscheinlichkeitsmaß

$\square$

Satz 9.81 (Stetigkeitssatz von Paul Lévy)
Seien $\mu, \mu_n, n \in \mathbb{N}$ Warscheinlichkeitsmaße 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 Warscheinlichkeitsmaßes* 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.

Beweis:

(a) Sei $X_n \sim \mu_n$

$$\begin{aligned}\varphi_n(u) &= E[\exp(iuX_n)] \\ &= E[\cos(uX_n)] + iE[\sin(uX_n)] \\ &\to E[\cos(uX)] + iE[\sin(uX)] = E[\exp(iuX)] \quad \text{für } X \sim \mu\end{aligned}$$

(b) Nur für $d = 1$ (Fall $d > 1$ vgl. Klenke, Theorem 15.23)

Behauptung: Sei μ Warscheinlichkeitsmaß auf $\mathcal{B}(\mathbb{R}), K > O$, dann: $\frac{1}{K}\int\limits_{-K}^{K}(1-\varphi_\mu(u))\mathrm{d}u \geq \mu\left(]-\frac{2}{K},\frac{2}{K}[^C\right)$,

denn:

$$\begin{aligned}\frac{1}{K}\int\limits_{-K}^{K}(1-\exp(iux))\mathrm{d}u &= \frac{1}{K}(2K - \int\limits_{-K}^{K}(\cos(ux))\mathrm{d}u = \frac{1}{K}(2K - 2\frac{\sin(ux)}{x}) \\ &= 2(1-\frac{\sin(ux)}{Kx}), \quad \text{für } x \neq 0\end{aligned}$$

$$\begin{aligned}\Longrightarrow \frac{1}{K}\int\limits_{-K}^{K}\int\limits_{\mathbb{R}}(1-\exp(iux))\,\mu(\mathrm{d}x)\,\mathrm{d}u &\overset{\text{Fubini}}{=} \int\limits_{\mathbb{R}}\frac{1}{K}\int\limits_{-K}^{K}(1-\exp(iux))\mathrm{d}u\,\mu(\mathrm{d}x) \\ &= 2\int\limits_{\mathbb{R}}\left(1-\frac{\sin(Kx)}{Kx}\right)\mu(\mathrm{d}x) \\ &\geq 2\int\limits_{]-\frac{2}{K},\frac{2}{K}[^C}(1-\frac{1}{|Kx|})\mu(\mathrm{d}x) \\ &\geq \int\limits_{]-\frac{2}{K},\frac{2}{K}[^C} 1\mu(\mathrm{d}x) = \underline{\mu\left(]-\frac{2}{K},\frac{2}{K}[^C\right)}\end{aligned}$$

Behauptung: μ ist gleichgradig stetig

- denn für $\varepsilon > 0$, da ψ stetig in 0, gilt: $\frac{1}{2K}\int\limits_{-K}^{K}(1-\psi(u))\mathrm{d}u \xrightarrow{K\searrow 0} 1-\psi(0) = 0$
- wähle $K > 0$ so dass: $\left|\frac{1}{2K}\int\limits_{-K}^{K}(1-\psi(u))\mathrm{d}u\right| < \varepsilon$

$\Rightarrow$ nach majorisierter Konvergenz gilt: $\exists N \in \mathbb{N}$ so dass $\forall n \geq N$ gilt:
$\left| \frac{1}{2K} \int\limits_{-K}^{K} (1 - \varphi_n(u)) \mathrm{d}u \right| < 2\varepsilon$

$\Rightarrow$ nach Obigem also:
$\mu_n(] - \frac{2}{K}, \frac{2}{K}]^C) \leq 4\varepsilon, \quad \forall n \geq N$

Also gil mit Helly (Satz 9.80) : Es existiert ein Warscheinlichkeitsmaß ν auf $\mathcal{B}^1$ und eine Teilfolge $(\mu_{n_k})_{k \in \mathbb{N}}$, so dass: $\mu_{n_k} \xrightarrow[k \to \infty]{\omega} \nu$

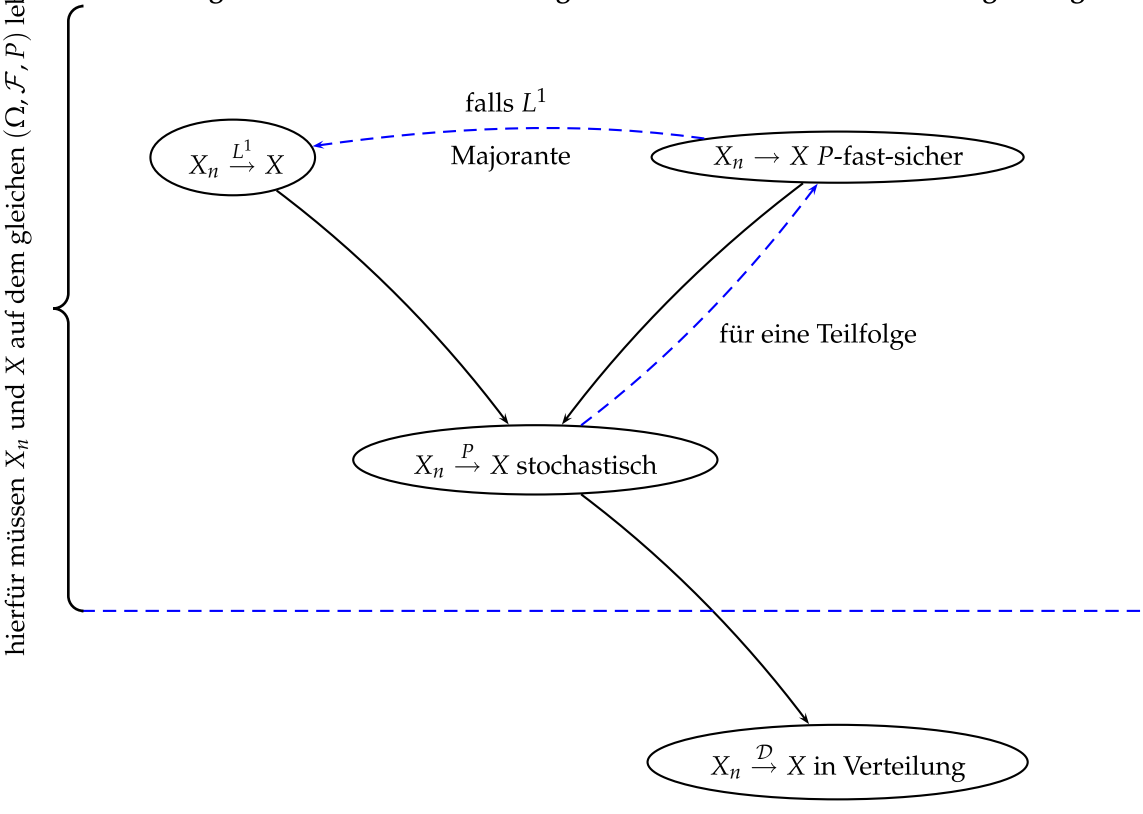
$\varphi_{n_k}(u) \to \varphi_\nu(u)$, $u \in \mathbb{R}$

$\psi = \varphi_\nu$

und jede Teilfolge von $(\mu_n)_{n \in \mathbb{N}}$ hat eine schwach gegen ν konvergierende Unterfolge, daher: $\mu_n \xrightarrow{\omega} \nu$

$\square$

Bemerkung (betreffend Zusammenhänge zwischen verschiedenen Konvergenzbegriffen für Zufallsvariable)



Beispiel (zur Anwendung von Satz 9.81)

(a) $X_n \sim \mathcal{N}(\mu_n, \sigma_n), \quad \mu_n \in \mathbb{R}, \sigma_n > 0$ mit $\mu_n \to \mu \in \mathbb{R}, \sigma_n \to \sigma \geq 0$
$\Rightarrow \varphi_n(u) = \exp(iu\mu_n - \frac{1}{2}u^2\sigma_n^2) \xrightarrow{n \to \infty} \exp(iu\mu - \frac{1}{2}\sigma^2 u^2)$
$X_n \xrightarrow{\mathcal{D}} \mathcal{N}(\mu, \sigma^2)$
Falls $\sigma = 0$ (degenerierter Fall) also $X_n \xrightarrow{\mathcal{D}} \delta_\mu$

(b) $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ und $X_n \xrightarrow{\mathcal{D}} X$
$\Rightarrow \mu = \lim\limits_{n\to\infty} \mu_n$ und $\sigma = \lim\limits_{n\to\infty} \sigma_n$ existieren und $X \sim \mathcal{N}(\mu, \sigma^2)$

(c) Also für $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ gilt:
$X_n \xrightarrow{\mathcal{D}} X \Leftrightarrow \mu = \lim\limits_{n\to\infty} \mu_n, \quad \sigma = \lim\limits_{n\to\infty} \sigma_n$ existieren und $X \sim \mathcal{N}(\mu, \sigma^2)$.
Die Normalverteilungsfamilie ist also abgeschlossen bezüglich Verteilungskonvergenz

Übung:
Zeigen Sie analoge Aussage für Poisson oder exponentialverteilte Zufallsvariablen.

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

denn:
für $X \sim$Poisson(λ) gilt: $\varphi_X(u) = exp(\lambda(e^{iu} - 1))$, damit folgt:

$$\begin{aligned}
\varphi_{Z_n}(u) &= E[\exp(iu\frac{1}{\sqrt{n}})(X_n - n)] \\
&= e^{-iu\sqrt{n}} \varphi_{X_n}(\frac{u}{\sqrt{n}}) \\
&= e^{-iu\sqrt{n}} \exp(n(e^{\frac{iu}{\sqrt{n}}} - 1)) \\
&= \exp\left(-iu\sqrt{n} + iu\sqrt{n} + n\left(-\frac{u^2}{2n} - i\frac{u^3}{3! n^{\frac{3}{2}}} + \frac{n^4}{4! n^2} + \dots\right)\right) \\
&\overset{*}{=} \exp\left(-\frac{u^2}{2} + \underbrace{n \cdot o\left(\frac{1}{n}\right)}_{\to 0 \text{ für } n\to\infty}\right) \xrightarrow{n\to\infty} \exp\left(-\frac{u^2}{2}\right)
\end{aligned}$$

$$\Longrightarrow \quad \varphi_{Z_n}(u) \xrightarrow{n\to\infty} \varphi_Z(u) \text{ für } Z \sim \mathcal{N}(0,1).$$

$*$ weil: $\left|\sum\limits_{k=3}^{\infty} \dfrac{(iu)^k}{k! n^{\frac{k}{2}}}\right| \le \dfrac{1}{n^{\frac{3}{2}}} \sum\limits_{k=3}^{\infty} |\dots| \le \dfrac{1}{n^{\frac{3}{2}}} \underbrace{\left(\sum\limits_{k=3}^{\infty} \dfrac{|u|^k}{k!}\right)}_{<\infty} \Rightarrow n\left(\sum\limits_{k=3}^{\infty} \dfrac{(iu)^k}{k! n^{\frac{k}{2}}}\right) \xrightarrow{n\to\infty} 0$

Beispiele (zur Motivation des zentralen Grenzwertsatzes:)

- Aus vorigem Beispiel:
  Seien $(Y_n)_{n\in\mathbb{N}}$ i.i.d. (identisch, unabhängig verteilt)$\sim$Poisson(1)
  $\Rightarrow X_n := \sum\limits_{j=1}^{n} Y_j \sim$Poisson$(n)$
  denn: $\bigotimes\limits_{j=1}^{n}$Poisson$(1)$ $=$Poisson(n)
  $\Rightarrow Z_n = \dfrac{X_n - n}{\sqrt{n}} = \dfrac{\sum\limits_{j=1}^{n} Y_j - nE[Y_1]}{\sqrt{nV(X_1)}} \xrightarrow{n\to\infty} \mathcal{N}(0,1)$

- Betrachte i.i.d. Folge von Zufallsvariablen $(Y_n)_{n\in\mathbb{N}}$ mit $P[Y_j = +1] = 1 - P[Y_j = -1] = \frac{1}{2}$
  $Z_n = \frac{\sum_{j=1}^n Y_j - 0}{\sqrt{n}}$ hat die charakteristische Funktion

$$\begin{aligned}
\varphi_{Z_n}(u) &= \left(\varphi_{Y_1}\left(\frac{u}{\sqrt{n}}\right)\right)^n \\
&= \left(\frac{1}{2}\left(e^{\frac{iu}{\sqrt{n}}} + e^{-\frac{iu}{\sqrt{n}}}\right)\right)^n \\
&= \left(\cos\left(\frac{u}{\sqrt{n}}\right)\right)^n \\
&= \exp\underbrace{\left(n \cdot \ln\left(\cos\left(\frac{u}{\sqrt{n}}\right)\right)\right)}_{\xrightarrow[n\to\infty]{\text{l'Hospital}} -\frac{u^2}{2}} \xrightarrow[n\to\infty]{} \exp\left(-\frac{u^2}{2}\right) = \varphi_Z(u)
\end{aligned}$$

$$\implies Z_n \xrightarrow{\mathcal{D}} Z \sim \mathcal{N}(0,1)$$

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 (gleichverteilter?) 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:

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

Hilfslemma 9.83
Sei $c_n \in \mathbb{C}$ mit $c_n \xrightarrow[n\to\infty]{} c$ mit $c > 0$ (d.h. $c \in \mathbb{R}^+$). Dann gilt:

$$\left(1 - \frac{c_n}{n}\right)^n \xrightarrow[n\to\infty]{} e^{-c}$$

Beweis: [von Satz 9.82 (zentraler Grenzwertsatz)]
X_n hat endliche zweite Momente $\implies$ die charakteristische Funktion $\varphi := \varphi_{X_n-\mu}$ ist zweimal stetig differenzierbar

$\Longrightarrow$ (Taylorentwicklung)

$$\varphi(x) = \varphi(0) + \varphi'(0) \cdot x + \varphi''(0) \cdot \frac{x^2}{2} + x^2 \cdot h(x)$$

mit $h(x) \xrightarrow[x \to 0]{} 0$

$\Longrightarrow$

$$\varphi_{S_n^*}(u) = \left(\varphi_{X_1-\mu}\left(\frac{u}{\sigma\sqrt{n}}\right)\right)^n = \left(\varphi(0) + \varphi'(0) \cdot \frac{u}{\sigma\sqrt{n}} + \varphi''(0) \cdot \frac{1}{2} \cdot \frac{u^2}{\sigma^2 \cdot n} + \frac{u^2}{\sigma^2 \cdot n} \cdot h\left(\frac{u}{\sigma\sqrt{n}}\right)\right)^n$$

hierbei ist

$$\begin{aligned}
&\varphi'(u) = i \cdot E[(X_1 - \mu) \cdot \exp(iu(X_1 - \mu))], \\
&\varphi'(0) = 0, \\
&\varphi''(u) = (-1)E[(X_1 - \mu)^2 \cdot \exp(iu(X_1 - \mu))] \\
&\varphi''(0) = -\sigma^2
\end{aligned}$$

$\Longrightarrow$

$$\begin{aligned}
\varphi_{S_n^*}(u) &= \left(\underbrace{\varphi(0)}_{=1} + \underbrace{\varphi'(0)}_{=0} \cdot \frac{u}{\sigma\sqrt{n}} + \underbrace{\varphi''(0) \cdot \frac{1}{2} \cdot \frac{u^2}{\sigma^2 \cdot n}}_{-\frac{1}{2}\cdot\frac{u^2}{n}} + \frac{u^2}{\sigma^2 \cdot n} \cdot h\left(\frac{u}{\sigma\sqrt{n}}\right)\right)^n \\
&= \left(1 - \frac{1}{2} \cdot \frac{u^2 + \overbrace{2\frac{u^2}{\sigma^2} \cdot h\left(\frac{u}{\sigma\sqrt{n}}\right)}^{\xrightarrow{n\to\infty} 0}}{n}\right)^n \xrightarrow[\text{Hilfslemma 9.83}]{n\to\infty} \mathrm{e}^{-\frac{u^2}{2}}
\end{aligned}$$

nach dem Levy'schen Stetigkeitssatz folgt:

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

□

Beweis: [des Lemmas 9.83]

a) Behauptung:
Für $z_1, \ldots, z_n$ und $w_1, \ldots, w_j$ in $\mathbb{C}$ mit $|z_j| \leq 1,\ |w_j| \leq 1\ \forall j$ gilt:

$$\left|\prod_{j=1}^n z_j - \prod_{j=1}^n w_j\right| \leq \sum_{j=1}^n |z_j - w_j|$$

denn per Induktion sehen wir:

IA: ✓

IS: $n \curvearrowright n+1$:
Betrachte

$$\begin{aligned}
\left|\prod_{j=1}^{n+1} z_j - \prod_{j=1}^{n+1} w_j\right| &\leq \underbrace{\left|\prod_{j=1}^{n+1} z_j - z_{n+1}\prod_{j=1}^{n} w_j\right|}_{=|z_{n+1}|\cdot\left|\prod_{j=1}^n z_j - \prod_{j=1}^n w_j\right|} + \underbrace{\left|z_{n+1}\prod_{j=1}^{n} w_j - \prod_{j=1}^{n+1} w_j\right|}_{=\left|\prod_{j=1}^n w_j\right|\cdot|z_{n+1}-w_{n+1}|} \\
&\overset{|w_j|,|z_j|\leq 1}{\leq} \underbrace{\left|\prod_{j=1}^{n} z_j - \prod_{j=1}^{n} w_j\right|}_{\leq \sum_{j=1}^n |z_j - w_j|} + |z_{n+1} - w_{n+1}| \qquad \checkmark
\end{aligned}$$

b) Behauptung:
Für $b \in \mathbb{C}$ mit $|b| \leq 1$ gilt:

$$\left|e^{-b} - (1-b)\right| \leq |b|^2$$

denn:

$$\begin{aligned}\left|e^{-b} - (1-b)\right| &\leq \left|\sum_{k=2}^{\infty} \frac{(-1)^k \cdot b^k}{k!}\right| \\ &\leq \sum_{k=2}^{\infty} \left|\frac{(-1)^k \cdot b^k}{k!}\right| \\ &\leq |b|^2 \cdot \sum_{k=2}^{\infty} \left(\frac{1}{2}\right)^k \\ &\leq |b|^2\end{aligned}$$

c) Behauptung:
Sei $c_n \in \mathbb{C},\ \ c_n \to c,\ \ \mathbb{R} \ni c > 0$, dann gilt:

$$\left(1 - \frac{c_n}{n}\right)^n \xrightarrow[n\to\infty]{} e^{-c}$$

denn:

$$\left|\left(1 - \frac{c_n}{n}\right)^n - e^{-c_n}\right| \overset{\text{Teil a)}}{\leq} \sum_{j=1}^{n} \left|1 - \frac{c_n}{n} - e^{-\frac{c_n}{n}}\right|$$

für n groß genug (so dass $\left|1 - \frac{c_n}{n}\right| \leq 1$ und $\left|e^{-\frac{c_n}{n}}\right| \leq 1$).
Das kann weiter abgeschätzt werden mittels Teil b) durch:

$$\left|\left(1 - \frac{c_n}{n}\right)^n - e^{-c_n}\right| \leq \sum_{j=1}^{n} \left|\frac{c_n}{n}\right|^2 = \frac{1}{n}\,|c_n|^2 \xrightarrow[n\to\infty]{} 0$$

$$\Longrightarrow \left(1 - \frac{c_n}{n}\right)^n \xrightarrow[n\to\infty]{} e^{-c}$$

$$\left(\text{mit Dreiecksungleichung, wo } \left|e^{-c_n} - e^{-c}\right| \xrightarrow[n\to\infty]{} 0 \text{ klar}\right)$$

$\square$

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

$$S_n \equiv \frac{1}{n} \sum_{j=1}^{n} X_j \xrightarrow[n\to\infty]{} \mu = E[X_1] \qquad P\text{-fast-sicher}$$

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

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

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

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

In diesem Sinne gilt: die Konvergenzrate ist $\sqrt{n}$!

Bemerkung
Es gibt Verallgemeinerungen des zentralen Grenzwertsatzes:

→ die Annahme der "identischen Verteilungen" kann abgeschwächt werden; wichtig für Anwendungen, da dieses Kriterium meist nicht erfüllt ist

→ sehr genaue Bedingungen für zentrale Grenzwertsatzaussage ("Lindeberg-Bedingungen"), vgl. Klenke

→ es gibt auch eine mehrdimensionale Version (→ Klenke)

→ Beispiel für Anwendung des zentralen Grenzwertsatzes:
Bestimmung von "Vertauensbereichen" (Konfidenzbereichen) für unbekannte Verteilungsparameter, die aus Daten geschätzt werden (Übungsblatt 11)

Erinnerung:
Seien $X_1, X_2, \ldots$ i.i.d. $\mathbb{R}$-wertige Zufallsvariablen. Dann ist die empirische Verteilungsfunktion F_n (bzgl. $X_1, \ldots, X_n$) definiert als:

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{X_k \leq x\}}$$

Wir wissen von Glivenko-Cantelli,

$$\sup_{x \in \mathbb{R}} |F(x) - F_n(x)| \xrightarrow[n \to \infty]{} 0 \quad P\text{-fast-sicher}$$

dass die empirische Verteilungsfunktion F_n gleichmässig fast-sicher gegen die Verteilungsfunktion F der X_k konvergiert.

Frage: Wie schnell konvergiert $F_n(x)$ gegen $F(x),\ x \in \mathbb{R}$? → Wende zentralen Grenzwertsatz an!

Antwort: mit Rate $\sqrt{n}$, denn für $x \in \mathbb{R}$, und $Y_k = \mathbb{1}_{\{X_k \leq x\}}$ i.i.d. Bernoulli($F(x)$)

$$\begin{aligned} \sqrt{n}\,(F_n(x) - F(x)) &= \sqrt{n} \left(\frac{1}{n} \left(\sum_{k=1}^{n} Y_k - E[Y_1] \right) \right) \\ &= \underbrace{\frac{\left(\frac{1}{n} \sum_{k=1}^{n} Y_k\right) - E[Y_1]}{\sqrt{n \cdot V(Y_1)}}}_{\xrightarrow[\text{zent. Grenzwertsatz}]{} \mathcal{N}(0,1)} \cdot \sqrt{V(Y_1)} \xrightarrow[n \to \infty]{\text{zent. Grenzwertsatz}} \mathcal{N}(0, \sigma^2) \end{aligned}$$

mit $\sigma^2 = V(Y_1) = F(x)\,(1 - F(x))$

Bemerkung
Mann kann zeigen:
Die Differenz der empirischen zur theoretischen Verteilungsfunktion konvergiert bei Skalierung mit $\sqrt{n}$ gegen eine "Brown'sche Brücke", vgl. mathematische Statistik/ stochastische Analysis.

Bemerkung
Beachte, dass σ^2 maximal für $F(x)$ nahe $\frac{1}{2}$ wird!

10 Grundbegriffe der Schätztheorie

[aus der mathematischen Statistik]

Definition 10.84 (statistisches Modell)
Sei $\mathcal{X}$ eine beliebige Menge, $\mathcal{F}$ eine σ-Algebra auf $\mathcal{X}$, 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**.*

Definition 10.85
*Für eine zu schätzende Funktion $g: \quad \Theta \to \mathbb{R}$ heißt eine messbare Abbildung $T: \quad \mathcal{X} \to g(\Theta) \equiv \operatorname{Im}(g)$ ein **Schätzer** von $g(\vartheta)$.*

Beispiel
Gegeben: Teich mit unbekannter Anzahl N von Fischen
Ziel: N schätzen!
Wir fangen N_w Fische, die wir mit weißem Punkt markieren und wieder aussetzen;
Tage später fangen wir wieder n Fische, und beobachten, dass davon k eine weiße Markierung haben.

statistisches Modell:

$$\mathcal{X} = \{0, 1, \ldots, n\} \quad \text{oder} \quad \mathcal{X} = \mathbb{N}_0$$
$$\Theta = \mathbb{N},$$

wollen schätzen: $g(\vartheta) = \vartheta$, d.h. $g = \mathrm{Id}$
für $\vartheta = N$:

$$P_\vartheta \equiv P_N = \begin{array}{l}\text{Verteilung der markierten Fische}\\ \text{in dem Fang mit Größe n, wenn}\\ N \text{ Fische im Teich sind}\end{array}$$

also

$$P_N(\{k\}) = \frac{\binom{N_w}{k} \cdot \binom{N-N_w}{n-k}}{\binom{N}{n}}$$

d.h. hypergeometrische Verteilung. Wir schätzen nun N nach dem Maximum-Likelihood-Prinzip, als dasjenige N, welches $P_N(\{k\})$ maximiert. Hierzu:

$$\begin{aligned}\frac{P_N(\{k\})}{P_{N-1}(\{k\})} &= \frac{\cancel{\binom{N_w}{k}} \cdot \binom{N-N_w}{n-k}}{\binom{N}{n}} \cdot \frac{\binom{N-1}{n}}{\cancel{\binom{N_w}{k}} \cdot \binom{N-1-N_w}{n-k}} \\ &= \frac{N-n}{N} \cdot \frac{N-N_w}{N-N_w-n+k}\end{aligned}$$

$$\begin{aligned}\implies \frac{P_N(\{k\})}{P_{N-1}(\{k\})} > 1 &\iff (N-n)\cdot(N-N_w) > N(N-N_w-n+k) \\ &\iff -n(N-N_w) > N(-n+k) \\ &\iff \frac{nN_w}{k} > N \\ &\iff N^* = \lfloor \frac{nN_w}{k} \rfloor\end{aligned}$$

wobei $\lfloor g \rfloor := \max\{m \in \mathbb{Z} \mid m \leq g\}$ (kleinste Zahl, die größer ist; sogenannte Gauß-Klammern).

Also wollen wir als Maximum-Likelihood-Schätzer für N gerade $T := \widehat{N} := \lfloor \frac{nN_w}{k} \rfloor$.

<u>zur Definition von ML-Schätzer:</u>
Sei μ ein Maß auf $(\mathcal{X}; \mathcal{F})$ so dass

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

$$(\text{“}\mathrm{d}P_\vartheta = \rho_\vartheta \mathrm{d}\mu\text{”})$$

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*** $g(\vartheta)$

Bemerkung

1. Der Maximum-Likelihood-Schätzer braucht im Allgemeinen nicht eindeutig zu sein.
2. Die mathematische Statistik zeigt: Maximum-Likelihood-Schätzer haben oft (asymptotisch) gute Eigenschaften
3. Obige Definitionen a) bis d) sind bzgl. dem Maß μ; Kanonisch sind meist die Fälle:
   a) $\mu = \sum_{x \in \mathcal{X}} \delta_x$ Zählmaß, falls $\mathcal{X}$ abzählbar
   b) $\mu =$ Lebesgue-Maß, falls $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{X} \in \mathcal{B}(\mathbb{R}^d)$ und $\mathcal{F}$ die Spur-σ-Algebra auf $\mathcal{X}$

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

$$L_k(\vartheta) = \rho_\vartheta(k) = \binom{n}{k} \vartheta^k (1 - \vartheta)^{n-k}$$

Betrachte den Fall $n \neq k \neq 0$ (sonst Sonderbetrachtung)

$$\begin{aligned} L_k'(\vartheta) &= \binom{n}{k} \cdot \left(k\vartheta^{k-1}(1-\vartheta)^{n-k} - \vartheta^k (n-k)(1-\vartheta)^{n-k-1} \right) \\ &= \binom{n}{k} \cdot \vartheta^{k-1}(1-\vartheta)^{n-k-1} \left(k(1-\vartheta) - (n-k)\vartheta \right) \end{aligned}$$

$L_k'(\vartheta) \stackrel{!}{=} 0$ liefert $\hat{\vartheta} = \dfrac{k}{n}$ als Maximum-Likelihood-Schätzer für $\vartheta = p$, denn Maximum für $k \neq 0, n$ kann nicht bei $\vartheta = 0$ oder $\vartheta = 1$ liegen, und

$$\begin{aligned} & & k(1-\vartheta) - (n-k)\vartheta &= 0 \\ &\iff & k - n\vartheta &= 0 \\ &\iff & \vartheta &= \frac{k}{n} \end{aligned}$$

Erinnerung

$(\underset{\downarrow}{\overset{\text{Stichprobenraum}}{\mathcal{X}}}, \mathcal{F}, \{P_\vartheta\}_{\vartheta \in \Theta})$
$X := Id_{\mathcal{X}}$,
$X : x \mapsto x, \quad x \hat{=}$Ihre Beobachtung

Maximum-Likelihood-Ansatz für Schätzer $\hat{\vartheta}$ für ϑ:

$$\hat{\vartheta}(x) = \underset{\vartheta \in \Theta}{\operatorname{argmax}}\, \rho_\vartheta(x) = \underset{\vartheta \in \Theta}{\operatorname{argmax}}\, L_x(\vartheta)$$

wobei

$$dP_\vartheta = \rho_\vartheta d\mu$$

2 typische Fälle für μ

A) $\mathcal{X}$ abzählbar, $\mathcal{F} = \mathcal{P}(\mathcal{X}) \equiv 2^{\mathcal{X}}$, $P_\vartheta(A) = \sum\limits_{x \in A} \rho_\vartheta(x), \quad A \subset \mathcal{X}$
$\implies dP_\vartheta = \rho_\vartheta d\mu$ bzgl. Zählmaß $\mu(A) = \sum\limits_{x \in \mathcal{X}} \delta_x(A)$

B) $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{F}$ = Spur-σ-algebra von $\mathcal{B}^d(\mathbb{R})$ auf $\mathcal{X}$, [z.B.:$\mathcal{X} = \mathbb{R}^d, \mathcal{F} = \mathcal{B}^d$]
$(\mathcal{X} \in \mathcal{B}^d(\mathbb{R}))P_\vartheta$ absolutstetig bzgl. Lebesguemaß $\quad \mu : dP_\vartheta = \underset{\text{Dichte der absolutstetigen Verteilung } P_\vartheta}{\underset{\uparrow}{\rho_\vartheta}} d\mu$

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

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 Maximum-Likelihood-Schätzer für Erfolgswahrscheinlichkeit $\rho = \vartheta \in [0,1]$ bei n Münzwürfen mit einer möglicherweise unfairen Münze.
$\hat{g}(x) = \frac{x}{n}, \quad x =$# von Zahl, $\quad X(x) = x \overset{P_\vartheta}{\sim}$ Binomial(n, ϑ)
$\implies E_\vartheta\left[\frac{X}{n}\right] = \frac{1}{n} n\vartheta = \vartheta \quad \forall \vartheta \in \Theta$
$\implies$ erwartungstreu

Definition 10.88
*Eine Folge von Schätzern $\hat{g}_n$ für $g(\vartheta), \vartheta \in \Theta$ heißt **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 "im Allgemeinen" eine konsistente Folge liefern (vgl. Georgii, Satz 7.30).

Beispiel (wieder der unfaire Münzwurf)
Seien $X_1, X_2, \dots$ i.i.d. Bernoulli(ϑ) (d.h. $X_i \in \{0,1\}$), $\vartheta \in [0,1] =: \Theta$

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

$(\vartheta_n)_{n\in\mathbb{N}}$ konsistente Schätzerfolge.

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 $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ ein erwartungstreuer Schätzer, da $E_\vartheta[\bar{X}_n] = \frac{1}{n} \cdot n \cdot m_\vartheta$, und konsistent, da

$$\bar{X}_n \overset{n\to\infty}{\longrightarrow} m_\vartheta \quad P\text{-fast sicher}$$

nach Gesetz der großen Zahlen.

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 $\sigma^2\vartheta$, denn für ϑ beliebig, aber fix:

$$
\begin{aligned}
E_\vartheta[s_n^2] &= \frac{1}{n-1} E_\vartheta\left[\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right] \overset{Y_i := X_i - m_\vartheta}{=} \frac{1}{n-1} E_\vartheta\left[\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2\right] \\
&= \frac{1}{n-1} E_\vartheta\left[\sum_{i=1}^{n}\left(Y_i^2 - 2Y_i \frac{1}{n}\sum_{j=1}^{n} Y_j\right) + \frac{1}{n}\sum_{i,j=1}^{n} Y_i Y_j\right] \\
&= \frac{1}{n-1}\left(n\sigma_\vartheta^2 - \cancel{n}\cdot 2\frac{1}{\cancel{n}}\sigma_\vartheta^2 + \frac{1}{\cancel{n}}\cancel{n}\sigma_\vartheta^2\right) = \sigma_\vartheta^2 \quad \checkmark
\end{aligned}
$$

Bemerkung
$\bar{X}_n$ und s_n^2 sind auch konsistent, vgl. für s_n^2 (vgl. Georgii, Übung)

Beispiel (Maximum-Likelihood-Schätzer im absolutstetigem Fall)
$X_1, \ldots, X_n$ 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 $X = (X_1, \ldots, X_n)$ bzgl. Lebesguemaß μ

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

Fall 1 $\sigma = \sigma_0$ bekannt, $m \in \mathbb{R}$ unbekannt

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

$\mathcal{L}_x(\vartheta) = \log(L_x(\vartheta)) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2$

$\frac{\partial}{\partial m}\mathcal{L}_x(\vartheta) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - m) \overset{!}{=} 0$

$\implies \widehat{m} = \frac{1}{n}\sum_{i=1}^{n} x_i$ maximiert $\mathcal{L}_x((m, \sigma_0))$, ist Maximum-Likelihood-Schätzer!

Fall 2 σ und m unbekannt

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

Als Maximum-Likelihood-Schätzer errechnen sich ($\frac{\partial}{\partial m}\mathcal{L}_x(\vartheta) \overset{!}{=} 0$, $\frac{\partial}{\partial \sigma}\mathcal{L}_x(\vartheta) \overset{!}{=} 0 \implies \ldots$) analog zu Fall 1:

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

und

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

(nicht erwartungstreu, vgl. oben jedoch "asymptotisch erwartungstreu")

Bemerkung
Statt "Punktschätzer" für unbekannte Parameter betrachtet man auch "Bereichsschätzer" (Vertauensbereiche, Konfidenzintervalle) (→ letzte Übungsaufgabe)

Elementare Testtheorie

Idee: Teste (Null)Hypothese(Θ_0) vs. 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 $P_0 \equiv P_{\vartheta_0}$?
((gegen bzw. für) $P_1 \equiv P_{\vartheta_1}$?)

Definition 10.89
$(\mathcal{X}, \mathcal{F}, P_\vartheta)$, $\Theta = \Theta_0 \dot\cup \Theta_1$.
Ein Test von Nullhypothese Θ_0 vs. Alternative Θ_1 ist gegeben durch einen Ablehnungsbereich $R \subset \mathcal{X}(R \in \mathcal{F})$.

$$\begin{aligned} x \in R &\curvearrowright \Theta_0 \quad \underline{\textit{abgelehnt}} \\ x \notin R &\curvearrowright \Theta_0 \quad \underline{\textit{nicht abgelehnt}} \end{aligned}$$

Beispiel
Medikamententest:

$$\begin{aligned} &\Theta_0 \widehat{=} \text{Medikament wirkt wie Placebo} \\ &\Theta_0 \widehat{=} \text{Medikament wirkt (besser)} \end{aligned}$$

Fehlerwahrscheinlichkeiten eines Tests

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

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

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

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

Im Weiteren:

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

Definition 10.90
der Test R (von engl. rejection) für Θ_0 vs. Θ_1 hat ***(Irrtums)niveau*** *$\alpha \in [0,1]$, falls*

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

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

Bemerkung

- typisch z.B. $\alpha = 1\%$,
  Niveau 1% sichert, dass die Wahrscheinlichkeit für Fehler 1. Art kleiner als 1% ist.
- Die Macht gibt an, wie wahrscheinlich sich die Alternative unter einem Maß P_{ϑ_1} aus der Alternative $\vartheta_1 \in \Theta_1$ "durchsetzt".

Satz 10.91 (und Definition, Newmann-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 gegebenem 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 $R := R_c := \{q \geq c\}$, $c > 0$, ***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. $P_1(R') \leq P_1(R)$*

Beweis:
nach Konstruktion gilt:

- $\alpha = P_0(R) = P_0(R \cup R') + P(R \setminus R')$
- $\alpha \geq P_0(R') = P_0(R \cup R') + P_0(R' \setminus R)$

$$\begin{aligned} \Longrightarrow \quad & P_0(R \setminus R') \geq P_0(R' \setminus R) \qquad (*) \\ \Longrightarrow \quad & P_1(R') - P_1(R) = \int_{\mathcal{X}} (\mathbb{1}_{R'} - \mathbb{1}_R) \rho_1 \mathrm{d}\mu \\ & = \int_{\mathcal{X}} \left(\mathbb{1}_{R'} - \mathbb{1}_{\{q \geq c\}} \right) \underbrace{\frac{\rho_1}{\rho_0}}_{=q} \cdot \rho_0 \mathrm{d}\mu \\ & = \int_{\mathcal{X}} \Bigg(\underbrace{q \cdot \mathbb{1}_{R' \setminus \{q \geq c\}}}_{\leq c \mathbb{1}_{R' \setminus \{q \geq c\}}} - \underbrace{q \cdot \mathbb{1}_{\{q \geq c\} \setminus R'}}_{\geq c \mathbb{1}_{\{q \geq c\} \setminus R'}} \Bigg) \rho_0 \mathrm{d}\mu \\ & \leq c \left(P_0(R' \setminus R) - P_0(R \setminus R') \right) \overset{(*)}{\leq} 0 \\ \Longrightarrow \quad & P_1(R') \leq P_1(R) \end{aligned}$$

□

Beispiel
Betrachte $X_1, \dots, X_n$ 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:

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

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

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

$\Longrightarrow$ optimale Tests haben die Form

$$\begin{aligned} R = \{x \mid q(x) \geq c_1\} &\overset{\downarrow c_1 = \exp\left(c_2 - \frac{1}{2}nm^2\right)}{=} \{x \mid \langle x, m\cdot\mathbb{1}\rangle \geq c_2\}, \quad c_i > 0, \\ &= \{x \mid \sum_{i=1}^n x_i \cdot m \geq c_2\} \\ &= \{x \mid m\sum_{i=1}^n x_i \geq c_2\} \end{aligned}$$

Das **Niveau** von R unter P_0:

$$\begin{aligned} m\sum_{i=1}^n X_i &\sim \mathcal{N}(0, m^2 n) \text{ unter } P_0 \\ \Longrightarrow \quad P_0(R) &= 1 - \Phi\left(\frac{c_2}{\sqrt{nm^2}}\right) \\ &= \Phi\left(-\frac{c_2}{\sqrt{nm^2}}\right), \end{aligned}$$

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

$$\begin{aligned} m\sum_{i=1}^n X_i &\sim \mathcal{N}(m^2 n, m^2 n) \text{ unter } P_1 \\ \Longrightarrow \quad P_1(R) &= 1 - \Phi\left(\frac{c_2 - m^2 n}{\sqrt{nm^2}}\right) \\ &= \Phi\left(-\frac{c_2 - m^2 n}{\sqrt{nm^2}}\right) \end{aligned}$$

Bemerkung
Für vorgegebenen Niveauparameter $\alpha \in (0,1)$ kann man $c_2 = c_2(\alpha)$ so bestimmen, dass der Test $R = R(\alpha)$ das Niveau $P_0(R) = \alpha$ hat ($\rightarrow$ Übung).

Index

σ-Algebra
    Definition, 9
    Eigenschaften, 9
    kleinste, Definition, 10
σ-Stetigkeit, 13
σ-Subadditivität, 13
σ-endlich
    Definition, 18
Übereinstimmung auf Erzeuger, 20

abgeschlossen
    bzgl. Differenzen, 19
    unter Limites, 19
Additivität
    endliche, 13
Algebra
    Definition, 18
asymptotisch
    Definition, 38

Bayes-Formel, 30
Bernoulli-Schema, 15
Borel-σ-Algebra
    Definition, 10
    Erzeuger, 10
Borel-Cantelli-Lemma, 39

Cantormenge, 7
Caratheodory
    Forstetzungssatz von, 20
charakteristische Funktion
    affiner Transformationen, 57
    Definition, 54
charakteristische Transformierte
    Definition, 54

degeneriert
    Definition, 63
Dichte
    Definition, 24
    Radon-Nykodym, 27
Dirac'sches Punktmaß, 24

Eindeutigkeitssatz, 20
empirische Verteilungsfunktion
    Definition, 51
Ereignisraum, 9
erwartungstreu
    Definition, 86
Erwartungswert
    Definition, 40

Faltung
    Definition, 58
Faltungsprodukt
    Definition, 58
Fatou
    Lemma von, 27

Gesetz der großen Zahlen
    schwaches, 46
    starkes, 49
gleichgradig straff
    Definition, 74
Glivenko-Cantelli
    Satz von, 52

Helly'sches Selektionsprinzip, 75

Indikatorfunktion, 13
Irrtumsniveau
    Definition, 89

Klassentheorem
    monotones, 19
Koch'sche Schneeflocke, 7
Kolmogorov
    0-1 Gesetz von, 38
Kolmogorov'sche Axiome, 9
konsistent
    Definition, 86
Konvergenz
    P-fast-sichere Konvergenz, Definition, 47
    Konvergenz in Verteilung, Definition, 69
    schwache Konvergenz, Definition, 69
    stochastische Konvergenz, Definition, 46
Korrelationskoeffizient
    Definition, 44
Kovarianz
    Definition, 42

Lévy
    Stetigkeitssatz von Paul Lévy, 76
Lebesguemaß, 22
Likelihood-Funktion
    Definition, 84
Likelihoodquotient
    Definition, 89
Log-Likelihood-Funktion
    Definition, 84

Münzwurf
    ∞-facher, 3
    n-facher, 3, 15
    einfacher, 3
Maß
    Definition, 18
Maßintegral
    Hauptresultate, Konvergenzsätze, 26
Macht
    Definition, 89
Maximum-Likelihood-Schätzer

Definition, 84
meßbarer Raum, 9
Moment
    Definition, 55
Monotonie, 13
Multiplikationsformel, 30

Newmann-Pearson
    Satz von, 89
Niveau
    Definition, 89

Poisson'scher Grenzwertsatz, 17
Potenzmenge
    Definition, 4
Prämaß
    Definition, 18
Produkt-σ-Algebra
    Definition, 12
Produktmaß
    diskretes
        Definition, 15

Riemann-Integral, 27

Satz über majorisierte Konvergenz, 27
Satz über monotone Konvergenz, 27
Schätzer
    Definition, 83
Spur-σ-Algebra
    Definition, 12
Standardabweichung
    Definition, 42
statistisches Modell
    Beispiel, 83
    Definition, 83
stetig
    absolut, 24

Test
    Definition, 88

Unabhängigkeit
    Definition, 31
Ungleichung
    Cauchy-Schwarz'sche Ungleichung, 41
    Hölder'sche Ungleichung, 41
    Jensen'sche Ungleichung, 42
    Marko'vsche Ungleichung, 41
    Minkowski Ungleichung, 41
    Tschebyschev-Ungleichung, 41
unkorreliert
    Definition, 42

Varianz
    Definition, 42
Verteilung
    Binomialverteilung, 16
    diskrete
        Beispiel, 15, 16
    empirische Verteilungsfunktion, Definition, 51
    Exponentialverteilung, 25
    Gamma-Verteilung, 25
    geometrische, 16
    multivariate Normalverteilung, Definition, 63
    Normalverteilung, 25
    Poisson-Verteilung, 16
    von Zufallsvariablen, 18
Verteilungsfunktion
    Beispiel, 24
    Definition, 22
    Eigenschaften, 22
Vitali
    Satz von, 4

Würfeln
    n-faches, 3
    einfaches, 3
Wahrscheinlichkeit
    bedingte, 29
    Formel totaler Wahrscheinlichkeit, 30
Wahrscheinlichkeitsmaß
    Definition, 9, 18
    Eigenschaften, 13
Wahrscheinlichkeitsraum
    Definition, 9
    diskreter, 14

Zähldichte, 14
zentraler Grenzwertsatz, 79
Zufallsvariable
    Gauß'sche Zufallsvariable, Definition, 63
Zufallvariable
    Definition, 17