Humboldt-Universität zu Berlin
Institut für Mathematik
Sommersemester 2010

# Stochastik I

**Prof. Becherer**

Bodo Graumann

19. Mai 2014



XeTeX Der Code wurde mit gvim sowie vim-latex erstellt und mit xelatex kompiliert – all das auf Gentoo Linux. Meinen Dank an die Freie Software Community und die TeX-Kollegen auf TeX.SX für ihre Hinweise und Unterstützung.
Bitte schreibt mir eure Kommentare und Verbesserungsvorschläge zu diesem Dokument! Ihr könnt mir entweder direkt mailen oder das Kontaktformular auf meiner Internetseite benutzen.

# Inhaltsverzeichnis

# Literatur

[Bau08] BAUM, Prof. H.: *Maßtheorie Skript*. http://www-irm.mathematik.hu-berlin.de/%7Ebaum/Skript/MIT-SS08.pdf, 2008

[Els07] ELSTRODT, J.: *Maß- und Integrationstheorie*. Springer, 2007

[Geo07] GEORGII, H.-O.: *Stochastik*. De Gruyter Verlag, 2007

[Kle08] KLENKE, A.: *Wahrscheinlichkeitstheorie*. Springer, 2008

[Kre05] KRENGEL, U.: *Einführung in die Wahrscheinlichkeitstheorie & Statistik*. 8. Vieweg, 2005

[Shi95] SHIRYAEV, Albert N.: *Probability*. Springer, 1995

[Sim82] SIMMONS, George F.: *Introduction to Topology and Modern Analysis*. Krieger Pub Co, 1982

[Sti03] STIRZAKER, D.: *Elementary Probability*. 2. Cambridge University Press, 2003

[Str85] STRASSER, H.: *Mathematical Theory of Statistics*. De Gruyter Verlag, 1985

Die Stochastik besteht im wesentlichen aus zwei Teile, die jedoch stark in einander greifen. Zum einen die Wahrscheinlichkeitstheorie, welche sich mit Wahrscheinlichkeitsmodellen beschäftigt und zum anderen die Statistik, die Häufigkeiten in der Praxis betrachtet.

Zuersteinmal wollen wir erklären, was ein Wahrscheinlichkeitsraum ist. Sei also ein Wahrscheinlichkeitsraum gegeben:

$$(\Omega, \mathcal{F}, \mathbb{P})$$

Dabei sind die $\Omega$ die möglichen Ausgänge bzw. Ergebnisse eines Experiments, $\mathcal{F} \subset \mathcal{P}(\Omega)$ bezeichnet die Ereignisse die wir unterscheiden und $\mathbb{P}$ ist ein Wahrscheinlichkeitsmaß über $\mathcal{F}$.

**Beispiel** einfaches Würfeln

Beim einfachen Würfeln haben wir $\Omega = \{1, 2, 3, 4, 5, 6\}$ wobei $\omega \in \Omega$ bedeutet, dass der Würfel die Augenzahl $\omega$ zeigt.

**Beispiel** n-maliges Würfeln

Würfeln wir nun n-mal, so ist $\Omega = \{1, \cdots, 6\}^n$ und $\omega \in \Omega$ zählt die Einzelergebnisse der n Würfe auf. Ein Ereignis wäre dann $A = \left\{ \omega \in \Omega \;\middle|\; \forall i, j \colon \omega_j = \omega_i \right\}$: „Alle Würfe ergeben die selbe Zahl“. Da wir die Gleichverteilung auf $\Omega$ erwarten, erhalten wir das *Laplace'sche Wahrscheinlichkeitsmaß*, dass jedem der endlich vielen Ausgänge $\omega \in \Omega$ die Wahrscheinlichkeit $\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}$ zuordnet und in diesem Fall ergibt:

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

**Beispiel** Münzwurf

Der einfache Münzwurf lässt sich wiefolgt modellieren:
$\Omega = \{0, 1\}$ wobei 0 Zahl bedeutet und 1 Kopf.

Dies lässt sich auch auf n-fachen Münzwurf erweitern: $\Omega = \{0, 1\}^n$.

Probleme entstehen aber, wenn man beliebig oft spielen will. Die Gleichverteilung auf $\mathcal{P}(\{0, 1\}^{\mathbb{N}})$ lässt sich nicht mehr intuitiv beschreiben.

**Beispiel** Aktienkurse

Wir betrachten die stetigen Funktionen $\Omega = C([0, T], \mathbb{R})$ und das Ereignis

$$A = \{ \omega \in \Omega \mid \forall 0 \leq t \leq T \colon \omega(t) \geq 5000 \}$$

**Ziel** Wir wollen jedem Ereignis $A$ eine Wahrscheinlichkeit $\mathbb{P}(A) \in [0, 1]$ zuordnen, sodass $\mathbb{P}$ als Abbildung von $\mathcal{F} \subseteq \mathcal{P}(\Omega) \to [0, 1]$ gewisse vorteilhafte Eigenschaften hat.

1 **Satz: Vitali**

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

1. $\mathbb{P}(\Omega) = 1$
2. $A_k \subseteq \Omega$, $k \in \mathbb{N}$, $i \neq j \Rightarrow A_i \cap A_j = \emptyset$, so gilt
$$\mathbb{P}(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)$$
3. Mit $T_k: \Omega \to \Omega$, $T_k(\omega)_i = \begin{cases} 1 - \omega_i & \text{für} \quad i = k \\ \omega_i & \text{sonst} \end{cases}$, $i \in \mathbb{N}$
und $A \in \mathcal{P}(\Omega)$ gilt $\forall k \in \mathbb{N}: \mathbb{P}(A) = \mathbb{P}(T_k(A))$

**Beweis** (1) Auf $\Omega$ definieren wir eine Äquivalenzrelation:

$$\omega \sim \omega' :\Leftrightarrow \exists n_0 \in \mathbb{N}: \forall k \geq n_0: \omega_k = \omega'_k$$

Glaubt man an das Auswahlaxiom, kann man ein Repräsentantensystem $A \subset \Omega$ auswählen.
Für $S = \left\{n_1, n_2, \cdots, n_k\right\} \subset \mathbb{N}$ endlich definieren wir $T_S := T_{n_1} \circ T_{n_2} \circ \cdots \circ T_{n_k}$ dann gilt

$$\Omega = \bigsqcup_{S=\left\{n_1,\cdots,n_k\right\}} T_S(A)$$

da $\omega \sim \omega' \Leftrightarrow \exists S: \omega = T_S(\omega')$ und $A$ Repräsentantensystem von $(\Omega, \sim)$ ist.
Jetzt gibt es folgende Fälle:

1. $\mathbb{P}(A) = 0 \Rightarrow \mathbb{P}(T_S(A)) = 0 \Rightarrow \mathbb{P}(\Omega) = 0$ ↯
2. $\mathbb{P}(A) > 0 \Rightarrow \mathbb{P}(T_S(A)) > 0 \Rightarrow \mathbb{P}(\Omega) = \infty$ ↯

□

# 1 Wahrscheinlichkeitsräume

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

2 **Bemerkung: Folgenräume und reelle Zahlen**

$\{0,1\}^{|\mathbb{N}|}$ ist mit $[0,1]$ gleichmächtig mittels der dualen Zahlendarstellung als Bijektion:

$$(\omega)_{i \in \mathbb{N}} \leftrightarrow \sum_{i=1}^{\infty} 2^{-i} \omega_i$$

wenn wir die abzählbarvielen Folgen der Form $\left(a_1, a_2, \cdots, 0, 1, 1, \cdots\right)$ und $\left(a_1, a_2, \cdots, 1, 0, 0, \cdots\right)$ paarweise identifizieren.

3 **Satz: Nichtexistenz der Gleichverteilung**

Es sei $\Omega := [0,1) = \mathbb{R} \mod 1$. Dann existiert kein $\mathbb{P}: \mathcal{P}(\Omega) \to [0,1]$ so dass

1. $\mathbb{P}(\Omega) = 1$
2. $\mathbb{P}(\bigsqcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)$
3. $\forall x \in \mathbb{R}, A \subset \Omega: \mathbb{P}(A) = \mathbb{P}(A + x)$

**Beweis** (3) Wir betrachten die Äquivalenzrelation

$$x \sim y :\Leftrightarrow x - y \in \mathbb{Q} \subset \mathbb{R}$$

Dazu sei $A \subset [0,1)$ eine Menge von Repräsentanten. (Dabei ist $|A| > |\mathbb{N}|$). Dann ist

$$\bigsqcup_{r \in [0,1) \cap \mathbb{Q}} (A + r) = [0,1)$$

und somit mit den geforderten Eigenschaften:

$$1 = \mathbb{P}(\Omega) = \sum_{x \in [0,1) \cap \mathbb{Q}} \mathbb{P}(A) \quad \lightning$$

$\square$

**Folgerung** Da es nicht möglich ist, die obigen „vernünftig" erscheinenden Forderungen auf der gesamten Potenzmenge zu erfüllen, müssen wir uns auf Teilmengen beschränken.

4 **Definition: „σ-Algebra, Ereignis"**

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

1. $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$
3. $A_k \in \mathcal{F}, k \in \mathbb{N} \Rightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$

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

5 **Definition: „Wahrscheinlichkeitsmaß (Колмогоров Axiome)"**

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

1. $\mathbb{P}(\Omega) = 1$
2. $A_k \in \mathcal{F}, k \in \mathbb{N} \Rightarrow \mathbb{P}(\bigsqcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)$

Wir bezeichnen $(\Omega, \mathcal{F}, \mathbb{P})$ dann als *Wahrscheinlichkeitsraum*.

6 **Bemerkung: Eigenschaften von σ-Algebren**

Ist $\mathcal{F}$ eine σ-Algebra auf $\Omega$, dann gilt

1. $\varnothing \in \mathcal{F}$
2. $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$
3. $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$
4. $A_k \in \mathcal{F}, k \in \mathbb{N} \Rightarrow \bigcap_{k=1}^{\infty} A_k \in \mathcal{F}$

7 **Lemma: erzeugte σ-Algebren**

Sei $\Omega \neq \varnothing$ und $G \subseteq \mathcal{P}(\Omega)$. Dann existiert eine kleinste σ-Algebra $\mathcal{A}_\sigma(G)$ auf $\Omega$ sodass $G \subseteq \mathcal{A}_\sigma(G)$.

8 **Definition: „Borel-σ-Algebra“**

$\mathcal{B}(\Omega) := \mathcal{A}_\sigma(\{\ M \subseteq \Omega, \text{offen}\ \})$ nenen wir die Borel-σ-Algebra und speziell $\mathcal{B}^n := \mathcal{B}(\mathbb{R}^n)$.

**Bemerkung** Die Borel-σ-Algebra kann von verschiedenen Erzeugersystem erzeugt werden. Zum Beispiel von der Menge aller offenen Mengen $\mathcal{G}$, der Menge aller abgeschlossenen Mengen $\mathcal{A}$ oder der Menge aller Hyperquader mit rationalen Koordinaten $\mathcal{K}$.

9 **Lemma: Erzeugung der Borel-σ-Algebra**

$$\mathcal{A}_\sigma(\mathcal{G}) = \mathcal{A}_\sigma(\mathcal{K}) = \mathcal{A}_\sigma(\mathcal{A})$$

**Beweis** (9) $\mathcal{A}_\sigma(\mathcal{G}) = \mathcal{A}_\sigma(\mathcal{A})$ folgt sofort aus der Abgeschlossenheit von σ-Algebren unter Komplement.
$\mathcal{A}_\sigma(\mathcal{G}) = \mathcal{A}_\sigma(\mathcal{K})$ gilt, da jede offene Menge eine abzählbare Vereinigung von solchen Hyperquadern ist und jeder Hyperquader sich als abzählbarer Schnitt offener Mengen darstellen lässt.

10 **Definition: „Spur-σ-Algebra“**

Sei $(\Omega, \mathcal{F})$ ein messbarer Raum, $\Omega' \subseteq \Omega$, dann ist $\mathcal{F}' := \mathcal{F} \cap \Omega' \subseteq \mathcal{F}$ eine σ-Algebra und heißt Spur-σ-Algebra.

11 **Definition: „Produkt-σ-Algebra“**

Sei $\Omega = \bigtimes_{i \in I} E_i$ für eine beliebige Indexmenge $I \neq \varnothing$ und $\pi_i$ die kanonischen Projektionen, sowie $(E_i, \mathcal{F}_i)$ messbare Räume. Dann ist die Produkt-σ-Algebra von $((E_i, \mathcal{F}_i))_{i \in I}$ definiert als

$$\bigotimes_{i \in I} \mathcal{F}_i := \mathcal{A}_\sigma\left(\left\{\ \pi_i^{-1}(A_i)\ \middle|\ A_i \in \mathcal{F}_i\ \right\}\right)$$

Bei gleichen $(E_i, \mathcal{F}_i)$ mit $\mathcal{F} = \mathcal{F}_i$ schreibt man auch $\mathcal{F}^I$ bzw. $\mathcal{F}^n$ für endliche Fälle.

**Bemerkung** Es gilt $\mathcal{B}^n = \left(\mathcal{B}^1\right)^n$.

12 **Satz: elementare Eigenschaften von Wahrscheinlichkeitsmaßen $\mathbb{P}$**

1. $\mathbb{P}(\emptyset) = 0$
2. $\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$
3. $A, B \in \mathcal{F}$, $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$.
4. $A_i \in \mathcal{F}$, $i \in \mathbb{N}$: $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$
5. $A_i \in \mathcal{F}$, $i \in \mathbb{N}$, $A_1 \supset A_2 \supset \cdots \Rightarrow \mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mathbb{P}(A_i)$.
6. $A_i \in \mathcal{F}$, $i \in \mathbb{N}$, $A_1 \subset A_2 \subset \cdots \Rightarrow \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mathbb{P}(A_i)$.
7. $A_i \in \mathcal{F}$, $i \in \mathbb{N}$, $\forall\omega$: $\lim_{i\to\infty} \mathbf{1}_{A_i}(\omega) = \mathbf{1}_A(\omega) \Rightarrow \lim_{i=1}^{\infty} \mathbb{P}(A_i) = \mathbb{P}(A)$

13 **Definition: „diskrete Wahrscheinlichkeitsräume“**

Ist in $(\Omega, \mathcal{F}, \mathbb{P})$ die Menge $\Omega$ abzählbar, dann nennen wir diesen Wahrscheinlichkeitsraum diskret.

14 **Satz: Zähldichte**

Sei $\Omega$ abzählbar und $\rho\colon \omega \to \Omega$ eine Folge in $\mathbb{R}_+$ mit $\sum_{\omega\in\Omega} \rho(\omega) = 1$. Dann existiert genau ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\Omega, \mathcal{P}(\Omega))$ mit $\mathbb{P}(A) = \sum_{\omega\in A} \rho(\omega)$ für alle $A \in \mathbb{P}(\Omega)$. Wir nennen $\rho$ die *Zähldichte* und die einzelnen $\rho(\omega)$ *Wahrscheinlichkeitsgewichte*.

15 **Definition: „diskretes Produktmaß“**

Für $n$ diskrete Wahrscheinlichkeitsräume mit Zählmaßen $\rho_i$ erhalten wir ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\left(\bigtimes_{i=1}^{n} \Omega_i, \bigotimes_{i=1}^{n} \mathcal{F}_i, \prod_{i=1}^{n} \rho_i\right)$. Dieses nennen wir *diskretes Produktmaß*. Für das Produkt $n$ gleicher Zählmaße $\rho$, schreiben wir $\mathbb{P} = \rho^{\otimes n}$.

**einige Wahrscheinlichkeitsmaße**

**Binomialverteilung** $\Omega = \{0, 1, \cdots, n\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $p(k) = \binom{n}{k} p^k (1-p)^{n-k} =: \mathrm{Bin}_{n,p}(k)$ mit $p \in [0, 1]$.

**Geometrische Verteilung** $\Omega = \mathbb{N}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $p \in (0, 1]$. $p(k) = p(1-p)^k$.

**Poisson Verteilung** $\Omega = \mathbb{N}$, $\lambda > 0$ „Intensität“, $p(k) = e^{-\lambda} \frac{\lambda^k}{k!} =: \mathrm{Poisson}_\lambda(k)$.

16 **Lemma: Zusammenhang zwischen Poisson- und Binomialverteilung**

Sei $p_n \in (0, 1]$ mit $\lim_{n\to\infty} np_n = \lambda > 0$, dann gilt für alle $k \in \mathbb{N}$:

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

17 **Definition: „Zufallsvariable“**

Seien $(\Omega, \mathcal{F})$ und $(\Omega', \mathcal{F}')$ messbare Räume. Eine Abbildung $X\colon \Omega \to \Omega'$ heißt *Zufallsvariable* falls

$$\forall A' \in \mathcal{F}' \colon X^{-1}(A') \in \mathcal{F}$$

Diese Forderung erwächst aus dem Bedürfnis

$$\mathbb{P}(\{ X \in A' \}) = \mathbb{P}(\{ \omega \in \Omega \mid X(\omega) \in A' \}) = \mathbb{P}(X^{-1}(A')) = \mathbb{P} \circ X^{-1}(A')$$

berechnen zu wollen. Hinreichend ist zum Beispiel $\mathcal{F} = \mathcal{P}(\Omega)$.

18 **Definition: „Wahrscheinlichkeitsverteilung“**

Sei $X\colon (\Omega, \mathcal{F}, \mathbb{P}) \to (\Omega', \mathcal{F}')$ eine Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Dann heißt $\mathbb{P}_X := \mathbb{P} \circ X^{-1}$ das *Bildmaß* von $X$ bzw. die *Verteilung* von $X$ und ist ein Wahrscheinlichkeitsmaß auf dem messbaren Raum $(\Omega', \mathcal{F}')$.

19 **Definition: „Algebra, (Prä)maß, σ-endlich“**

Sei $\Omega \neq \emptyset$.

- $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ heißt *Algebra* falls
  1. $\Omega \in \mathcal{A}$
  2. $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$
  3. $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$
- Eine Abbildung $\mu\colon \mathcal{A} \to [0, \infty]$ heißt *Prämaß* auf einer Algebra $\mathcal{A}$ falls $A_n \in \mathcal{A}$ mit $A = \bigsqcup_{n=1}^{\infty} A_n \in \mathcal{A} \Rightarrow \mu(A) = \sum_{n=1}^{\infty} \mu(A_n)$.
- Eine solches $\mu$ heißt *Maß* falls $\mathcal{A}$ eine σ-Algebra ist.
- Ein Maß $\mu$ heißt *σ-endlich* falls $\exists (A_n)_{n\in\mathbb{N}}, A_n \in \mathcal{A} \colon A_n \uparrow \Omega \wedge \mu(A_n) < \infty$.
- Ein Maß $\mu$ heißt *Wahrscheinlichkeitsmaß* falls $\mu(\Omega) = 1$.

**Beispiel** Auf $\Omega = \mathbb{R}$ ist $\mathcal{E} = \left\{ \bigsqcup_{k=1}^{\infty} \left(a_k, b_k\right] \cap \mathbb{R} \;\middle|\; n \in \mathbb{N}, -\infty \leq a_1 < b_1 < \cdots < a_n < b_n \leq \infty \right\}$ eine Algebra. Dann definieren wir das folgende Prämaß:

$$\lambda\left(\bigsqcup_{k=1}^{n} \left(a_k, b_k\right]\right) := \sum_{k=1}^{n} (b_k - a_k)$$

Dafür gilt dann sofort endliche Additivität.

20 **Satz: Monotone Klassentheorem für Mengen**

Sei $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ ein Mengensystem, welches abgeschlossen bezüglich endlichen Durchschnitten ist und $\Omega$ enthält. Sei weiterhin $\mathcal{A}$ das kleinste $\mathcal{C}$ umschließende Mengensystem, welches abgeschlossen bezüglich wachsenden Grenzwerten und Mengendifferenzen (genauer: $A, B \in \mathcal{A}, B \subset A \Rightarrow A \setminus B \in \mathcal{A}$) ist, dann ist bereits $\mathcal{A} = \sigma(\mathcal{C})$.

**Beweis** (20) Für $B \subseteq \Omega$ sei $\mathcal{A}_B := \{ A \in \mathcal{A} \mid A \cap B \in \mathcal{A} \}$. Dann ist $\mathcal{A}_B$ abgeschlossen bezüglich wachsenden Grenzwerten und Mengendifferenzen. Für $B \in \mathcal{C}$ gilt für alle $C \in \mathcal{C}$:

$$B \cap C \in \mathcal{C} \subseteq \mathcal{A} \Rightarrow C \in \mathcal{A}_B \qquad \Rightarrow \mathcal{C} \subseteq \mathcal{A}_B \subseteq \mathcal{A} \Rightarrow \mathcal{A}_B = \mathcal{A}$$

Für $B \in \mathcal{A}$ und $C \in \mathcal{C}$ gilt $B \in \mathcal{A}_C = \mathcal{A}$ und $B \cap C \in \mathcal{A} \Rightarrow C \in \mathcal{A}_B$. Damit gilt . $\mathcal{A}$ ist also abgeschlossen bezüglich endlicher Durchschnitte.

Auch gilt $\Omega \in \mathcal{C} \subseteq \mathcal{A}$ und damit ist $\mathcal{A}$ auch abgeschlossen gegenüber Komplementbildung. Nach der Abgeschlossenheit unter aufsteigenden Grenzwerten ist dann auch jede abzählbare Vereinigung aus Mengen in $\mathcal{A}$ selber in $\mathcal{A}$. Mit $\mathcal{A} \subseteq \mathcal{C}$ folgt dann die Behauptung.

21 **Korollar:**

Seien $\mathbb{P}, Q$ Wahrscheinlichkeitsmaße auf dem messbaren Raum $(\Omega, \mathcal{F})$, die auf einem unter Durchschnitt stabilen System $\mathcal{C}$ von $\mathcal{F} = \mathcal{A}_\sigma(\mathcal{C})$ übereinstimmen, dann gilt $\mathbb{P} = Q$ auf $\mathcal{F}$.

22 **Satz: Fortsetzungssatz von Carathéodory (1917)**

Für jedes σ-endliche Prämaß $\mu$ auf einer Algebra $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ existiert ein eindeutiges Maß $\tilde{\mu}$ auf der σ-Algebra $\mathcal{F} = \mathcal{A}_\sigma(\mathcal{A})$, welche mit diesem auf $\mathcal{A}$ übereinstimmt. Zudem ist dann $\tilde{\mu}$ selber σ-endlich.

23 **Lemma: Eindeutigkeitssatz**

Seien $\mu$ und $\nu$ σ-endliche Maße auf dem Maßraum $(\Omega, \mathcal{F})$ wobei $\mathcal{F} = \mathcal{A}_\sigma(\mathcal{E})$ für einen unter Durchschnittbildung stabilen Erzeuger $\mathcal{E}$ mit $\mu = \nu$ auf $\mathcal{E}$ und $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}$ mit $A_n \uparrow \Omega$ sowie $\mu(A_n) = \nu(A_n)$. Dann gilt auf ganz $\mathcal{F}$: $\mu = \nu$.

24 **Lemma:**

Sei $\mu$ ein Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und endlich auf Kompakta. Dann ist

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

monoton wachsend und rechtsstetig.

25 **Definition: „Verteilungsfunktion"**

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

$$F(x) := \mathbb{P}((-\infty, x])$$

26 **Korollar:**

Jede Verteilungsfunktion $F$ eines Wahrscheinlichkeitsmaßes ist rechtsstetig, monoton wachsend und es ist $\lim_{x\to\infty} F(x) = 1$ sowie $\lim_{x\to-\infty} F(x) = 0$.

27 **Satz:**

Sei $F\colon \mathbb{R} \to \mathbb{R}$ eine monoton wachsende, rechtsstetige Funktion. Dann existiert ein eindeutiges σ-endliches Maß $\mu$ auf $(\mathbb{R}, \mathcal{B})$ mit $\forall a, b \in \mathbb{R}, a < b\colon \mu((a,b]) = F(b) - F(a)$.

**Beispiel**

1. Für $F(x) = x$ erhalten wir damit das *Lebesguemaß* $\lambda$ auf $(\mathbb{R}, \mathcal{B})$.
2. Für $F(x) = \min(1, \max(x, 1))$ liefert der Satz uns die Gleichverteilung $U$ auf $([0,1], \mathcal{B}([0,1]))$.

**Beweis** (27)

**Eindeutigkeit** $\{ (a,b] \mid a, b \in \mathbb{R}, a < b \}$ ist ein unter Durchschnitt stabiler Erzeuger von $\mathcal{B}$, wodurch die Eindeutigkeit bereits gilt.

**Existenz** Wir wählen

$$\mathcal{E} := \left\{ \bigsqcup_{k=1}^{K} \left(a_k, b_k\right] \cap \mathbb{R} \;\middle|\; K \in \mathbb{N}, -\infty \leq a_1 < b_1 < a_2 < \dots < b_k \leq +\infty \right\}$$

Dieses $\mathcal{E}$ ist eine Algebra. Für $A := \bigsqcup_{k=1}^{K} \left(a_k, b_k\right] \in \mathcal{E}$ definieren wir

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

Damit ist $\mu$ additiv auf $\mathcal{E}$.

Seien nun $A_n := \bigsqcup_{k=1}^{K_n} \left(a_{k,n}, b_{k,n}\right]$ disjunkt. Mit $A^\infty := \bigsqcup_{n=1}^{\infty} A_n \in \mathcal{E}$ gilt dann $A^\infty = \bigcup_{k=1}^{K_\infty} \left(a_{k,\infty}, b_{k,\infty}\right]$ wobei $K_\infty < \infty$. Dann müssen wir zeigen

$$\sum_{k=1}^{K_\infty} F(a_{k,\infty}) - F(b_{k,\infty}) = \sum_{n=1}^{\infty} \sum_{k=1}^{K_n} \left(F(b_{k,n}) - F(a_{k,n})\right) = \sum_{k=1}^{K_\infty} \sum_{n=1}^{\infty} \sum_{\substack{j=1 \\ \left(a_{j,n}, b_{j,n}\right] \subseteq \left(a_{k,\infty}, b_{k,\infty}\right]}}^{K_n} F(b_{j,n}) - F(a_{j,n})$$

Nun zeigen wir die Gleichheit summandenweise. Für jedes $k$ gilt also wegen Monotonie und $\left(a_\infty, b_\infty\right] \supseteq \bigsqcup_{n=1}^{N} \left(a_n, b_n\right]$ die Richtung „$\geq$“.

Betrachten wir $O_n := \left(a_n, b_n + \delta_n\right) \supseteq \left(a_n, b_n\right]$ offen mit $\delta_n := \delta(\varepsilon)$ sodass $\mu(\left(a_n, b_n + \delta_n\right]) \leq \mu(\left(a_n, b_n\right]) + \varepsilon 2^{-n}$ für ein beliebiges $\varepsilon > 0$. $O_n$ ist eine offene Überdeckung von $\left[a_\infty + \delta_\infty, b_\infty\right] \subseteq \left(a_\infty, b_\infty\right]$ mit $\delta_\infty = \delta_\infty(\varepsilon)$ sodass $\mu(\left(a_\infty, a_\infty + \delta_\infty\right])$ wegen Rechtsstetigkeit. Dieses Intervall ist kompakt, also können wir eine endliche Teilüberdeckung finden:
$\exists N \in \mathbb{N}\colon \bigcup_{n=1}^{N} \left(a_n, b_n + \delta_n\right) \supseteq \left[a_\infty + \delta_\infty, b_\infty\right] \supseteq \left(a_\infty + \delta_\infty, b_\infty\right]$. Damit gilt weiter

$$\begin{aligned}
&\mu(\left(a_\infty, b_\infty\right]) = \mu(\left(a_\infty, a_\infty + \delta_\infty\right]) + \mu(\left(a_\infty + \delta_\infty, b_\infty\right]) \\
&\leq \varepsilon + \sum_{n=1}^{N} \underbrace{\mu(\left(a_n, b_n + \delta_n\right])}_{\mu(\left(a_n, b_n\right]) + \varepsilon 2^{-n}} \leq 2\varepsilon + \sum_{n=1}^{\infty} \mu(\left(a_n, b_n\right])
\end{aligned}$$

Also ist auch „$\leq$“ da $\varepsilon$ beliebig gewählt war.

Somit ist $\mu$ ein Prämaß auf $\mathcal{E}$ und nach Fortsetzungssatz ein Maß auf $\mathcal{B} = \mathcal{A}_\sigma(\mathcal{E})$.

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

1. Sei $F(x) := \mathbf{1}_{[a,\infty)}$ mit $a \in \mathbb{R}$. Dann erhalten wir das Dirac-Punktmaß $\delta_a$ auf $a \in \mathbb{R}$.
2. Führt man noch mehrere endliche Zwischenstufen ein auf $a_i$ mit den Werten $p_i$, dann erhält man $\mu = \sum_{i=1}^n (p_i - p_{i-1})\delta_{a_i}$.

28 **Definition: „absolute Stetigkeit, Dichte“**

Existiert für ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\mathbb{R}, \mathcal{B})$ eine messbare Funktion $\rho: (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B}), \forall t \in \mathbb{R}: \rho(t) \geq 0$ sodass sich die Verteilungsfunktion von $\mathbb{P}$ darstellen lässt als $F(x) = \int_{-\infty}^x \rho(t)\mathrm{d}t$ so heißt $\rho$ die *Dichte* von $\mathbb{P}$ und $F$ ist *absolutstetig*.

29 **Lemma:**

Ist $F$ absolutstetig und es gilt $\int_{-\infty}^\infty \rho(t)\mathrm{d}t = 1$ für die Dichte $\rho$, dann ist $F$ die Verteilungsfunktion eines Wahrscheinlichkeitsmaßes $\mathbb{P}$ auf $(\mathbb{R}, \mathcal{B})$.

**Beispiel** für Wahrscheinlichkeitsmaße mit Dichten

1. Die Normalverteilung hat die Dichte $\varphi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ mit den Parametern $\mu$ und $\sigma$.
2. Die Gammverteilung besitzt ebenfalls eine Dichte. Diese können wir wiefolgt herleiten:

   Betrachten wir ein Modell für die Anzahl von Versicherungsschäden über einem Zeitintervall $(0, t]$. Dies können wir durch $\text{Poisson}_{\alpha t}(k) = e^{-\alpha t}\frac{(\alpha t)^k}{k!}$ modellieren. (Es wurde $\lambda = \alpha t$ gewählt.) $\alpha$ beschreibt die Proportionalität zwischen der Länge des Zeitintervalls und der erwarteten Anzahl an Schäden. Für das Ereignis, dass mindestens $r$ Schäden auftreten erhalten wir:

   $$\mathbb{P}(A) = 1 - \text{Poisson}_\lambda(\{0, 1, \cdots, r\}) = 1 - \sum_{k=0}^{r-1} \text{Poisson}_\lambda(k)$$

   $$= 1 - \sum_{k=0}^{r-1} e^{-\alpha t}\frac{(\alpha t)^k}{k!} = \int_0^t \frac{\alpha^r}{(r-1)!} x^{r-1} e^{-\alpha x}\mathrm{d}x$$

   Dann ist die Verteilung der Zeit bis zum $r$-ten Schaden die Gammaverteilung mit der Dichte

   $$\gamma_{\alpha,r}(x) = \mathbf{1}_{[0,\infty)}(x) \cdot \frac{\alpha^r x^{r-1}}{(r-1)!} \cdot e^{-\alpha x}$$

   Die *Gammafunktion* ist

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

Allgemein kann dies nicht analytisch ausgeschrieben werden. Jedoch erhält man für $r \in \mathbb{N}$: $\Gamma(r) = (r-1)!$.

3. Die *Exponentialverteilung* mit Parameter $\alpha > 0$ hat die Dichte $\rho_\alpha = \gamma_{\alpha,1}$.

**Erinnerung** an die Hauptresultate zum Maßintegral

Ist $(\Omega, \mathcal{F})$ ein messbarer Raum mit einem Maß $\mu\colon \mathcal{F} \to [0,\infty]$. Dann wollen wir ein Integralbegriff definieren. Wir bezeichnen dieses dann mit $\int_\Omega f \mathrm{d}\mu$ bzw. $\int_\Omega f(\omega)\mathrm{d}\mu(\omega)$. Dabei werden erst die Integrale über elementare Funktionen (Treppenfunktionen) definiert: $f = \sum_{k=1}^N \alpha_k \mathbf{1}_{A_k}$ wobei $a_k \in \mathbb{R}_+ \Rightarrow \int f \mathrm{d}\mu := \sum_{k=1}^N \alpha_k \mu(A_k)$. Ist $f \geq 0$ messbar, so definieren wir $\int f \mathrm{d}\mu := \sup_{\substack{g \text{ elementar} \\ 0 \leq g \leq f}} \int g \mathrm{d}\mu$. Für alle anderen messbaren Funktionen $f$ mit $\int |f| \mathrm{d}\mu$ endlich setzen wir dann $\int f \mathrm{d}\mu := \int \max(0, f)\mathrm{d}\mu - \int \max 0, -f \mathrm{d}\mu$. Diese Funktionen heißen dann Lebesgue-integrierbar ($f \in L^1(\mu)$ bzw. $L^1(\Omega, \mathcal{F}, \mu)$).

**Bemerkung**

1. Für eine Folge messbarer Funktionen $(f_n)_{n\in\mathbb{N}}$ gelten die folgenden Konvergenzsätze
   a) „Lemma von Fatou“: für $f_n \geq 0$ gilt
   $$\int \liminf_{n\to\infty} f_n \mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n \mathrm{d}\mu$$
   b) „monotone Konvergenz“: für $0 \leq f_1 \leq f_2 \leq \cdots$ gilt
   $$\lim_{n\to\infty} \int f_n \mathrm{d}\mu = \int \lim_{n\to\infty} f_n \mathrm{d}\mu$$
   c) „majorisierte Konvergenz“: $\exists g \in L^1(\mu)$: $\mu(\left\{ \omega \in \Omega \;\middle|\; \exists n \in \mathbb{N}\colon |f_n(\omega)| > g(\omega) \right\})$ und $f_n \to f_\infty$ $\mu$-fast überall gilt, dann ist $\lim_{n\to\infty} \int f_n \mathrm{d}\mu = \int f_\infty \mathrm{d}\mu$.
2. für $f\colon \mathbb{R} \to [0,\infty)$ Riemann integrierbar auf $\mathbb{R}$, gilt dass $f$ Lebesgue integrierbar ist und dass die Integrale übereinstimmen, wenn für $\mu$ das Lebesguemaß gewählt wird.

30 **Korollar:**

Sei $(\Omega, \mathcal{F}, \mu)$ mit $\mu\colon \mathcal{F} \to [0,\infty]$ ein Maßraum und $\rho\colon \Omega \to \mathbb{R}$ eine $(\mathcal{F}, \mathcal{B})$-messbare Funktion. Außerdem sei $\rho \geq 0$ und $\int \rho \mathrm{d}\mu = 1$. Dann ist $\mathbb{P}(A) := \int \mathbf{1}_A \rho \mathrm{d}\mu = \int_A \rho \mathrm{d}\mu$ für $A \in \mathcal{F}$ ein Wahrscheinlichkeitsmaß. Man nennt $\rho = \frac{d\mathbb{P}}{d\mu}$ die *Radon-Nikodym Dichte* von $\mathbb{P}$ bezüglich $\mu$ und nennt $\mathbb{P}$ absolutstetig bezüglich $\mu$ mit Dichte $\rho$.

# 2 bedingte Wahrscheinlichkeit und Unabhängigkeit

**Motivation** Wir betrachten eine Studie eines neuen Tests auf eine Krankheit. Es werden 1000 Versuchspatienten getestet. Daraus entsteht die *Kontingenztafel*:

| | negativ | positiv | |
|---|---|---|---|
| gesund | 970 | 20 | 990 |
| krank | 1 | 9 | 10 |
| | 97 | 29 | 1000 |

Fragen: Welche Diagnose muss bei positivem Testergebnis gestellt werden?

Der Anteil der Gesunden inter den positiv getesteten ist $\frac{20}{29} \approx 69\%$. Der Anteil der Kranken unter den positiv Getesteten ist $\frac{9}{29} \approx 31\%$. Andererseits gilt unter der Bedingung, dass ein negatives Testergebnis vorliegt, dass der Anteil der Gesunden $\frac{970}{971} \approx 99,9\%$ und der Anteil der Kranken $\frac{1}{971} \approx 0,1\%$ ist.

31 **Definition: „bedingte Wahrscheinlichkeit"**

Gegeben sei $(\Omega, \mathcal{F}, \mathbb{P})$ und $A, B \in \mathcal{F}$ mit $\mathbb{P}(B) > 0$. Dann heißt

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

die bedingte Wahrscheinlichkeit von $A$ unter $B$.

32 **Satz: Eigenschaften der bedingten Wahrscheinlichkeit**

Sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ gegeben mit $B \in \mathcal{F}$ und $\mathbb{P}(B) > 0$.

1. $Q(A) := \mathbb{P}(A|B)$ ist ein Wahrscheinlichkeitsmaß auf $\mathcal{F}$.

2. Formel von der totalen Wahrscheinlichkeit: Sei $B = \bigsqcup_{i \in I} B_i$ mit $\mathbb{P}(B_i) > 0$ und abzählbarem $I$, dan gilt

$$\mathbb{P}(A \cap B) = \sum_{i \in I} \mathbb{P}(B_i)\mathbb{P}(A|B_i)$$

3. Seien $A, B_i \in \mathcal{F}$ mit strikt positiven Wahrscheinlichkeiten, $I$ abzählbar, $B_i$ paarweise disjunkt und $\Omega = \bigsqcup_{i \in I} B_i$, dann gilt

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(B_j)\mathbb{P}(A|B_j)}{\sum_{i \in I} \mathbb{P}(B_i)\mathbb{P}(A|B_i)}$$

33 **Lemma: Multiplikationsformel**

Seien $A_i \in \mathcal{F}$ mit $\mathbb{P}(\bigcap_{i=1}^{n-1} A_i) > 0$, dann ist

$$\mathbb{P}(\bigcap_{i=1}^{n} A_i) = \mathbb{P}(A_1) \prod_{i=2}^{n} \mathbb{P}(A_i| \bigcap_{j=1}^{i-1} A_j)$$

34 **Definition: „Unabhängigkeit“**

1. Zwei Ereigniss $A, B \in \mathcal{F}$ heißen unabhängig, falls $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

2. Eine beliebige Familie von Ereignissen $(A_i)$ heißt unabhängig falls für jede endliche Teilmenge $(A_j)_{j\in J}$ gilt

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

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

4. Eine Familie von Zufallsvariablen $Y_i\colon (\Omega, \mathcal{F}) \to (\Omega_i, \mathcal{F}_i)$ heißt unabhängig, falls die σ-Algebren $\sigma(Y_i) := Y_i^{-1}(\mathcal{F}_i)$ unabhängig sind. (Dabei ist $\sigma(Y_i)$ die kleinste σ-Algebra auf $\Omega$, bezüglicher welcher $Y_i$ messbar ist.)

**Bemerkungen**

1. Sind $(A_i)_{i\in I}$ unabhängig, dann sind die $A_i$ paarweise unabhängig. Die umgekehrte Implikation gilt jedoch nicht.

2. $(A_i)_{i\in I}$ sind genau dann unabhängig, wenn die $(\{\emptyset, A_i, \bar{A}_i, \Omega\})_{i\in I}$ unabhängige σ-Algebren sind.

35 **Satz: Unabhängigkeitskriterium für Zufallsvariablen**

Gegeben sei eine Familie von Zufallsvariablen $Y_i\colon (\Omega, \mathcal{F}) \to (\Omega_i, \mathcal{F}_i)$, $i \in I$ und ein unter Durchschnitt stabiler Erzeuger $\mathcal{E}_i$ ($\mathcal{F}_i = \sigma(\mathcal{E}_i)$). Für endliche $J \subset I$ und $B_i \in \mathcal{E}_i$, $i \in J$ gelte

$$\mathbb{P}(\bigcap_{i\in J} Y_i^{-1}(B_i)) = \prod_{i\in J} \mathbb{P}(Y_i \in B_i)$$

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

36 **Korollar: Speziellere Unabhängigkeitskriterien**

Sei $(Y_i)_{i\in\{1,\cdots,n\}}$ eine endliche Familie von Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$, dann gilt:

1. (diskreter Fall) falls die $Y_i$ die Form $Y_i\colon (\Omega, \mathcal{F}) \to (\Omega_i, \mathcal{F}_i)$ haben mit $\Omega_i$ abzählbar und $\mathcal{F}_i = \mathcal{P}(\Omega_i)$, so sind die $(Y_i)_{i=1,\cdots,n}$ genau dann unabhängig, wenn

$$\forall \omega_i \in \Omega_i\colon \mathbb{P}(Y_i = \omega_i, i = 1, \cdots, n) = \prod_{i=1}^{n} \mathbb{P}(Y_i = w_i)$$

2. (reellwertiger Fall) falls die $Y_i$ die Form $Y_i\colon (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ haben, so sind die $(Y_i)_{i=1,\cdots,n}$ unabhängig genau dann, wenn

$$\forall c_i \in \mathbb{R}\colon \mathbb{P}(Y_i \le c_i, i = 1, \cdots, n) = \prod_{i=1}^{n} \mathbb{P}(Y_i \le c_i)$$

3. (reellwertiger, absolutstetiger Fall) sind die $Y_i$ wie zuvor und die $\mathbb{P}_{Y_i} = \mathbb{P} \circ Y_i^{-1}$ *absolutstetige* Verteilungen auf $(\mathbb{R}, \mathcal{B})$ mit Dichtefunktionen $\rho_i \colon (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$, so sind die $(Y_i)_{i=1,\cdots,n})$ genau dann unabhängig, wenn $Y \colon = (Y_1, \cdots, Y_n) \colon (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}^n)$ eine absolutstetige Verteilung $\mathbb{P} \circ Y^{-1}$ mit der Dichte $\rho_Y(y) \colon = \rho_Y(y_1, \cdots, y_n) = \prod_{i=1}^n \rho_i(y_i)$ ist (bis auf eine Lebesgue-Nullmenge).

**Beispiel** $Y_1, Y_2$ sind unabhängig und jeweis $N(0,1)$-verteilt genau dann, wenn $Y = (Y_1, Y_2)$ eine absolutstetige Verteilung mit Dichte $\rho_Y(y) = \rho_1(y)\rho_2(y) = \frac{1}{2\pi} e^{-\frac{y_1^2+y_2^2}{2}}$. Diese Verteilung von $Y$ ist gerade die Standerdnormalverteilung mit Mittelwertparameter $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ und Kovarianzmatrixparameter $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = N(\vec{0}, I_2)$.

**Bemerkungen**

1. Die gemeinsame Verteilung einer Familie von Zufallsvariablen $Y_i \colon (\Omega, \mathcal{F}) \to (\Omega_i, \mathcal{F}_i)$ ist die Verteilung $\mathbb{P} \circ Y^{-1}$ der Zufallsvariable $Y \colon = (Y_i)_{i \in I} \colon (\Omega, \mathcal{F}) \to (\bigtimes_{i \in I} \Omega_i, \bigotimes_{i \in I} \mathcal{F}_i)$.

2. Umgekehrt induziert eine mehrdimensionale Zufallsvariable $D$ eine Verteilung $\mathbb{P} \circ Y^{-1}$ auf dem Produktraum $(\bigtimes_{i \in I} \Omega_i, \bigotimes_{i \in I} \mathcal{F}_i)$.

3. Die gemeinsame Verteilung induziert die Randverteilungen (Einschränkungen der $Y_i$) als $\mathbb{P}_{Y_i}(B_i) = \mathbb{P}(Y_i \in B_i) = \mathbb{P}(Y_1 \in \Omega_1, \cdots, Y_i \in B_i, \cdots Y_n \in \Omega_n) = \mathbb{P} \circ Y^{-1}(\Omega_1 \times \cdots \times B_i \times \cdots \times \Omega_n)$.

37 **Satz: Existenz des Produktmaßes**

Sei $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)_{i \in I}$ eine Familie von Wahrscheinlichkeitsräumen, dann existiert genau ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\Omega \colon = \bigtimes_{i \in I} \Omega_i$, $\mathcal{F} \colon = \bigotimes_{i \in I} \mathcal{F}_i = \sigma(\pi_i, i \in I)$ (Die kleinste σ-Algebra bezüglich welcher alle Koordinatenprojektionen messbar sind.), so dass für alle endlichen Teilmengen $Y \subseteq I$ gilt

$$\forall A_i \in \mathcal{F}_i \colon \mathbb{P}\left(\bigcap_{i \in J} \pi_i^{-1}(A_i)\right) = \prod_{i \in J} \mathbb{P}_i(A_i)$$

$\mathbb{P}$ heißt das *Produktmaß* und wir schreiben $\mathbb{P} = \bigoplus_{i \in I} \mathbb{P}_i$.

**Bemerkung** Insbesondere gilt für die Randverteilung der $i$-ten Koordinate unter $\mathbb{P}$, dass mit $A_i \in \mathcal{F}_i$: $\mathbb{P}(\pi_i^{-1}(A_i)) = \mathbb{P}_i(A_i)$ ist, das heißt die $\mathbb{P}_i$ sind die Randverteilungen der Koordinatenprojektionen $\pi_i$ welche Zufallsvariablen auf dem Raum $(\Omega, \mathcal{F}, \mathbb{P})$ sind.

38 **Korollar: Existenz von Projektionen**

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

39 **Bemerkung:**

1. Wegen dem obigen Korrolar sind Zufallsvariablen genau dann unabhängig, wenn ihre gemeinsame Verteilung durch das Produktmaß der Einzelverteilungen gegeben ist. Das heißt die Randverteilungen gerade die Verteilungen der einzelnen Zufallsvariablen sind.

2. Grundidee der bedingten Verteilungen:

   Fall $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$

   a) diskreter Fall: Zähldichte $\rho(y_1, y_2)$, bedingte Verteilung von $Y_1$ gegenüber $Y_2$ ist beschrieben durch die bedingte Zähldichte

   $$\rho(y_1|y_2) = \frac{\rho(y_1, y_2)}{\sum_{y_i \in \Omega} \rho(y_i, y_2)} = \frac{\rho(y_1, y_2)}{\rho_{Y_2}(y_2)}$$

   b) absolutstetiger Fall: sei $Y = (Y_1, Y_2)$ absolutstetig mit Dichte $\rho(y_1, y_2)$

   die bedingte Dichte ist

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

   dann gilt z.B.

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

   c) allgemeiner Fall (Ausblick auf Stochastik 2 / Maßtheorie)

   Sei $Y_2$ eine Zuffalsvariable die Werte in einem „polnischen Raum“ (separabler, vollständiger matrischer Raum mit Borelscher σ-Algebra) annimmt. (z.B. $(\mathbb{R}^n, \mathcal{B}^n)$), dann existiert ein stochastischer Kern (oder Markov-Übergangskern) $K$. (Das heißt $K \colon \Omega_1 \times \mathcal{F}_2 \to [0, 1]$ mit

   i. $y_1 \mapsto K(y_1, A_2)$ ist $\mathcal{F}_1$-messbar für alle $A_2 \in \mathcal{F}_2$

   ii. $A_2 \mapsto K(y_1, A_2)$ ist ein Wahrscheinlichkeitsmaß für alle $y_1 \in \Omega_1$

   ) Für diesen Kern gilt $\mathbb{P}_{Y_1, Y_2} = \mathbb{P}_{Y_1} \otimes K$, das heißt

   $$\forall A \in \mathcal{F}_1 \otimes \mathcal{F}_2 \colon \mathbb{P}((Y_1, Y_2) \in A) = \int_{\Omega_1 \times \Omega_2} \mathbf{1}_A \mathbb{P}_{Y_1, Y_2}(\mathrm{d}y_1, \mathrm{d}y_2) = \int_{\Omega_1} \left( \int_{\Omega_2} \mathbf{1}_{A_{y_1}}(y_2) K(y_1, \mathrm{d}y_2) \right) \mathbb{P}_{Y_1}(\mathrm{d}y_1)$$

   mit der Sektion

   $$A_{y_1} := \left\{ y_2 \in \Omega_2 \;\middle|\; (y_1, y_2) \in A \right\} \in \mathcal{F}_2$$

   falls $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$.

   Z.B. für $A = A_1 \times A_2$ gilt $A_{y_1} = \begin{cases} A_2, & y_1 \in A_1 \\ \emptyset, & \text{sonst} \end{cases}$ ist

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

**Bemerkung** Analoge Aussagen gelten für die Dimensionen $n \geq 2$.

40 **Satz: Unabhängigkeit nach Abbildung**

Sei $(Y_i)_{i \in I}$ eine Familie von unabhängigen Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$, $Y_i\colon (\Omega, \mathcal{F}) \to (\Omega_i, \mathcal{F}_i)$, $I = \bigsqcup_{k \in K} I_k$ und sind $\varphi_k(\bigtimes_{i \in I_k} \Omega_i, \bigotimes_{i \in I_k} \mathcal{F}_i) \to (\tilde{\Omega}_k, \tilde{\mathcal{F}}_k)$, $k \in K$ messbare Abbildungen, dann sind die Zufallsvariablen $\tilde{Y}_k \colon= \varphi_k((Y_i)_{i \in I_k})$, $k \in K$ unabhängig.

# 3 Asymptotische Ereignisse

Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(Y_k)_{k \in \mathbb{N}}$, $Y_k\colon (\Omega, \mathcal{F}) \to (\Omega_k, \mathcal{F}_k)$ eine Folge von Zufallsvariablen.

41 **Definition: „asymptotische Ereignisse“**

Ein Ereignis $A \in \mathcal{F}$ heißt *asymptotisch* bezüglich $(Y_k)_{k \in \mathbb{N}}$ falls für alle $n \in \mathbb{N}$ ein $B_n \in \bigotimes_{k \geq n} \mathcal{F}_k$ existiert mit $A = \left((Y_k)_{k \geq n}\right)^{-1}(B_n)$. Wir schreiben $\mathcal{A}((Y_k)_{\mathbb{N}})$ für das System der asymptotischen Ereignisse bezüglich $(Y_k)_{\mathbb{N}}$.

**Bemerkung** $\mathcal{A}((Y_k)_{\mathbb{N}})$ ist eine σ-Algebra.

**Beispiele**

1. $A = \bigcap_{k=1}^{\infty} \bigcup_{l \geq k} \left\{ \omega \mid Y_l(\omega) \in A_l \in \mathcal{F}_l \right\}$
2. $A = \left\{ \omega \mid \lim_{N \to \infty}(\frac{1}{N} \sum_{i=1}^{N} Y_i(\omega)) \text{ existiert und nimmt Werte im Intervall } [a, b] \text{ an} \right\}$

42 **Satz: 0-1 Gesetz von Колмогоров**

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

**Beweis** (42) Wir betrachten die Projektionen $\pi_i\colon \bigtimes_{k \in \mathbb{N}} \Omega_k \to \Omega_i, (\omega_k)_{\mathbb{N}} \mapsto \omega_i$ und

$$\mathcal{E} \colon= \left\{ \bigcap_{i=1}^{n} \pi_i^{-1}(A_i) \;\middle|\; A_i \in \mathcal{F}_i, i = 1, \cdots, n, n \in \mathbb{N} \right\}$$

dabei ist $\mathcal{E}$ stabil unter Durchschnittbildung und erzeugt $\bigotimes_{k \in \mathbb{N}} \mathcal{F}_k = \sigma(\pi_i | i \in \mathbb{N})$. Dann existiert für jedes $n \in \mathbb{N}$ ein $B_n \in \bigotimes_{k \geq n} \mathcal{F}_k$, sodass $A = ((Y_k)_{k \geq n})^{-1}(B_n) = \left\{(Y_k)_{k \geq n} \in B_n\right\}$. A ist unabhängig von $\left(Y_k\right)_{k \in \mathbb{N}}^{-1}(E)$, $E \in \mathcal{E}$, also auch für jedes $E \in \bigotimes_{k \in \mathbb{N}} \mathcal{F}_k$. Somit ist $A$ auch unabhängig von sich selbst, also $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A)$. □

43 **Satz: Borel-Cantelli-Lemma**

Es ist $(\Omega, \mathcal{F}, \mathbb{P})$ gegeben und $A_k, k \in \mathbb{N}$ eine Folge von Ereignissen sowie $A \colon= \limsup_k A_k = \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k$, dann gilt:

1. $\sum_{k=0}^{\infty} \mathbb{P}(A_k) < \infty \Rightarrow \mathbb{P}(A) = 0$
2. sind die $A_k, k \in \mathbb{N}$ unabhängige Ereignisse mit $\sum_{k=0}^{\infty} \mathbb{P}(A_k) = +\infty \Rightarrow \mathbb{P}(A) = 1$

**Beweis** (43)

1. $\forall n \colon A \subseteq \bigcup_{k\geq n} A_k \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(\bigcup_{k\geq n} A_k) \leq \sum_{k=n}^{\infty} \mathbb{P}(A_k) \xrightarrow{n\to\infty} 0$
2. $\bar{A} = \bigcup_{n=0}^{\infty} \bigcap_{k\geq n} \bar{A}_k$, dann gilt

$$\mathbb{P}(\bar{A}) = \sum_{n=0}^{\infty} \mathbb{P}(\bigcap_{k=n}^{\infty} \bar{A}_k) = \sum_{n=0}^{\infty} \lim_{m\to\infty} \mathbb{P}(\bigcap_{k=n}^{m} \bar{A}_k) = \sum_{k=0}^{\infty} \lim_{m\to\infty} \prod_{k=n}^{m} \underbrace{\underbrace{\mathbb{P}(\bar{A}_k)}_{=1-\mathbb{P}(A_k)}}_{\leq e^{-\mathbb{P}(A_k)}} \leq \sum_{n=0}^{\infty} \lim_{m\to\infty} e^{-\sum_{k=n}^{m} \mathbb{P}(A_k)} = 0$$

**Beispiel** „Monkey typing typewriter“: Seien $X_1, X_2, \ldots$ unabhängige Laplace-Zufallsvariablen auf $\{a, \cdots, z, A, \cdots, Z\} \cup \mathit{Punktation}$. Dann ist die Wahrscheinlichkeit, dass irgendwann ein beliebiges Wort, Ihr Name oder Goethes „Faust“ vorkommt gleich 1.

# 4 Erwartungswert und Varianz

44 **Definition: „Erwartungswert“**

Sei $X$ eine Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in $(\mathbb{R}, \mathcal{B})$. Ist $X \geq 0$ oder $X \in L^1(\mathbb{P})$ (das heißt $\int_\Omega |X|^1 \mathrm{d}\mathbb{P} < \infty$), so heißt $\mathbb{E}(X) := \int_\Omega X \mathrm{d}\mathbb{P}$ der *Erwartungswert* von $X$.

**Bemerkung**

- $L^p(\mathbb{P}) := L^p(\Omega, \mathcal{F}, \mathbb{P}) = \left\{ X \colon (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}) \;\middle|\; \int |X|^p \mathrm{d}\mathbb{P} < \infty \right\}$ für $p \in [1, \infty)$ ist ein Banachraum mit der Norm $\|X\|_p = (\int |X|^p \mathrm{d}\mathbb{P})^{\frac{1}{p}}$ und für $p = 2$ ein Hilbertraum mit $\langle X_1, X_2 \rangle = \int X_1 X_2 \mathrm{d}\mathbb{P}$.
- Man sagt oft auch „Mittelwert“ von $X$ bezüglich $\mathbb{P}$ zu $\mathbb{E}(X)$.
- Wichtige Eigenschaften des Erwartungswerts folgen aus Eigenschaften des Maßintegrals; etwa Linearität, Monotonie und die Konvergenzaussagen.

**Beispiel** Liegt ein diskreter Wahrscheinlichkeitsraum mit dem Zählmaß vor, so ist der Erwartungswert $\mathbb{E}(X) = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\})$.

45 **Lemma:**

Sei $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ die Verteilung von $X \colon (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ und $f$ eine messbare Funktion $f \colon (\Omega', \mathcal{F}') \to (\mathbb{R}, \mathcal{B})$ mit $f \geq 0$ oder $f \circ X \in L^1(\mathbb{P})$. Dann gilt $\mathbb{E}(f(X)) = \int_\Omega f(X(\omega)) \mathrm{d}\mathbb{P}(\omega) \stackrel{!}{=} \int_{\Omega'} f(x) \mathrm{d}\mathbb{P}_X(x)$

**Beweis** (45) für $f = \mathbf{1}_A$, $A \in \mathcal{F}'$ gilt $\int_\Omega \mathbf{1}_A(X)\mathrm{d}\mathbb{P} = \mathbb{P}(X \in A) = \mathbb{P}_X(A)$ ✓
Damit gilt die Behauptung für alle elementaren Funktionon $f = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k}$, $A_k \in \mathcal{F}'$, $a_k \in \mathbb{R}$. Weil jede messbare Funktion $f \geq 0$ monoton durch elementare Funktionen $f_n$ approximierbar ist mit $0 \leq f_n \leq f$, folgt die Behauptung für $f \geq 0$ mittels messbarer Konvergenz. Für $f \in L^1(\mathbb{P}_X)$ folgt die Behauptung dann über $f = f^+ - f^-$.

46 **Korollar:**

Sei $X$ eine Zufallsvariable mit absolutstetiger Verteilung und Dichte $\rho$ sowie $f$ eine messbare reelle Funktion für die $Y := f \circ X \geq 0$ oder $Y \in L^1(\mathbb{P})$ ist, so gilt $\mathbb{E}(f(x)) = \int_{\Omega'} f(x)\rho(x)\mathrm{d}x$ .

47 **Satz: wichtige Ungleichungen**

1. Markovsche Ungleichung
$$\forall p \in [1, \infty) : \forall \varepsilon > 0 : \mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}(|X|^p)}{\varepsilon^p}$$
2. den Spezialfall von 1 mit $p = 2$ nennt man auch Čebyšëv Ungleichung
3. exponentielle Markov Ungleichung
$$\mathbb{P}(\alpha X \geq \varepsilon) \leq \frac{\mathbb{E}(e^{\alpha X})}{e^\varepsilon}$$
4. Cauchy-Schwarz Ungleichung: für $X, Y \in L^2(\mathbb{P})$ gilt $XY \in L^1(\mathbb{P})$ und $\mathbb{E}(XY) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$
5. Höldersche Ungleichung: $X \in L^p(\mathbb{P})$, $Y \in L^q(\mathbb{P})$ mit $p \in (1, \infty)$, $q > 1$ sodass $\frac{1}{p} + \frac{1}{q} = 1$, dann gilt $\mathbb{E}(|XY|) \leq (\mathbb{E}(|X|)^p)^{\frac{1}{p}}(\mathbb{E}(|Y|^q))^{\frac{1}{q}}$.
6. Minkowski Ungleichung: $X, Y \in L^p(\mathbb{P})$, $p \in [1, \infty)$, dann ist $\|X+Y\|_{L^p} \leq \|X\|_{L^p} + \|Y\|_{L^p}$

**Beweis** (47)

1. $\mathbb{E}(|X|^p) \geq \mathbb{E}(\varepsilon^p \mathbf{1}_{\{|X| \geq \varepsilon\}}) = \varepsilon^p \mathbb{P}(|X| \geq \varepsilon)$

Die anderen Beweise werden hier nicht geführt.

48 **Lemma: Jensensche Ungleichung**

Ist $X$ eine reelle Zufallsvariable und $f : \mathbb{R} \to \mathbb{R}$ eine konvexe Funktion, $X, f(X) \in L^1(\mathbb{P})$, dann gilt
$$f(\mathbb{E}(X)) \leq \mathbb{E}(f(X))$$

**Beweis** (48) Da $f$ konvex ist, lässt sie sich darstellen als $f(x) = \sup_y(\alpha_y x + \beta_y)$, $x \in \mathbb{R}$ (supremum affiner Funktionen) mit passenden $\alpha_y, \beta_y \in \mathbb{R}$ darstellen, also gilt
$$\mathbb{E}(f(X)) \geq \sup_y(\alpha_y \mathbb{E}(X)^y + \beta y) = f(\mathbb{E}(X))$$

49 **Lemma: Erwartungswert unabhängiger Zufallsvariablen**
Sind $X, Y \in L^2(\mathbb{P})$ unabhängig, dann ist $\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$.

**Bemerkung** Für eine $\mathbb{R}^n$-wertige Zufallsvariable $X$ ist $\mathbb{E}(X) = (\mathbb{E}(X_i))_{i=1,\cdots,n}$ koordinatenweise definiert.

50 **Definition: „Varianz, Kovarianz“**
Für $X, Y \in L^2(\mathbb{P})$ gilt:

1. $\mathrm{Var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ ist die *Varianz* von $X$. $\sqrt{\mathrm{Var}(X)}$ ist die *Standardabweichung* (oder Streuung) von $X$.
2. $\mathrm{Cov}(X, Y) := \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ die *Kovarianz* von $X$ und $Y$.
3. Falls $\mathrm{Cov}(X, Y) = 0$ ist, heißen $X$ und $Y$ *unkorreliert*.

51 **Lemma:**
Seien $X, Y, X_1, X_2, \cdots \in L^2(\mathbb{P})$, $a, b, c, d \in \mathbb{R}$. Dann gilt

1. $\mathrm{Cov}(aX + b, cY + d) = a \cdot c \,\mathrm{Cov}(X, Y)$, insbesondere ist $\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X)$.
2. $\mathrm{Cov}(X, Y) \leq \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$
3. $\sum_{k=1}^n X_k \in L^2$ und $\mathrm{Var}(\sum_{k=1}^n X_k) = \sum_{k=1}^n (\mathrm{Var}(X_k) + \sum_{j \neq k} \mathrm{Cov}(X_j, X_k))$
4. Sind $X$ und $Y$ unabhängig, dann sind sie auch unkorreliert.

**Bemerkung**

1. Ist $X$ eine $\mathbb{R}$-wertige Zufallsvariable, $X \in L^2(\mathbb{P})$ mit $\mathrm{Var}(X) > 0$, dann heißt
$$\tilde{X} := \frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}}$$
*standardisiert*. ($\mathbb{E}(\tilde{X}) = 0$, $\mathrm{Var}(\tilde{X}) = 1$)
2. Für $X$ $\mathbb{R}^n$-wertig, $X \in L^2(\mathbb{P})$, das heißt $\forall i \colon X_i \in L^2$ heißt $(\mathrm{Cov}(X_i, X_j))_{i,j=1,\cdots,n}$ die *Varianz / Covarianzmatrix* von $X$.

**Beispiele**

1. für Varianzberechnung: Seien $X_1, \cdots, X_n$ iid (independent identically distributed) Bernoulli(p) Zufallsvariablen. Dann ist

$$X = \sum_{i=1}^{n} X_i \sim \mathrm{Bin}_{n,p}$$

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(X_i) = np$$

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \underbrace{0}_{\sum Cov} = n\,\mathrm{Var}(X_1) = np(1-p)$$

2. für $X \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}, \sigma^2 > 0$ ist $\mathbb{E}(X) = \mu$

**Bemerkung** Im allgemeinen impliziert Unkorreliertheit von $X, Y$ nicht deren Unabhängigkeit.

**Gegenbeispiele**

1. $U \sim \mathcal{U}((0, 2\pi))$ also gleichverteilt auf $(0, 2\pi)$. Dann wählen wir $X := \sin U$, $Y := \cos U$ dann ist $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ und $\mathrm{Cov}(X, Y) = 0$ aber sicher gilt keine Unabhängigkeit wegen $X^2 + Y^2 = 1$.

2. Sei $X \sim \mathcal{N}(0, 1)$ und $Y := X^2 - 1$. Dann ist $\mathbb{E}(Y) = 0$ also $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) = \mathbb{E}(X^3 - X) = 0$.

**Bemerkung** Falls $X$ $\mathbb{R}^n$-wertig, $X \in L^2(\mathbb{P})$, $\Sigma := \mathrm{Cov}(X, X)$, $Y := AX + b$ wobei $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ so gilt $\mathrm{Cov}(Y, Y) = A\Sigma A^T$, denn $\mathrm{Cov}(Y_i, Y_j) = Cov((AX)_i, (AY)_i)$ (sehe Übung)

52 **Definition: „Korrelationskoeffizient“**

Seien $X, Y \in L^2(\mathbb{P})$ mit $\mathrm{Var}(X), \mathrm{Var}(Y) > 0$, dann heißt

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

*Korrelation* von $X$ und $Y$. (Dies wird oft mit $\rho(X, Y)$ bezeichnet.)

53 **Lemma:**

Seien $X, Y$ wiederum wie eben. Dann gilt

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

2. Ist $\mathbb{E}(X) = 0$ so folgt

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

für $a^* = \mathrm{Cov}(X, Y)/\mathrm{Var}(X) = \mathrm{Corr}(X, Y)\sqrt{\mathrm{Var}(Y)/\mathrm{Var}(X)}$ und $b^* = \mathbb{E}(Y)$ sowie

$$\min_{a,b} \mathbb{E}(|Y - (aX + b)|^2) = \mathrm{Var}(Y)(1 - (\mathrm{Corr}(X, Y))^2)$$

## 4.1 Die Gesetze der großen Zahlen

**Vorbemerkung** klassische Formelierung der Čebyšëv-Ungleichung: $Y \in L^2(\mathbb{P}), \varepsilon > 0 \Rightarrow \mathbb{P}(|Y - \mathbb{E}(Y)| \geq \varepsilon) \leq \frac{\mathrm{Var}(Y)}{\varepsilon^2}$.

54 **Definition: „stochastische Konvergenz“**

Gegeben ist ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und $\mathbb{R}$-wertige Zufallsvariablen $Y, (Y_n)_{n\in\mathbb{N}}$. Dann sagen wir $(Y_n)_{n\in\mathbb{N}}$ *konvergiert stochastisch* (bzw. „in Wahrscheinlichkeit“ oder „in $\mathbb{P}$“), falls

$$\forall \varepsilon > 0\colon \lim_{n\to\infty} \mathbb{P}(|Y_n - Y| \leq \varepsilon) = 1$$

Man schreibt auch $Y_n \xrightarrow{\mathbb{P}} Y$.

55 **Satz: schwaches Gesetz der großen Zahl**

Seien $(X_n)_{n\in\mathbb{N}}$ paarweise unkorrelierte Zufallsvariablen aus $L^2(\mathbb{P})$ mit $\sup_n \mathrm{Var}(X_n) \leq c$ für ein $c \in \mathbb{R}$. Dann gilt für $\varepsilon > 0$:

$$\mathbb{P}(|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i))| \geq \varepsilon) \leq \frac{\sup_n \mathrm{Var}(X_n)}{n\varepsilon^2} \to 0$$

**Bemerkung**

1. Dies ist eine Form der stochastischen Konvergenz mit $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i)) \xrightarrow{\mathbb{P}} 0$.
2. Falls $\forall k\colon \mathbb{E}(X_k) = \mathbb{E}(X_1)$, so gilt $\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\mathbb{P}} \mathbb{E}(X_1)$

**Beweis** (55) $Y_n := \frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i)) \in L^2$ nach der Minkowski-Ungleichung und es gilt $\mathbb{E}(Y_n) = 0$. Dann gilt mit der Čebyšëv-Ungleichung:

$$\mathrm{Var}(Y_n) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_k) \leq \frac{1}{n^2}nc \to 0$$

$$\Rightarrow Y_n \xrightarrow{\mathbb{P}} 0$$

□

56 **Definition: „fast-sichere Konvergenz“**

Seien $Y, (Y_n)_{n\in\mathbb{N}}$ reellwertige Zufallsvariablen auf dem selben $(\Omega, \mathcal{F}, \mathbb{P})$. Dann konvergiert $Y_n$ $\mathbb{P}$-fast-sicher gegen $Y$ falls sie nur auf einer Nullmenge nicht konvergiert. Das heißt

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

Man sagt auch $Y_n$ konvergiert $\mathbb{P}$-fast-überall.

**Beispiele**

1. *Monte Carlo Integration*

   Wir betrachten eine messbare Funktion $f\colon [0,1] \to [0,c]$ mit $c \in \mathbb{R}^+$ (z.B. 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

   $$\mathbb{P}(|\frac{1}{n}\sum_{k=1}^{n} f(X_k) - \int_{[0,1]^d} f\mathrm{d}x\,| \geq \varepsilon) = \mathbb{P}(|\frac{1}{n}\sum_{k=1}^{n} f(X_k) - \mathbb{E}(f(X_1))| \geq \varepsilon) \leq \frac{\operatorname{Var}(X_1)}{n\varepsilon^2} \xrightarrow{n\to\infty} 0$$

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

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

   $$\mathbb{E}(f(\frac{1}{n}\sum_{k=1}^{n} X_k)) = \sum_{k=0}^{n} f(\frac{k}{n})\operatorname{Bin}_{n,p}(k) = \sum_{k=0}^{n} f(\frac{k}{n})\binom{n}{k}p^k(1-p)^{n-k} =: \underbrace{f_n(p)}_{\text{Bernstein Polynom } n\text{-ten Grades}}$$

   Dann behaupten wir

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

   $f$ ist stetig also gleichmäßig stetig auf dem Kompaktum $[0,1]$:
   $\forall \varepsilon > 0\colon \exists \delta > 0\colon \forall x, y\colon |x-y| < \delta \Rightarrow |f(x) - f(y) \leq \varepsilon|$. Dann folgt weiter für beliebiges $\varepsilon > 0$.

   $$|f_n(p) - f(p)| = |\mathbb{E}(f(\frac{1}{n}\sum_{k=1}^{n} X_k) - f(p))|$$

   $$\leq \mathbb{E}(|f(\frac{1}{n}\sum_{k=1}^{n} X_k) - f(p)|(\mathbf{1}_{\{|\frac{1}{n}\sum_{k=1}^{n} X_k - p| < \delta\}} + \mathbf{1}_{\{|\frac{1}{n}\sum_{k=1}^{n} X_k - p| \geq \delta\}}))$$

   $$\leq \varepsilon + 2\|f\|_\infty \frac{1}{n\delta^2} p(1-p) \searrow_{n\to\infty} \varepsilon$$

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

□

**Bemerkung** Der Begriff der $\mathbb{P}$-fast-überallen Konvergenz ist wohldefiniert, denn $\left\{\, \omega \in \Omega \;\middle|\; \lim_{n\to\infty} Y_n(\omega) = Y(\omega) \,\right\}$ ist messbar:

$$\left\{\, \omega \in \Omega \;\middle|\; \lim_{n\to\infty} Y_n = Y \,\right\} = \left\{\, \omega \in \Omega \;\middle|\; \forall n \in \mathbb{N}\colon \exists k \in \mathbb{N}\colon \forall l \geq k\colon |Y_k(\omega) - Y(\omega)| \leq \frac{1}{n} \,\right\}$$

$$= \bigcap_{n\in\mathbb{N}} \bigcup_{k\in\mathbb{N}} \bigcap_{l=k}^{\infty} \left\{ |Y_l - Y| \leq \frac{1}{n} \right\} \in \mathcal{F}$$

57 **Lemma:**

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

**Beweis** (57)

$$1 = \mathbb{P}(Y_n \to Y) = \mathbb{P}(\bigcap_{n=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{l=k}^{\infty} \left\{ |Y_l - Y| \leq \frac{1}{n} \right\} = \lim_{n\to\infty} \mathbb{P}(\bigcup_{k=1}^{\infty} \bigcap_{l=k}^{\infty} \left\{ |Y_l - Y| \leq \frac{1}{n} \right\})$$

$$\Rightarrow 1 = \mathbb{P}(\bigcup_{k=1}^{\infty} \bigcap_{l=k}^{\infty} \left\{ |Y_l - Y| \leq \frac{1}{n} \right\} = \lim_{k\to\infty} \mathbb{P}(\bigcap_{l=k}^{\infty} \left\{ |Y_l - Y| \leq \frac{1}{n} \right\})$$

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

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

**Bemerkung** Im Allgemeinen gilt die Umkehrung nicht — stochastische Konvergenz impliziert nicht die $\mathbb{P}$-fast-sichere Konvergenz.

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

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

Dann gilt $\mathbb{P}(|Y_k - 0| > \varepsilon) \leq \frac{1}{2^n}$ für $2^n \leq k \leq 2^{n+1}$. Also konvergiert $Y_k$ $\mathbb{P}$-stochastisch gegen 0 jedoch nicht $\mathbb{P}$-fast-überall. Es gilt sogar $\limsup_k Y_k(\omega) = 1$ und $\liminf_k Y_k(\omega) = 0$. Das heißt $Y_k$ konvergiert nirgends.

58 **Satz: starkes Gesetz der großen Zahl**

Seien unkorrelierte $(X_k)_{k\in\mathbb{N}}$ reellwertige Zufallsvariablen mit $\sup_n \operatorname{Var}(X_n) < \infty$, dann gilt

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

**Beweis** (58) O.B.d.A. gelte $\mathbb{E}(X)_n = 0$ (sonst betrachte $X_n' = X_n - \mathbb{E}(X_n)$). Zuerst werden wir die Konvergenzaussage für $Y_{n^2}$ zeigen. Für ein beliebiges $\varepsilon > 0$ gilt nach Čebyšëv:

$$\mathbb{P}(\underbrace{\{ |Y_{n^2}| > \varepsilon \}}_{=:A_n(\varepsilon)}) \leq \frac{c}{n^2\varepsilon^2}$$

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

Wir können nun also Satz 43 (Borel-Cantelli-Lemma) anwenden und erhalten dass die Wahrscheinlichkeit, dass nur endlich viele $A_n(\varepsilon)$ eintreten 1 ist:

$$\mathbb{P}(\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\bar{A_n}(\varepsilon)) = 1$$

$$\Rightarrow \exists N \in \mathcal{F}, \mathbb{P}(N) = 0 \colon \forall \omega \in \bar{N} \colon \exists n \in \mathbb{N} \colon \forall m \geq n \colon |Y_{m^2}(\omega)| \leq \varepsilon$$

$$\Rightarrow \limsup_{n\to\infty} |Y_{n^2}(\omega)| \leq \varepsilon \quad \text{fast überall}$$

$$\Rightarrow \mathbb{P}(\bigcap_{k=1}^{\infty}\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\bar{A_n}\left(\frac{1}{k}\right)) = 1$$

und für jedes $\omega$ aus der letzteren Menge gilt $\lim_{n\to\infty} Y_{n^2(\omega)} = 0$. Also gilt diese Aussage $\mathbb{P}$-fast-sicher.

Nun zeigen wir dass sogar $Y_n \to 0$ $\mathbb{P}$-fast-sicher gilt. Für jedes $m \in \mathbb{N}$ gibt es ein $n(m) \in \mathbb{N}$ sodass $n^2 \leq m < (n+1)^2$ dann

$$\mathbb{P}(|mY_m - n^2 Y_{n^2}| \geq n^2\varepsilon) \leq \frac{1}{\varepsilon^2 n^4} \operatorname{Var} \sum_{k=n^2+1}^{m} X_k \leq \frac{c(m-n^2)}{\varepsilon^2 n^4} \leq \infty$$

$$\Rightarrow \sum_{m=1}^{\infty} \mathbb{P}(|mY_m - n(m)^2 Y_{n(m)^2}| \geq n(m)^2\varepsilon) \leq \sum_{n=1}^{\infty} \frac{c(m-n(m)^2)}{\varepsilon^2 n(m)^4}$$

$$= \frac{c}{\varepsilon^2}\sum_{n=1}^{\infty}\sum_{m=n^2}^{(n+1)^2-1}\frac{m-n^2}{n^4} = \frac{c}{\varepsilon^2}\sum_{n=1}^{\infty}\sum_{k=1}^{2n}\frac{k}{n^4} = \frac{c}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{2n(2n+1)}{2}\cdot\frac{1}{n^4} < \infty$$

mit Satz 43 (Borel-Cantelli-Lemma) Folgt, analog zum vorigen Schritt dass

$$\mathbb{P}(\lim_{m\to\infty} |\frac{m}{n(m)^2} Y_m - Y_{n(m)^2}| = 0) = 1$$

Das heißt die Konvergenz gilt $\mathbb{P}$-fast-sicher. Zusammen mit der im ersten Teil des Beweises gezeigten Konvergenzaussage von $Y_{n^2}$ folgt dann

$$\lim_{m\to\infty}\frac{m}{n(m)^2} Y_m = 0 \quad \mathbb{P}\text{-fast-sicher} \qquad \wedge \qquad \lim_{m\to\infty}\frac{m}{n(m)^2} = 1$$

und somit gilt die Behauptung.

**Bemerkung** Das starke Gesetz der großen Zahl gilt auch unter schwächeren Voraussetzungen als oben angegeben. 1981 hat Etemadi die Konvergenz für $X_i \in L^1(P)$ unkorreliert und identisch verteilt gezeigt.

59 **S**

eien $X_1, X_2, \cdots$ iid Zufallsvariablen in $L^2(P)$, $\mathbb{R}$-wertig. Für jede Realisierung $\omega \in \Omega$ heißt

$$F_n(x)(\omega) := \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{(-\infty,x]}(X_k(\omega))$$

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

$$\mathbb{P}_n := \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k(\omega)}$$

auf $(\mathbb{R}, \mathcal{B})$. Dann sind $Y_k := \mathbf{1}_{(-\infty,x]}(X_k)$ iid Zufallsvariablen, $Bernoulli(F(x))$-verteilt, wobei $F$ die Verteilungsfunktion der einzelnen $X_k$ ist. Nach dem Gesetz der großen Zahlen gilt dann

$$\lim_{n\to\infty} F_n(x) = \frac{1}{n} \sum_{k=1}^{n} Y_k = F(x)$$

Das heißt die empirischen Verteilungsfunktionen konvergieren Punktweise gegen die Verteilungsfunktion $F$ aus der die iid Ziehungen stammen.

60 **Satz: Gliwenko-Cantelli**

Seien $(X_k)_{k\in\mathbb{N}}$ reellwertige Zufallsvariablen iid auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Verteilungsfunktion $F(x)$ und $F_n$ die empirische Verteilungsfunktion der $X_1, \cdots, X_n$. Dann gilt

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

**Beweis** (60)

$$Y_n(x) := \mathbf{1}_{(-\infty,x]}(X_n), \quad Z_n(x) := \mathbf{1}_{(-\infty,x)}(X_n)$$

Dann sind $(Y_n)_\mathbb{N}$ und $(Z_n)_\mathbb{N}$ jeweils iid Folgen von Zufallsvariablen und sind jeweils Bernoulli-Folgen mit Erfolgswahrscheinlichkeit $F(x)$ bzw. $F(\nearrow x) := \lim_{\tilde{x}\nearrow x} F(\tilde{x})$ ($F_n(\nearrow x)$ analog)

$$\mathbb{E}Y_n(x) = F(x), \mathbb{E}Z_n(x) = F(\nearrow x)$$

nach Satz 58 (starkes Gesetz der großen Zahl) gilt dann

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} Y_k(x) \xrightarrow{n\to\infty} F(x) F_n(\nearrow x) = \frac{1}{n} \sum_{k=1}^{n} Z_k(x) \xrightarrow{n\to\infty} F(\nearrow x)$$

Sei $F(-\infty) := 0$, $F(+\infty) := 1$. Fixiere nun $N \in \mathbb{N}$ und definiere

$$x_j := \inf \left\{ x \in \bar{\mathbb{R}} \;\middle|\; F(x) \geq \frac{j}{N} \right\}$$

$$R_n := \max_{j=0,\cdots,N} \left\{ |F_n(x_j) - F(x_j)| + |F_n(\nearrow x_j| - F(\nearrow x_j) \right\}$$

$$\Rightarrow \lim_{n\to\infty} R_n = 0 \quad \mathbb{P}-\text{fast überall}$$

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

$$F_n(x) \leq F_n(\nearrow x_j) \leq F_n(\nearrow x_j) + R_n \leq F(x) + R_n + \frac{1}{N}$$
$$F_n(x) \geq F_n(x_{j-1}) \geq F_n(x_{j-1}) - R_n \geq F(x) - R_n - \frac{1}{N}$$
$$\Rightarrow \limsup_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \leq \frac{1}{N} + \limsup_{n\to\infty} R_n \leq \frac{1}{N}$$

Da das $N$ beliebig gewählt war wurde die Behauptung bewiesen.

# 5 Charakteristische Funktionen

**Grundidee** Wir wollen Wahrscheinlichkeitsmaße auf $(\mathbb{R}^n, \mathcal{B}^n)$ durch (komplexwertige) Funktionen eindeutig charakterisieren und damit nützliche Aussagen über Maße mit funktionentheoretischen Mitteln erhalten.

**Notation** Im folgenden verwenden wir das euklidische Skalarprodukt $\langle x, y\rangle := \sum_{k=1}^d x_k y_k$.

61 **Definition: „charakteristische Funktion eines Maßes“**
Sei $\mu$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^n, \mathcal{B}^n)$. Dann heißt $\hat{\mu}$:

$$\hat{\mu}: \mathbb{R}^d \Rightarrow \mathbb{C}, \quad u \mapsto \int_{\mathbb{R}^d} e^{i\langle u,x\rangle} \mathrm{d}\mu(x)$$

die *charakteristische Transformation* (oder „Fourier Transformation“) von $\mu$.
Für eine Zufallsvariable $X$ mit Werten in $\mathbb{R}^d$ heißt

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

die *charakteristische Funktion* von $X$.

**Bemerkung**

$$\hat{\mu}(u) = \int_{\mathbb{R}^d} e^{i\langle u,x\rangle} \mathrm{d}\mu(x) = \int_{\mathbb{R}^d} \cos(\langle u, x\rangle) \mathrm{d}\mu(x) + i \int_{\mathbb{R}^d} \sin(\langle u, x\rangle) \mathrm{d}\mu(x)$$
$$\varphi_X(u) = \mathbb{E}(\cos(\langle u, X\rangle)) + i\mathbb{E}(\sin(\langle u, X\rangle))$$

62 **Lemma: Eigenschaften der charakteristischen Transformation**
Sei $\mu$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^d, \mathcal{B}^d)$. Dann ist $\hat{\mu}$ eine beschränkte Funktion auf $\mathbb{R}^d$ mit $\hat{\mu}(0) = 1$.

**Beweis** (62) $\hat{\mu}(0) = 1$ ist klar. $\hat{\mu}$ ist beschränkt da

$$|\hat{\mu}(u)| \leq \int_{\mathbb{R}^d} \underbrace{|e^{i\langle u,x\rangle} \mathrm{d}\mu(x)|}_{=1} = 1$$

Stetigkeit gilt wegen der Beschränktheit des Integranden nach beschränkter Konvergenz (Lebesgue).

63 **Definition: „Momente einer Zufallsvariablen“**
Ist $X$ eine Zufallsvariable mit werten in $\mathbb{R}^d$, dann heißt $\mathbb{E}(|X|^m)$, $m \in \mathbb{N}$ das $m$-te Moment von $X$.

64 **Satz: Beziehung zwischen Momenten und charakteristischer Funktion**
Sei $X$ eine $\mathbb{R}^d$-wertige Zufallsvariable mit einem endlichen $m$-ten Moment. Dann ist die charakteristische Funktion $\varphi_X$ von $X$ $m$ mal stetig partiell differenzierbar und

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

**Beweis** (64) Sei $\mu := P_X$ die Verteilung auf $\mathcal{B}^d$ und $\int_{\mathbb{R}^d} |x|^m \mathrm{d}\mu(x) < \infty$, das heißt $|x|^m \in L^1(\mu)$. Wir betrachten den Fall $m = 1$.
Wir betrachten also den Differenzenquotienten der Richtungsableitung. Für $\lim_{n\to\infty} t_n = 0$:

$$\frac{\hat{\mu}(u + t_n e_j) - \hat{\mu}(u)}{t_n} = \int_{\mathbb{R}^d} \frac{1}{t_n} e^{i\langle u,x\rangle}(e^{i\langle t_n e_j,x\rangle} - 1)\mathrm{d}\mu(x)$$

$$\lim_{n\to\infty} \frac{1}{t_n}(e^{i\langle t_n e_j,x\rangle} - 1) = \lim_{n\to\infty} \frac{\cos(t_n x_j) - 1}{t_n} + \frac{i\sin(t_n x_j)}{t_n} = -x_j \sin(0) + ix_j \cos(0) = ix_j$$

$$|\frac{1}{t_n}(e^{i\langle t_n e_j,x\rangle} - 1)| \leq 2|x| \in L^1(\mu)$$

$$\Rightarrow \lim_{n\to\infty} \frac{\hat{\mu}(u + t_n e_j) - \hat{\mu}(u)}{t_n} = \int_{\mathbb{R}^d} e^{i\langle u,x\rangle}(ix_j)\mathrm{d}\mu(x) = i\mathbb{E}(X_j e^{i\langle u,X\rangle}) = \frac{\partial}{\partial x_j}\varphi_X(u)$$

nach majorisierter Konvergenz. Die Stetigkeit folgt analog zu obigem Lemma. Weiter folgt die Behauptung für andere $m$ mit vollständiger Induktion über $m$.

**Beispiel**

1. $X \sim Bernoulli(p)$:

$$\varphi_X(u) = \mathbb{E}(e^{iuX}) = (1-p)e^0 + pe^{iu} = pe^{iu} + 1 - p$$

2. $X \sim \mathrm{Bin}(n, p)$:

$$\varphi_X(u) = \mathbb{E}(e^{iuX}) = \mathbb{E}(e^{iu\sum_{k=1}^n X_k}) = \prod_{k=1}^n \mathbb{E}(e^{iuX_k}) = (pe^{iu} + 1 - p)^n$$

3. $X \sim \text{Poisson}(\lambda)$:

$$\varphi_X(u) = \sum_{k=0}^{\infty} e^{iuk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda(e^{iu}-1)}$$

4. $X \sim \mathcal{U}([-a,a])$: $\varphi_X(u) = \frac{1}{au}\sin(au)$

5. $X \sim \mathcal{N}(0,1)$:

$$\varphi_X(u) = \mathbb{E}(e^{iuX}) = \int_{\mathbb{R}} \cos(ux)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,\mathrm{d}x + \underbrace{i\int_{\mathbb{R}} \sin(ux)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,\mathrm{d}x}_{=0}$$

$$\varphi'_X(u) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} -\sin(ux)xe^{-\frac{x^2}{2}}\,\mathrm{d}x = -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} u\cos(ux)e^{-\frac{x^2}{2}}\,\mathrm{d}x = -u\varphi_X(u)$$

$$\Rightarrow \frac{\varphi'_X(u)}{\varphi_X(u)} = -u \Rightarrow \ln(\varphi_X(u)) = -\frac{u^2}{2} + c \Rightarrow \varphi_X(u) = e^{-\frac{u^2}{2}+c} = e^{-\frac{u^2}{2}}$$

65 **Lemma: charakteristische Funktion von affinen Transformationen**
Sei $X$ eine $\mathbb{R}^d$-wertige Zufallsvariable und $Y := AX + b$ mit $A \in \mathbb{R}^{m\times d}$ und $b \in \mathbb{R}^m$. Dann gilt

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

**Beispiel** Sei $X \sim \mathcal{N}(\mu,\sigma^2), \mu \in \mathbb{R}, \sigma \neq 0$ und ist $X = \mu + \sigma Y$ für $Y \sim \mathcal{N}(0,1)$, dann gilt $\varphi_X(u) = \exp(iu\mu - \sigma^2\frac{u^2}{2})$.

**Beispiel** Seien $X_1, \cdots, X_d$ iid $X_k \sim \mathcal{N}(0,1)$. Dann heißt $X = (X_1, \cdots, X_d)$ standardnormalverteilt in $\mathbb{R}^d$ und $\varphi_X(u) = \exp(-\frac{|u|^2}{2})$.

## 5.1 Summe von unabhängigen Zufallsvariablen

Wir werden sehen, dass wir für Summen von unabhängigen Zufallsvariablen die charakteristische Funktion sehr einfach berechnen lassen und zudem den Begriff der *Faltung* einführen.

66 **S**
eien $X, Y\colon \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Verteilungen $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$, $\mathbb{P}_Y = \mathbb{P} \circ Y^{-1}$ auf $\mathcal{B}^1$. Dann heißt die Verteilung von $Z := X + Y$ die *Faltung* (bzw. *Faltungsprodukt*) von $\mathbb{P}_X$ und $\mathbb{P}_Y$ und man schreibt $\mathbb{P}_Z = \mathbb{P}_X * \mathbb{P}_Y$. Sie ist gegeben durch

$$\mathbb{P}_Z(A) = (\mathbb{P}_X * \mathbb{P}_Y)(A) := \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbf{1}_A(x+y)\mathrm{d}\mathbb{P}_x(x)\,\mathrm{d}\mathbb{P}_y(y)$$

**Beweis** (66) Da $X$ und $Y$ unabhängig sind ist $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$ damit gilt für $g\colon \mathbb{R}^2 \to \mathbb{R}$ integrierbar (d.h. $g \in L^1(\mathbb{P}_X \otimes \mathbb{P}_Y)$ oder $g \geq 0$)

$$\mathbb{E}(g(X,Y)) = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x,y) \mathrm{d}\mathbb{P}_X(x)\, \mathrm{d}\mathbb{P}_Y(y)$$

$$\xRightarrow{g(x,y)=\mathbf{1}_A(x+y)} \mathbb{E}(X+Y) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_A(x+y) \mathrm{d}\mathbb{P}_X\, \mathrm{d}\mathbb{P}_Y = \mathbb{P}(X+Y \in A) = \mathbb{P}(Z \in A) = P_Z(A)$$

67 **Korollar:**

Seien $X, Y$ unabhängige Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in $\mathbb{R}$, $Z := X + Y$. Dann ist

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

**Beweis** (67) Wie im obigen Satz erhalten wir

$$\mathbb{E}(e^{i\langle u, X+Y\rangle}) = \mathbb{E}(e^{i\langle u,X\rangle} \cdot e^{i\langle u,Y\rangle}) = \mathbb{E}(e^{i\langle u,X\rangle})\mathbb{E}(e^{i\langle u,Y\rangle})$$

68 **Bemerkung: Unabhängigkeit**

Es reicht aber nicht $\varphi_{X+Y}(u) = \varphi_X(u)\varphi_Y(u)$ um zu schließen, dass $X$ und $Y$ unabhängig sind.

69 **Satz: Existenz von Dichten**

Seien $X, Y$ unabhängige Zufallsvariablen und $Z := X + Y$.

1. Hat zudem $X$ eine Dichte $f_X$, so hat $Z$ eine Dichte auf $(\mathbb{R}, \mathcal{B})$ und

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

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

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

**Beweis** (69)

1.

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_A(x+y) \mathrm{d}\mathbb{P}_X(x)\, \mathrm{d}\mathbb{P}_Y(y) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_A(x+y) f_X(x) \mathrm{d}x\, \mathrm{d}\mathbb{P}_Y(y)$$

$$\overset{z=x+y}{=} \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_A(z) f_X(z-y) \mathrm{d}z\, \mathrm{d}\mathbb{P}_Y(y) = \int_{\mathbb{R}} \mathbf{1}_A(z) \int_{\mathbb{R}} f_X(z-y) \mathrm{d}\mathbb{P}_Y(y)\, \mathrm{d}z$$

$$\Rightarrow f_Z(z) = \int_{\mathbb{R}} f_X(z-y) \mathrm{d}\mathbb{P}_Y(y)$$

Die andere Form folgt mittels Vertauschung von $X$ und $Y$ nach Symmetrie.

2.

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

70 **Satz: Charakteristik-Eigenschaft**

Ist $X$ eine Zufallsvariable in $(\mathbb{R}^d, \mathcal{B}^d)$. Dann charakterisiert $\varphi_X$ bzw. $\hat{\mathbb{P}}_X$ die Verteilung $\mathbb{P}_X$ von $X$ auf $\mathcal{B}^d$ das heißt sind $\mu_1, \mu_2$ Wahrscheinlichkeitsmaße auf $\mathcal{B}^d$, dann gilt $\hat{\mu}_1 = \hat{\mu}_2$ nur wenn $\mu_1 = \mu_2$.

**Beweis** (70) (Siehe zum Beispiel Sim82, Seite 160). Dieser Beweis benutzt die lokal kompakte Version des Stone-Weierstraß-Theorems.
Wir betrachten die Funktion

$$f(\sigma, x) := \frac{1}{(\sqrt{2\pi\sigma^2})^d}\exp(-\frac{|x|^2}{2\sigma^2})$$
$$\hat{f}(\sigma, u) := \exp(-\frac{|u|^2}{2\sigma^2})$$

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

$$\varphi_X(u) = \mathbb{E}(e^{i\langle u, X\rangle}) = \prod_{j=1}^{d} \varphi_{X_j}(u_j) = \prod_{j=1}^{d} \exp(-\frac{u_j^2\sigma^2}{2}) = \hat{f}(\sigma, u)$$
$$\Rightarrow f(\sigma, u-v) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}}\hat{f}(\sigma, \frac{u-v}{\sigma^2}) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}}\mathbb{E}(e^{i\langle \frac{u-v}{\sigma^2}, X\rangle})$$

Seien $\mu_1, \mu_2$ Wahrscheinlichkeitsmaße auf $(\mathbb{R}^d, \mathcal{P}^d)$ mit $\hat{\mu}_1 = \hat{\mu}_2 =: \hat{\mu}$.

$$\int_{\mathbb{R}^d} f(\sigma, u-v)\mathrm{d}\mu_1 = \int_{\mathbb{R}^d} \frac{1}{(\sqrt{2\pi}\sigma)^d} \underbrace{(\int_{\mathbb{R}^d} f(\sigma, x)e^{i\langle u, x\rangle}\mathrm{d}x\,)}_{=\hat{f}(\sigma, \frac{u-v}{\sigma^2}} \mathrm{d}\mu_1(x) = \int_{\mathbb{R}^d} f(\sigma, u-v)\mathrm{d}\mu_1(u)$$
$$= \int_{\mathbb{R}^d} \frac{1}{(\sqrt{2\pi}\sigma)^d} f(\sigma, x)\hat{\mu}_1(x)\mathrm{d}x$$

Analog ergibt sich diese Formel mit $\mu_2, \hat{\mu}_2$.

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

gilt für alle Funktionn $g$ aus dem Vektorraum $\mathcal{H}$ der durch $\left\{\, f(\sigma, \cdot - v) \,\middle|\, \sigma > 0, v \in \mathbb{R}^d \,\right\}$ aufgespannt wird. $\mathcal{H}$ teilt einzelne Punkte in $\mathbb{R}^d$. Somit zeigt Stone-Weierstraß, dass $\mathcal{H}$ dicht bezüglich gleichmaßiger Konvergenz in der Banachalgebra $C_0(\mathbb{R}^d)$ liegt, wobei $C_0(\mathbb{R}^d)$ der Banachraum der stetigen Funktionen $\mathbb{R}^d \to \mathbb{R}$ mit Konvergenz bezüglich der Supremumsnorm ist, welche „gegen $\infty$ verschwinden". (d.h. $\forall \varepsilon > 0 \exists$ Kompaktum $K \subseteq \mathbb{R}^d, g \in C_0(\mathbb{R}^d)$: $|g| \leq \varepsilon$ auf $\mathbb{R}^d \setminus K$).

Jede Indikatorfunktion auf Rechtecken kann monoton approximiert werden durch Funktion aus $\mathcal{H}$ und dort die Maße übereinstimmen, müssen sie auch auf der gesammten Borell-σ-Algebra übereinstimmen.

**Bemerkung** Will man etwas Konstruktives zur Berechnung des Maßes aus seiner charakteristischen Funktion, so braucht man Ergebnise aus der Fourieranalysis zur *Fouriertransformierteninversion*. Es ergibt sich zum Beispiel im Eindimensionalen für eine Zufallsvariable $X: \Omega \to \mathbb{R}$:

$$\mathbb{P}_X((a,b)) + \frac{1}{2}\mathbb{P}_X(\{a,b\}) = \lim_{T \nearrow \infty} \frac{1}{2\pi} \int_{[-T,T]} \frac{e^{-iua} - e^{-iub}}{iu} \varphi_X(u) \mathrm{d}u$$

für beliebige $a, b \in \mathbb{R}, a < b$. Die Funktion

$$G(b) := \lim_{a \searrow -\infty} \mathbb{P}_X((a,b)) + \frac{1}{2}\mathbb{P}_X(\{a,b\}) = \mathbb{P}_X((-\infty,b)) + \frac{1}{2}\mathbb{P}_X(\{b\})$$

bestimmt die Verteilung von $X$ bereits eindeutig: $P(X \leq b) = F_X(b) = G(b) + \frac{1}{2}(G(b) - G(\nearrow b))$.

(Siehe auch Shi95, Paragraph 12, Theorem 3)

71 **Satz: Unabhängigkeit anhand charakteristischer Funktionen**

Zufallsvariablen $X = (X_1, \cdots, X_d)$ auf $(\Omega, \mathcal{F}, \mathbb{P})$ und Werten in $\mathbb{R}^d$ sind genau dann unabhängig, wenn

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

## 5.2 Normalverteilungen

72 **Definition: „multidimensionale Normalverteilung"**

Eine $\mathbb{R}^d$-wertige Zufallsvariable $X = (X_1, \cdots, X_d)$ auf $(\Omega, \mathcal{F}, \mathbb{P})$ heißt *Gaußsche Zufallsvariable* oder *multidimensional normalverteilt* falls für jedes $a \in \mathbb{R}^d$ die Linearkombination $\langle a, X \rangle = \sum_{k=1}^{d} a_k X_k$ eindimensional normalverteilt sind.

**Bemerkung** Unter Umständen sind die Linearkombinationen degeneriert normalverteilt mit Varianz 0, das heißt die Punktmaße auf $\mu$ in $\mathbb{R}^1$.

73 **Satz: charakteristische Funktion der Normalverteilung**

Ist $X$ eine $\mathbb{R}^d$-wertige Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$, dann ist $X$ genau dann multidimensional normalverteilt, wenn ihre charakteristische Funktion die Form

$$\forall u \in \mathbb{R}^d : \varphi_X(u) = \exp(i\langle u, \mu \rangle - \frac{1}{2}\langle u, Qu \rangle)$$

hat, mit $\mu \in \mathbb{R}^d$ und $Q$ eine symmetrische nichtnegative semi-definit $d \times d$-Matrix. Außerdem ist dann $Q$ die Koverianzmatrix und $\mu$ der Erwartungswertvektor.

**Beweis** (73) Hat die charakteristische Funktion von $X$ die gegebene Form so betrachten wir $Y := \langle a, X\rangle$ und erhalten für die $v \in \mathbb{R}^1$:

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

Wegen der Eindeutigkeit muss dann $Y \sim \mathcal{N}(\langle a, \mu\rangle, \langle a, Qa\rangle)$ sein. Die Momente erhalt wir aus Satz 64 (Beziehung zwischen Momenten und charakteristischer Funktion):

$$\mathbb{E}(X_i X_j) = \frac{1}{i^2} \cdot \frac{\partial}{\partial x_i \partial x_j}\varphi_X(u)|_{u=0} = (-1)(-\mu_i\mu_j - Q_{ij}) = \mu_i\mu_j + Q_{ij}$$
$$\Rightarrow \operatorname{Cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}X_i\mathbb{E}X_j = Q_{ij}$$

Ist auf der anderen Seite $X$ bereits als normalverteilt gegeben, so ergibt mit $Y := \langle a, X\rangle = \sum_{k=1}^d a_k X_k$ und $Q := \operatorname{Cov}(X)$:

$$\mathbb{E}(Y) = \mathbb{E}(\langle a, X\rangle) = \langle a, \mathbb{E}(X)\rangle = \langle a, \mu\rangle \wedge \mu = \mathbb{E}(X) \wedge \operatorname{Var}(Y) = a^T Qa = \langle a, Qa\rangle$$
$$\Rightarrow Y \sim \mathcal{N}(\langle a, \mu\rangle, \langle a, Qa\rangle) \Rightarrow \varphi_Y(v) = \exp(iv\langle a, \mu\rangle - \frac{1}{2}v^2\langle a, Qa\rangle)$$
$$\Rightarrow \varphi_Y(1) = \varphi_{\langle a,X\rangle}(1) = \mathbb{E}(\exp(i\langle a, X\rangle)) = \varphi_X(a)$$

Also hat $\varphi_X$ die behauptete Form.

**Beispiel** Seien $X_1, \cdots, X_d$ unabhängige Zufallsvariablen $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$. Dann ist $X = (X_1, \cdots, X_d)$ multivariat normalverteilt, denn

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

mit $\mu = (\mu_1, \cdots, \mu_d)$ und $Q = \operatorname{diag}(\sigma_1^2, \cdots, \sigma_d^2)$.

74 **Satz: Unabhängigkeit einer Normalverteilung**

Ist $X: \Omega \to \mathbb{R}^d$ (multidimensional normalverteilt), dann sind die Komponenten $X_j$ genau dann unabhängig, wenn sie unkorreliert sind, das heißt wenn $\operatorname{Cov}(X)$ eine Diagonalmatrix ist.

**Beweis** (74) Die Hinrichtung ist gerade das obige Beispiel. Die Rückrichtung gilt, da die charakteristischen Funktionen bereits gleich sind, wenn die Kovarianzmatrix Diagonalform hat.

75 **Bemerkung: Fortsetzbarkeit der charakteristischen Funktion**

Setzt man die charakteristische Funktion $\varphi_X(z) = \mathbb{E}(e^{\langle z,x\rangle})$ auf ganz $\mathbb{C}$ fortsetzt, so kann man zeigen, dass sie beispielsweise in $\mathbb{R}^1$ auf einem Streifen $\{ z \in \mathbb{C} \mid |\operatorname{Re} z| < c \}$ holomorph ist. (vergleiche Str85, I, Paragraph 5)

76 **Lemma: Simulation von multidimensionalen Normalverteilungen**

Sei $X \sim \mathcal{N}(\mu, Q)$. Dann existieren unabhängige univariate $Y_1, \cdots, Y_d$, $Y_j \sim \mathcal{N}(0, \lambda_j^2)$ sodass $X = y + AY$ für $\mu = \mathbb{E}(X)$ und eine orthogonale Matrix $A$ gilt.

**Beweis** (76) Sei $Q = \mathrm{Cov}(X)$ symmetrisch, nichtnegativ semidefinit. Dann ist $Q = A\Lambda A^T$ mit einer orthogonalen Matrix $A$ und einer Diagonalmatrix $\Lambda = \mathrm{diag}(\lambda_1^2, \cdots, \lambda_d^2)$ sowie $\lambda_j \geq 0$. Setzen wir dann $Y := A^T(X-\mu)$, so erfüllt $Y$ die Forderungen.

Hat $Q$ nicht vollen Rang, so reicht sogar $\Lambda = \mathrm{diag}(\lambda_1^2, \cdots, \lambda_{\mathrm{rk}\, Q}^2, 0, \cdots, 0)$.

**Bemerkung** Eine normalverteilte Zufallsvariable $X \sim \mathcal{N}(\mu, Q)$ hat genau dann eine Dichte wenn $\det Q \neq 0$. Dies gilt mit $X = y + \bar{A}Y$, $Y \sim \mathcal{N}(0, I_d)$, $\bar{A} = A\,\mathrm{diag}(\lambda_1, \cdots, \lambda_d)$. Dann hat $Y$ die Dichte

$$\frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{|y|^2}{2}} =: f_Y(y)$$

$$\Rightarrow f_X(x) = |\det \bar{A}^{-1}| f_Y(\bar{A}^{-1}(x-\mu)) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{\det Q}} \exp(-\frac{1}{2}\langle x-\mu, Q^{-1}(x-\mu)\rangle)$$

# 6 Konvergenz in Verteilung / schwache Konvergenz

77 **Definition: „schwache Konvergenz“**

Sei $(E, d)$ ein metrischer Raum mit der Borelschen σ-Algebra $\mathcal{B}(E)$.

1. Seien $\mu, (\mu_n)_{n\in\mathbb{N}}$ Wahrscheinlichkeitsmaße auf $(E, \mathcal{B})$. Dann *konvergiert* $\mu_n$ *schwach* gegen $\mu$ falls für alle stetigen, beschränkten Funktionen $f : E \to E$ gilt:

   $$\lim_{n\to\infty} \int_E f \mathrm{d}\mu_n = \int_E f \mathrm{d}\mu$$

   Wir schreiben $\mu_n \to \mu$ oder $\mu_n \xrightarrow{w} \mu$.

2. Seien $X, (X_n)_{n\in\mathbb{N}}$ Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ bzw. $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$, dann sagt man $X_n$ *konvergiert in Verteilung* gegen $X$ falls $\mathbb{P}_{X_n} = \mathbb{P}_n \circ X_n^{-1}$ schwach gegen $\mathbb{P}_X$ konvergiert. Wir schreiben $X_n \xrightarrow{\mathcal{D}} X$.

78 **Satz: schwache Konvergenz im Reellen**

Seien $X, (X_n)_{n\in\mathbb{N}}$ reellwertige Zufallsvariablen mit Verteilungsfunktionn $F_X$ und $F_{X_n}$. Dann sind die folgenden Aussagen äquivalent:

1. $X_n \xrightarrow{\mathcal{D}} X$
2. $F_{X_n}(c) \to F_X(c)$ an allen Stellen $c$ an denen $F_X$ stetig ist.

**Beweis** (78)

*a)* $\Rightarrow$ *b)* Sei $c$ eine Stetigkeitsstellen von $F_X$. Wähle dann Folgen stetiger und beschränkter Funktionen $(g_m)_{m\in\mathbb{N}}, (h_m)_{m\in\mathbb{N}}$ sodass

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

Dann gilt diese Ungleichung auch für die Erwartungswerte von $X_n$ und für die von $g_m$ und $h_m$ ist die Konvergenz wegen schwacher Konvergenz von $X_n$ bekannt:

$$F_X(c-\frac{1}{m}) \leq \mathbb{E}(g_m(X)) \leq F_{X_n}(c) \leq \mathbb{E}(h_m(X)) \leq F_X(c+\frac{1}{m})$$

Da $F_X$ bei $c$ stetig ist, gilt dann $F_{X_n}(c) \to F_X(c)$.

*b)* $\Rightarrow$ *a)* Wir wählen $f \in C_b(\mathbb{R},\mathbb{R})$ (stetig und beschränkt) und $\varepsilon > 0$ beliebig. $F_Y$ hat höchstens abzählbar viele Sprungstellen. Also können wir eine Zerlegung der reellen Achse in $-\infty < c_1 < c_2 < \cdots < c_m < \infty$ so finden, dass $F_Y$ bei den $c_i$ stetig ist, $F_Y(c_1) < \varepsilon$ und $F(c_m) > 1-\varepsilon$ sowie

$$\sup_{x\in[c_i,c_{i+1}]} |f(x) - f(c_i)| \leq \varepsilon$$

Dies ist möglich, da $f$ auf $[c_1, c_m]$ gleichmäßig stetig ist. Dann können wir abschätzen:

$$\mathbb{E}(f(Y_n)) = \mathbb{E}(f(Y_n)\mathbf{1}_{\{\ \omega\in\omega \mid Y_n(\omega)\leq c_i \vee Y_n(\omega)>c_m\ \}}) + \sum_{i=2}^{m} \mathbb{E}(f(Y_n)\mathbf{1}_{\{\ \omega\in\Omega \mid Y_n(\omega)\in(c_{i-1},c_i]\ \}})$$

$$\leq \|f\|_\infty 2\varepsilon + \sum_{i=2}^{m} (f(c_i)+\varepsilon)(F_{Y_n}(c_i)) - F_{Y_n}(c_{i-1})$$

$$\Rightarrow \lim_{n\to\infty} \mathbb{E}(f(Y_n)) \leq 2\varepsilon\|f\|_\infty + \underbrace{\mathbb{E}(f(Y)) + 2\varepsilon + \|f\|_\infty 2\varepsilon}_{\geq \mathbb{E}(\sum_{i=2}^m f(c_i)\mathbf{1}_{\{Y\in(c_{i-1},c_i]\}})} = \mathbb{E}(f(Y)) + 2\varepsilon(2\|f\|_\infty + 1)$$

$$\Rightarrow \limsup_{n\to\infty} \mathbb{E}(f(Y_n)) \leq \mathbb{E}(f(Y))$$

Analog mit $-f$ statt $f$ liefert $\liminf_{n\to\infty} \mathbb{E}(f(Y_n)) \geq \mathbb{E}(f(Y))$ und damit die Behauptung.

## 6.1 Beziehungen zu anderen Konvergenzarten

79 **Satz: fast-sichere Konvergenz und Konvergenz in Verteilung**
Seien $Y_n, Y\colon \Omega \to \mathbb{R}^d$ Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ und $Y_n \to Y$ $\mathbb{P}$-fast-sicher. Dann gilt auch $Y_n \xrightarrow{\mathcal{D}} Y$.

**Beweis** (79) Sei $f \in C_b(\mathbb{R}^d, \mathbb{R})$. Dann gilt wegen der Stetigkeit $f(Y_n) \to f(Y)$ $\mathbb{P}$-fast-sicher und $|f(Y_n)|\|f\|_\infty \in L^1(\mathbb{P})$. Mit dem Satz der majorisierten Konvergenz folgt dann $\lim_{n\to\infty} \mathbb{E}(f(Y_n)) = \mathbb{E}(f(Y))$.

80 **Satz: Konvergenz von Unterteilfolgen**
Sind $Y_n, Y\colon \Omega \to \mathbb{R}^d$ Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ so sind die folgenden Aussagen äquivalent:

1. $Y_n \xrightarrow{\mathbb{P}} Y$
2. Jede Teilfolge $(Y_{n_k})_{k\in\mathbb{N}}$ hat eine Unterteilfolge $(Y_{n_{k_l}})_{l\in\mathbb{N}}$ so, dass $\lim_{l\to\infty} Y_{n_{k_l}} = Y$ $\mathbb{P}$-fast-sicher.

**Beweis** (80)

*b)* $\Rightarrow$ *a)* Wir nehmen an $Y_n \xrightarrow{\mathbb{P}} Y$ gilt nicht. Dann gilt

$$\exists \varepsilon > 0, \delta > 0, n_k\colon \mathbb{E}(|Y_{n_k} - Y| \geq \varepsilon) \geq \delta$$

$Y_{n_k}$ konvergiert jedoch fast-sicher gegen $Y$ und damit $Y_{n_k} \xrightarrow{\mathbb{P}} Y$ ↯

*a)* $\Rightarrow$ *b)* Sei $(Y_{n_k})$ eine Teilfolge von $Y_n$ so gilt für alle $\varepsilon > 0$: $\lim_{k\to\infty} \mathbb{P}(|Y_{n_k} - Y| \geq \varepsilon) = 0$ also

$$\exists K_1 \in \mathbb{N}\colon \forall k \geq K_1\colon \mathbb{P}(|Y_{n_k}| \geq \frac{1}{2}) \leq \frac{1}{2}$$
$$\exists K_2 \in \mathbb{N}\colon \forall k \geq K_2\colon \mathbb{P}(|Y_{n_k}| \geq \frac{1}{2^2}) \leq \frac{1}{2^2}$$
$$\dots$$

Dann definieren wir für eine Unterteilfolge $Z_j = Y_{n_{k_j}}$ die $A_j := \left\{ |Z_{j+1} - Z_j| \geq \frac{1}{2^j} \right\}$. Damit ist $\sim_{j=1}^{\infty} \mathbb{P}(A_j) = \sum_{j=1}^{\infty} \frac{1}{2^j}$. Mit Satz 43 (Borel-Cantelli-Lemma) wissen wir nun, dass nur endlich viele der $A_j$ eintreten, das heißt die Wahrscheinlichkeit dass $Z_j$ eine Cauchyfolge ist, ist 1 und somit gilt die Behauptung.

81 **Satz: Majorisierte Konvergenz mit stochastischer Konvergenz**
Seien $Y_n, Y\colon \Omega \to \mathbb{R}^d$ Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit $Y_n \to Y$ $\mathbb{P}$-fast-sicher und es gelte $|Y_n| \leq Z$ für ein $Z \in L^1(\mathbb{P})$. Dann gilt $Y_n \xrightarrow{L^1} Y$, das heißt $\|Y_n - Y\|_{L^1} = \mathbb{E}(|Y_n - Y|) \to 0$ und $Y \in L^1(\mathbb{P})$. Insbesondere also $\mathbb{E}(Y_n) \to \mathbb{E}(Y)$.

**Beweis** (81) Angenommen $Y_n \xrightarrow{L^1} Y$ gilt nicht, das heißt $\exists \varepsilon > 0, Y_{n_k}\colon \forall k\colon \mathbb{E}(|Y_{n_k} - Y|) \geq \varepsilon$. Nach Satz 80 (Konvergenz von Unterteilfolgen) können wir annehmen dass $Y_{n_k} \to Y$ $\mathbb{P}$-fast-sicher, daher gilt nach dem klassischen Theorem der majorisierten Konvergenz dass $Y_{n_k} \xrightarrow{L^1} Y$ ↯.

82 **Satz: Stochastische Konvergenz und Konvergenz in Verteilung**
Seien $Y, Y_n\colon \Omega \to \mathbb{R}^d$ Zufallsvariablen mit $Y_n \xrightarrow{\mathbb{P}} Y$ so gilt $Y_n \xrightarrow{\mathcal{D}} Y$.

**Beweis** (82) Sei $f \in C_b(\mathbb{R}^d, \mathbb{R})$, dann gilt $f(Y_n) \xrightarrow{L^1} f(Y)$. Wegen Satz 80 (Konvergenz von Unterteilfolgen) und Satz 81 (Majorisierte Konvergenz mit stochastischer Konvergenz) folgt dann

$$\|f(Y_n)\| \leq \|f\|_\infty \in L^1(\mathbb{P}) \Rightarrow f(Y_n) \xrightarrow{L^1} f(Y)$$

also $\mathbb{E}(f(Y_n)) \to \mathbb{E}(f(Y))$.

**Bemerkung** Die Umkehrung gilt nicht, betrachten wir beispielsweise das folgende Gegenbeispiel:

$X \sim \mathcal{N}(0,1)$, $Y_n := (-1)^n X$. Dann ist $Y_n \xrightarrow{\mathcal{D}} P_X$ aber $Y_n$ konvergiert nicht stochastisch.

83 **Lemma:**

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

**Bemerkung:** $F$ induziert über $\mu((a,b]) = F(b) - F(a)$ ein Maß auf $\mathcal{B}$, welches im Allgemeinen jedoch kein Wahrscheinlichkeitsmaß zu sein braucht.

**Beweis** (83) Wir zeigen zuerst, dass es eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ gibt, sodass $\forall q\colon H(q) := \lim_{k\to\infty} F_{n_k}$ existiert.

Dazu wählen wir eine Nummerierung der rationalen Zahlen $\mathbb{Q} = \left\{ q_m \mid m \in \mathbb{N} \right\}$. Da für jedes $q$: $F_n(q) \in [0,1]$ ist, können wir nach Bolzano-Weierstraß konvergente Teilfolgen wählen sodass $F_{n_k^1}(q_1)$ konvergiert, $F_{n_k^2}(q_1)$, $F_{n_k^2}(q_2)$ konvergieren, also $F_{n_k^j}$ bei $q_1, \cdots, q_j$ konvergiert. Wir wählen dann die Diagonalfolge $G_k := F_{n_k^k}$ und wissen somit, dass $G_k$ auf ganz $\mathbb{Q}$ konvergiert. Dann definieren wir $H(q) := \lim_{k\to\infty} G_k(q)$ für $q \in \mathbb{Q}$ und $H$ ist wachsend auf $Q$ sowie $H(q) \in [0,1]$.

Dann setzen wir $H$ auf ganz $\mathbb{R}$ zu $F$ fort:

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

Damit ist $F$ wachsend und rechtsstetig. Es bleibt zu zeigen, dass auch $\lim_{k\to\infty} G_k(c) = F(c)$ für alle Stetigkeitsstellen $c$ von $F$.

Sei $c$ also eine beliebige Stetigkeitsstelle von $F$ und $\varepsilon > 0$, dann existieren $r, s \in \mathbb{Q}$ mit $r < c < s$ und

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

$$\Rightarrow \limsup_{k\to\infty} G_k(c) \leq \limsup_{k\to\infty} G_k(s) \overset{s\in\mathbb{Q}}{=} H(s) \leq F(s) \leq F(c) + \varepsilon$$

$$\liminf_{k\to\infty} G_k(c) \geq \liminf_{k\to\infty} G_k(r) = H(r) = F(r) \geq F(c) - \varepsilon$$

Da $\varepsilon > 0$ beliebig gewählt war, gilt somit $\lim_{k\to\infty} G_k(c) = F(c)$.

84 **Definition: „gleichgradige Straffheit“**

Eine Familie von Wahrscheinlichkeitsmaßen $(\mu_i)_{i\in I}$ heiß *gleichgradig straff* falls

$$\forall \varepsilon > 0 \colon \exists M > 0 \colon \forall i \in I \colon \mu_i \left( \overline{(-M, +M]} \right) < \varepsilon$$

**Beispiel** für eine nicht gleichgradig straffe Folge: $\mu_n := \mathcal{U}([0,n])$ (Gleichverteilung). Dann ist der Grenzwert das Nullmaß.

**Beispiel** für gleichgradig straffe Mengen Familien:

1. Ist $I$ endlich, so ist $(\mu_i)_{i\in I}$ gleichgradig straff
2. Oft ist $I = \mathbb{N}$. Dan ist die Folge $(\mu_n)_{n\in\mathbb{N}}$ genau dann gleichgradig straff, wenn für jedes $N \in \mathbb{N}$ die Folge $(\mu_n)_{n\geq N}$ gleichgradig straff ist.
3. Sind $(\mu_i)_{i\in I_1}$ und $(\mu_i)_{i\in I_2}$ gleichgradig straffe Familien, dann ist auch $(\mu_i)_{i\in I_1\cup I_2}$ gleichgradig straff.
4. Falls $\mu_n \xrightarrow{w} \mu$ gilt, so ist $(\mu_n)_{n\in\mathbb{N}}$ gleichgradig straff.

85 **Satz: Hellysches Selektionsprinzip**

Sei $(\mu_n)_{n\in\mathbb{N}}$ eine gleichgradig stetige Folge von Wahrscheinlichkeitsmaßen auf $\mathcal{B}^1$. Dann gibt es eine Teilfolge $(\mu_{n_k})_{k\in\mathbb{N}}$ und ein Wahrscheinlichkeitsmaß $\mu$ sodass $\mu_{n_k} \xrightarrow[k\to\infty]{w} \mu$.

**Beweis** (85) Seien

1. $F_n$ die Verteilungsfunktionen der $\mu_n$,
2. $G_k$ und $F$ wie aus Lemma 83
3. $\nu_k$ seien die durch $G_k$ über $\nu_k((a,b]) := G_k(b) - G_k(a)$ definierten Maße

Dann müssen wir zeigen, dass $F$ die Verteilungsfunktion eines Wahrscheinlichkeitsmaßes ist.

Sei $\varepsilon > 0$, dann existiert ein $M > 0$ sodass für jedes $n$ gilt: $\mu_n(\overline{(-M, +M)}) < \varepsilon$. Wähle nun $y > 1$ so, dass $F$ stetig bei $y$ und $-y$ ist. Dann gilt

$$(1 - F(y)) + F(-y) = \lim_{k\to\infty} \underbrace{(1 - G_k(y)) + G_k(-y)}_{\nu_k(\overline{(-y,y]})} \leq \limsup_{k\to\infty} \nu_k(\overline{(-M, +M]}) \leq \varepsilon$$

Als $\mu$ wählen wir also das zu $F$ gehörige Wahrscheinlichkeitsmaß.

86 **Satz: Stetigkeitssatz von Paul Lévy**

Seien $\mu, \mu_n, n \in \mathbb{N}$ Wahrscheinlichkeitsmaße auf $(\mathbb{R}^d, \mathcal{B}^d)$ mit charakteristischen Funktionen $\varphi = \hat{\mu}, \varphi_n = \hat{\mu}_n$. Dann gilt

1. $\mu_n \xrightarrow{w} \mu \quad \Rightarrow \quad \forall u \in \mathbb{R}^d: \varphi_n(u) \to \varphi(u)$
2. Falls $\forall u \in \mathbb{R}^d: \lim_{n\to\infty} \varphi_n(u) = \psi(u)$ für eine Funktion $\psi: \mathbb{R}^d \to \mathbb{C}$, die in 0 stetig ist, dann ist $\varphi$ die charakteristische Funktion eines Wahrscheinlichkeitsmaßes $\nu$ auf $\mathcal{B}^d$ und $\mu_n \xrightarrow{w} \nu$.

**Bemerkung** In der Tat kann man in *a*) sogar die stärkere Aussage zeigen, dass $\varphi_n$ auf Kompakta gleichmäßig gegen $\varphi$ konvergiert.

**Beweis** (86)

1. Sei $X \sim \mu, X_n \sim \mu_n$, dann ist

$$\begin{aligned}\varphi_n(u) &= \mathbb{E}(\exp(iuX_n)) = \mathbb{E}(\cos(uX_n)) + i\mathbb{E}(\sin(uX_n)) \\ &\xrightarrow{n\to\infty} \mathbb{E}(\cos(uX)) + i\mathbb{E}(\sin(uX)) = \mathbb{E}(\exp(iuX)) = \varphi(u)\end{aligned}$$

**Beispiel**

1. Seien $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ Zufallsvariablen für beliebige $\mu_n \in \mathbb{R}, \sigma_n > 0$ mit $\mu_n \to \mu \in \mathbb{R}$. Dann gilt

$$\begin{aligned}&\varphi_n(u) = \exp(iu\mu_n - \frac{1}{2}u^2\sigma_n^2) \xrightarrow{n\to\infty} \exp(iu\mu - \frac{1}{2}\sigma^2u^2) \\ &\Rightarrow X_n \xrightarrow{\mathcal{D}} \mathcal{N}(\mu, \sigma^2)\end{aligned}$$

2. Seien $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ und $X_n \xrightarrow{\mathcal{D}} X$, dann wisen wir:

$$\begin{aligned}&\forall u \in \mathbb{R}: \lim_{n\to\infty} \varphi_{X_n}(u) = \varphi_X \\ &\Rightarrow \mu = \lim_{n\to\infty} \mu_n \wedge \sigma = \lim_{n\to\infty} \sigma_n \wedge X \sim \mathcal{N}(\mu, \sigma^2)\end{aligned}$$

3. Zusammen bedeutet das: $X_n \xrightarrow{\mathcal{D}} X$ genau dann wenn $\mu = \lim_{\to\infty} \mu_n, \sigma = \lim_{n\to\infty} \sigma_n$ existieren und $X \sim \mathcal{N}(\mu, \sigma^2)$. Die Normalverteilungsfamilie ist also abgeschlossen bezüglich Verteilungskonvergenz.

**Übung** Zeigen Sie analoge Aussagen für Poisson- oder Exponentialverteilte Zufallsvariablen.

**Beispiel** für die Anwendung von Satz 86 (Stetigkeitssatz von Paul Lévy):
Betrachte die Zufallsvariablen $X_n \sim \text{Poisson}(\lambda_n)$, $\lambda_n := n$, $Z_n := \frac{X_n - \mathbb{E}(X_n)}{\sqrt{\mathbb{V}(X_n)}} = \frac{X_n - n}{\sqrt{n}}$. Dann gilt

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

denn:

$$\varphi_{Z_n}(u) = \mathbb{E}\left(\exp(iu\frac{X_n - n}{\sqrt{n}})\right) = e^{-iu\sqrt{n}}\varphi_{X_n}(\frac{u}{\sqrt{n}}) = \varphi_{\mathcal{N}(0,1)}(u) = e^{-iu\sqrt{n}}\exp(n(e^{iu/\sqrt{n}} - 1))$$

$$\overset{\text{Taylor}}{=} \exp\left(-iu\sqrt{n} + iu\sqrt{n} + n\left(-\frac{u^2}{2n} - i\frac{u^3}{3!\,n^{1.5}} + \frac{u^4}{4!\,n^2} + \cdots\right)\right) = \exp\left(-\frac{u^2}{2} + n\,\mathrm{o}\left(\frac{1}{n}\right)\right) \to e^{-\frac{u^2}{2}}$$

weil

$$|n\sum_{k=3}^{\infty}\frac{(iu)^k}{k!\,n^{\frac{k}{2}}}| \le \frac{1}{\sqrt{n}}\sum_{k=3}^{\infty}|\frac{(iu)^k}{k!}n^{\frac{k}{2}}| \le \frac{1}{\sqrt{n}}\underbrace{\sum_{k=3}^{\infty}\frac{|u|^k}{k!}}_{<\infty} \to 0$$

**Motivation zum Zentralen Grenzwertsatz** Man kann die $X_n \sim \text{Poisson}(n)$ aus obigem Beispiel auch als $X_n = \sum_{i=1}^n Y_i$ mit unabhängigen $Y_i \sim \text{Poisson}(1)$ erzeugen.

Oder betrachtet man zum Beispiel eine Folge iid Zufallsvariablen $Y_j$ mit $\mathbb{P}(Y_j = 1) = 1 - \mathbb{P}(Y_j = -1) = \frac{1}{2}$. Dann hat $Z_n := \frac{\sum_{j=1}^n Y_j - 0}{\sqrt{n}}$ die charakteristische Funktion

$$\varphi_n(u) = \left(\varphi_{Y_1}\left(\frac{u}{\sqrt{n}}\right)\right)^n = \left(\frac{1}{2}e^{i\frac{u}{\sqrt{n}(+1)}} + \frac{1}{2}e^{i\frac{u}{\sqrt{n}}(-1)}\right) = \cos^n\left(\frac{u}{\sqrt{n}}\right) = \exp(n\ln(\cos\left(\frac{u}{\sqrt{n}}\right)))$$

$$\xRightarrow{\text{l'Hôpital}} \lim_{n\to\infty}\varphi_n(u) = \lim_{n\to\infty}\exp(\frac{-\sin\left(\frac{u}{\sqrt{n}}\right)(\frac{-u}{2n^{1.5}})}{-\frac{1}{n^2}\cos\left(\frac{u}{\sqrt{n}}\right)}) = \exp(-\frac{u^2}{2}) = \varphi_Z(u)$$

für $Z \sim \mathcal{N}(0,1)$. Und somit gilt wieder nach Satz 86 (Stetigkeitssatz von Paul Lévy): $Z_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

**Bemerkung** Der Zentrale Grenzwertsatz zeigt nun, dass eine entsprechende Aussage wie in den vorigen Beispielen allgemeiner für beliebige zugrundeliegende Verteilungen der unabhängigen Summanden $Y_j$ gilt. Etwas genauer gilt, dass Summen „vieler“ unabhängiger gleichgroßer Zufallsgrößen approximativ Gauß-verteilt sind und nach Standardisierung standard-Gauß-verteilt.

87 **Satz: Zentraler Grenzwertsatz**

Seien $(X_n)_{n\in\mathbb{N}}$ iid und reellwertig mit $\mu := \mathbb{E}(X_n)$ sowie $\sigma^2 := \operatorname{Var}(X_n) \in (0,\infty)$.

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

Dann gilt: $S_n^* \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

88 **Lemma: Hilfsaussage**

Sei $(C_n)_{n\in\mathbb{N}}, c_n \in \mathbb{C}$ mit $c := \lim_{n\to\infty} c_n \in (0,\infty) \subseteq \mathbb{R}$, dann gilt

$$\lim_{n\to\infty}\left(1-\frac{c_n}{n}\right)^n = e^{-c}$$

**Beweis (Satz 87 (Zentraler Grenzwertsatz))** $X_n$ hat endliche erste und zweite Momente, also ist die charakteristische Funktion $\varphi := \varphi_{X_n-\mu}$ zweimal stetig differenzierbar. Dann ergibt die Taylorentwicklung:

$$\varphi(u) = \varphi(0) + \underbrace{\varphi'(0)}_{=0}u + \varphi''(0)\frac{u^2}{2} + \mathcal{O}_{u\to 0}(u^3)$$

$$\Rightarrow \varphi_{S_n^*}(u) = \left(\varphi_{X_1-\mu}\left(\frac{u}{\sigma\sqrt{n}}\right)\right)^n = \left(\varphi(0) + \varphi''(0)\frac{u^2}{2\sigma^2}n + \mathcal{O}_{u\to 0}\left(\left(\frac{u}{\sigma\sqrt{n}}\right)^3\right)\right)^n$$

$$= \left(1 - \frac{\frac{1}{2}u^2 - \mathcal{O}\left(\frac{u^3}{\sigma^3\sqrt{n}}\right)}{n}\right) \xrightarrow{\text{Lemma 88 (Hilfsaussage)}} e^{-\frac{u^2}{2}}$$

**Beweis (Lemma 88 (Hilfsaussage))**

- Wir zeigen zuerst induktiv, dass $z_1,\cdots,z_n,w_1,\cdots,w_n \in \mathbb{C}$ mit $|z_j| \le 1, |w_j| \le 1$ gilt

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

  Der Induktionsanfang für $n = 1$ ist offensichtlich erfüllt. Der Induktionsschritt ergibt sich als:

  $$\left|\prod_{j=1}^{n+1} z_j - \prod_{j=1}^{n+1} w_j\right| \le \left|\prod_{j=1}^{n+1} z_j - z_{n+1}\prod_{j=1}^{n} w_j\right| + \left|z_{n+1}\prod_{j=1}^{n} w_j - \prod_{j=1}^{n+1} w_j\right|$$

  $$= \underbrace{|z_{n+1}|}_{\le 1}\underbrace{\left|\prod_{j=1}^{n} z_j - \prod_{j=1}^{n} w_j\right|}_{\le\sum_{j=1}^{n}|z_j-w_j|} + \underbrace{\left|\prod_{j=1}^{n} w_j\right|}_{\le 1}|z_{n+1} - w_{n+1}| \le \sum_{j=1}^{n}|z_j - w_j| + |z_{n+1} - w_{n+1}|$$

- Weiter gilt für $b \in \mathbb{C}$ mit $|b| \leq 1$:

$$|e^{-b} - (1-b)| = |\sum_{k=2}^{\infty} \frac{(-b)^k}{k!}| \leq \sum_{k=2}^{\infty} |\frac{(-b)^k}{k!}| \leq |b|^2 \sum_{k=1}^{\infty} \frac{1}{2^k} \leq |b|^2$$

- Sei nun $(c_n)_{n\in\mathbb{N}}$ wie vorausgesetzt, dann gilt für hinreichend große $n$ sodass $|1 - \frac{c_n}{n}| \leq 1$:

$$|(1-\frac{c_n}{n})^n - e^{-c_n}| \leq \sum_{j=1}^{n} |1 - \frac{c_n}{n} - e^{-\frac{c_n}{n}}| = n|1 - \frac{c_n}{n} - e^{-\frac{c_n}{n}}|$$

$$\leq n|\frac{c_n}{n}|^2 = |\frac{c_n^2}{n}| \xrightarrow{n\to\infty} 0$$

$$\Rightarrow \lim_{n\to\infty} (1 - \frac{c_n}{n}^n) = e^{-c}$$

**Bemerkung** Seien $X_n$ iid und reellwertig. Aus dem starken Gesetz der Großen Zahlen wissen wir

$$S_n := \frac{1}{n} \sum_{j=1}^{n} X_j \xrightarrow{n\to\infty} \mu := \mathbb{E}(X_1) \quad \mathbb{P}\text{-fast-sicher}$$

**Frage** Wie schnell konvergiert dies?

**Vorüberlegung** Für die 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 $\lim\sup_{n\to\infty} n^\alpha |Y_n| =: c < \infty$ ist. Ein solches $\alpha$, sodass $n^\alpha |S_n - \mu|$ $\mathbb{P}$-fast-sicher gegen ein $c \in \mathbb{R}$ konvergiert gibt es nicht! Allerdings zeigt Satz 87 (Zentraler Grenzwertsatz), dass Konvergenz in Verteilung gegen eine endliche Zufallsvariable vorliegt:

$$\sqrt{n}(S_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$$

In diesem Sinne kann man sagen die Konvergenzordnung ist $\sqrt{n}$.

**Bemerkung** Es gibt Verallgemeinerungen von Satz 87 (Zentraler Grenzwertsatz).

- Die Annahme der identischen Verteilung kann abgeschwächt werden. Dies ist gerade für Anwendungen wichtig.
- Sehr scharfe Bedingungen für Satz 87 (Zentraler Grenzwertsatz) sind beispielsweise die *Lindenberg-Bedingungen*. (Kle08)
- Es gibt auch eine mehrdimensionale Version. (Kle08)
- Beispiel für die Anwendung von Satz 87 (Zentraler Grenzwertsatz): Bestimmung von „Vertrauensbereichen“ (Konfidenzbereichen) für unbekannte Verteilungsparameter die aus Daten geschätzt werden.