Skript zur Vorlesung

Wahrscheinlichkeitstheorie I

von

Prof. Dr. Karl-Theodor Sturm

Universität Bonn

Inhaltsverzeichnis

Teil I

W-Räume und Zufallsvariablen

Einleitung

Literaturhinweise

- R.Durrett: Probability Theory and Examples, Wadsworth & Brooks/Cole
- H.Bauer: Maßtheorie, De Gruyter
- H.Bauer: Wahrscheinlichkeitstheorie, 4. Auflage, De Gruyter, 1991

Ziel der Wahrscheinlichkeitstheorie

- Einsichten in „vom Zufall“ gesteuerte Abläufe (Glücksspiel, Molekularbewegung), Formulierung von Gesetzmäßigkeiten.
- Die Wahrscheinlichkeitstheorie erklärt **nicht** was der Zufall ist

Beispiel:
Roulette: 38 Felder (18 rot, 18 schwarz, 2 grün),
Einsatz: 1 DM.
Gewinn nach einem Spiel: P(S_1=1 DM)= $\frac{18}{38}$, P(S_1=-1 DM)=$\frac{20}{38}$
Zu erwartender Gewinn pro Spiel: E[S_1] $= \mu = 0,0526$ DM.
Sei S_n der Gewinn nach n Spielen, $\frac{S_n}{n}$ der mittlere Gewinn pro Spiel nach n Spielen. (E[S_n] $= n\mu$. Es wird gezeigt (s.u.)):
Mit dem **Starken Gesetz der Großen Zahlen** folgt hier mit 100% Wahrscheinlichkeit:

$$\frac{S_n}{n} \to \mu \text{ für } n \to \infty$$

Mit dem **Zentralen Grenzwertsatz** folgt z. B.:

$$\begin{array}{lll} \text{mit 30\% Wahrscheinlichkeit:} & S_{100} \geq 0 & E(S_{100}) = -5,26 \\ \text{mit 99\% Wahrscheinlichkeit:} & S_{10.000} \geq -296 & E(S_{10.000}) = -526 \end{array}$$

Methode (A.N. Kolmogorov 1933)
Das mathematische Modell zur Beschreibung zufälliger Ereignisse ist ein *Wahrscheinlichkeitsraum (= normierter Maßraum)*

Kapitel 1

Wahrscheinlichkeitsräume

Definition 1.1. *Ein* **W-Raum** *ist ein Tripel* $(\Omega, \mathcal{A}, P)$*, wobei* Ω *eine beliebige nicht-leere Menge ist,* $\mathcal{A}$ *eine* $\sigma-$*Algebra auf* Ω *und* P *ein W-Maß auf* $(\Omega, \mathcal{A})$.

Ω ist die Menge der (im Modell) möglichen Fälle, der möglichen Ausgänge des Zufallsexperiments.

Beispiele 1.2.

1. Münzwurf	Ω_1	$= \{\text{Kopf,Zahl}\}$
2. Würfeln	Ω_2	$= \{1, 2, 3, 4, 5, 6\}$
3. n-maliges Würfeln	Ω_3	$= \{\omega = (x_1, \cdots, x_n) : x_i \in \{1, \cdots, 6\}\}$
4. unendlich oft würfeln	Ω_4	$= \{\omega = (x_1, x_2, \cdots) : x_i \in \{1, \cdots, 6\}\} = \Omega_2^{\mathbb{N}}$
5. Zufallszahl in $[0, 1]$	Ω_5	$= [0, 1]$
6. stetige stochastische Prozesse (Zeit $\mathbb{R}_+$, Zustandsraum $\mathbb{R}$) („Brownsche Molekularbewegung“)	Ω_6	$= \mathcal{C}(\mathbb{R}_+)$

$\mathcal{A}$ ist die Menge der (im Modell zugelassenen, in Betracht gezogenen, bewertbaren) Ereignisse.

Definition 1.3. $\mathcal{A}$ *heißt* $\sigma-$**Algebra** *auf* Ω*, falls gilt:*

$$\mathcal{A} \subset \mathcal{P}(\Omega),\ \Omega \in \mathcal{A}$$

$$A \in \mathcal{A} \Longrightarrow A^c = \Omega \setminus A \in \mathcal{A}$$

$$A_i \in \mathcal{A}\ (\forall i \in \mathbb{N}) \Longrightarrow \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$$

Das Paar $(\Omega, \mathcal{A})$ *wird auch* **Meßraum** *genannt.*
Die Mengen $A \in \mathcal{A}$ *heißen* ***meßbar***

Bemerkungen:
a) σ steht für abzählbare Summe.
b) Mit $A_i \in \mathcal{A}\, (i \in \mathbb{N})$ ist auch

$$\bigcap_{i \in \mathbb{N}} A_i = \left(\bigcup_{i \in \mathbb{N}} A_i^c \right)^c \in \mathcal{A}$$

ebenso

$$\bigcup_n \bigcap_{i \geq n} A_i \in \mathcal{A} \qquad \text{u.ä.}$$

c) Stets ist die Potenzmenge $\mathcal{P}(\Omega)$ eine $\sigma-$Algebra auf Ω.
d) Sind $\mathcal{A}_i$ für $i \in I$ (=beliebige, nicht-leere Indexmenge) $\sigma-$Algebren auf Ω, so auch $\bigcap\limits_{i \in I} \mathcal{A}_i$ (i.a. jedoch <u>nicht</u> $\bigcup\limits_{i \in I} \mathcal{A}_i$).

Konstruktion von $\sigma-$Algebren: Für jedes $\mathcal{A}_0 \subset \mathcal{P}(\Omega)$ ist

$$\sigma(\mathcal{A}_0) := \bigcap_{\substack{\mathcal{A}_0 \subset \mathcal{A} \\ \sigma-\text{Algebra auf } \Omega}} \mathcal{A}$$

die kleinste σ-Algebra auf Ω, die $\mathcal{A}_0$ enthält. Sie heißt auch die <u>„von $\mathcal{A}_0$ erzeugte $\sigma-$Algebra“</u>.

Beispiele 1.4.
a) In der Situation von obiger Bemerkung d):

$$\bigvee_{i \in I} \mathcal{A}_i := \sigma \left(\bigcup_{i \in I} \mathcal{A}_i \right)$$

b) Sind $(\Omega_1, \mathcal{A}_1)$ und $(\Omega_2, \mathcal{A}_2)$ Meßräume, so definiert

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma \left(\{ A_1 \times A_2 : \; A_1 \subset \mathcal{A}_1, \, A_2 \subset \mathcal{A}_2 \} \right)$$

eine $\sigma-$Algebra auf $\Omega_1 \times \Omega_2$ (<u>„Produkt-$\sigma-$Algebra“</u>).
($\{ A_1 \times A_2 : \; A_i \in \mathcal{A}_i \,, i = 1, 2 \}$ ist i.a. <u>keine</u> $\sigma-$Algebra!)
c) Sei Ω topologischer Raum mit $\mathcal{A}_0$ =Menge der offenen Mengen in Ω.
$\mathcal{B}(\Omega) = \sigma(\mathcal{A}_0)$ heißt die **Borelsche $\sigma-$Algebra** auf Ω. Die $A \in \mathcal{B}(\Omega)$ heißen <u>Borel-Mengen</u>, dazu gehören offene Mengen, abgeschlossene Mengen,

F_σ-Mengen (=abzählbare Vereinigungen abgeschlossener Mengen), G_σ-Mengen (=abzählbarer Durchschnitt offener Mengen), $G_{\delta,\sigma}$, $F_{\sigma,\delta}$, $\cdots$.
Für $\Omega = \mathbb{R}$ ist jedoch $\mathcal{B}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$.
(Denn: sei c = Mächtigkeit von $\mathbb{R} \Rightarrow c$ = Mächtigkeit von $\mathcal{B}(\mathbb{R})$
$\Rightarrow 2^c >$ Mächtigkeit von $\mathcal{B}(\mathbb{R})$)

Unterschied zwischen Topologie und $\sigma-$Algebra:
Sei $\mathcal{A}_0 \subset \mathcal{P}(\Omega)$ beliebig. Sei $\tau(\mathcal{A}_0)$ die von $\mathcal{A}_0$ erzeugte Topologie und $\sigma(\mathcal{A}_0)$ die von $\mathcal{A}_0$ erzeugte $\sigma-$Algebra.
Dann gilt:

$$\begin{aligned} A_i \in \mathcal{A}_0 \ (i \in I) \ \Rightarrow \bigcup_{i\in I} A_i \ &\in \ \tau(\mathcal{A}_0) \quad \text{für beliebige I} \\ &\in \ \sigma(\mathcal{A}_0), \quad \text{falls I abzählbar} \\ A \in \mathcal{A}_0 \ \Rightarrow \ A^c \ &\in \ \sigma(\mathcal{A}_0) \end{aligned}$$

Beispiele 1.5. • $\Omega = \mathbb{R}$, $\mathcal{A}_0 = \{\, \{x\} : x \in \mathbb{R}\,\}$

$$\Rightarrow \sigma(\mathcal{A}_0) = \{\, A \subset \mathbb{R} : A \quad \text{abzählbar oder} \quad A^c \quad \text{abz.}\} \subseteq \tau(\mathcal{A}_0) = \mathcal{P}(\mathbb{R})$$

• $\Omega = \mathbb{R}$, $\mathcal{A}_0 = \{\,]x, y[: x, y \in \mathbb{Q}\} \Rightarrow \sigma(\mathcal{A}_0) = \mathcal{B}(\mathbb{R}) \supseteq \tau(\mathcal{A}_0) = \tau(\mathbb{R})$

denn: Offenbar $\mathcal{A}_0 \subseteq \{I \in \mathcal{P}(\mathbb{R}) : I \text{ offen }\} = \tau$
$\Rightarrow \sigma(\mathcal{A}_0) \subseteq \sigma(\tau) =: \mathcal{B}(\mathbb{R})$
Umgekehrt: $\forall I \in \tau : \quad \exists I_n \in \mathcal{A}_0 : I = \bigcup_{n\in\mathbb{N}} I_n$
$\Rightarrow I \in \sigma(\mathcal{A}_0)$
$\Rightarrow \sigma(\tau) = \sigma(\mathcal{A}_0)$. $\square$

Probabilistische Interpretation:

$A \in \mathcal{A}$	heißt **Ereignis** („A tritt ein“),
z. B. $\emptyset =$	unmögliches Ereignis
$\Omega =$	sicheres Ereignis
$A = \{\omega\} =$	Elementarereignis
$A^c = \Omega \backslash A$	„A tritt nicht ein“
$\bigcup A_i$	„mindestens eines der A_i tritt ein“
$\bigcap A_i$	„jedes der A_i tritt ein“
$\bigcup_n \bigcap_{i\geq n} A_i$	„alle außer endlich vielen der A_i treten ein“ oder „schließlich alle A_i treten ein“
$\bigcap_n \bigcup_{i\geq n} A_i$	„unendlich viele der A_i treten ein“

Beispiele 1.6. (vgl. Beginn des Kapitels)

ad 1) (Münzwurf): $A = \{\text{Kopf}\}$ „Kopf liegt oben“

ad 2) (Würfeln): $A = \{2, 4, 6\}$ „eine gerade Zahl liegt oben“

ad 3) ($n-$mal Würfeln): „k-mal die 6“:

$$A = \{\ \omega = (x_1, \cdots, x_n) :\ x_i \in \{1, \cdots, 6\},\ |\{i = 1, \cdots, n :\ x_i = 6\}| = k\ \}$$

($|\cdot|$ =Mächtigkeit=Anzahl der Elemente)

„gewürfelte Augenzahl ist im Mittel ≤ 3“:

$$A = \{\omega = (x_1, \cdots, x_n) :\ \frac{1}{n}\sum_{i=1}^{n} x_i\ \leq 3\}$$

ad 4) ($\infty\times$Würfeln) „niemals die 6“:

$$A = \{\omega = (x_1, x_2, \cdots) :\ x_i \neq 6\ \forall\, i \in \mathbb{N}\}$$

ad 5) (Zufallszahl): „Zahl liegt zwischen $\frac{1}{4}$ und $\frac{1}{3}$ (ausschließlich)“

$$A =]\tfrac{1}{4}, \tfrac{1}{3}[$$

ad 6) (stoch. Prozeß) „Teilchen (Kurs,...) überschreitet den Wert 2“

$$A = \{\omega \in \mathcal{C}(\mathbb{R}_+) :\ \omega(t) > 2 \text{ für (mind.) ein } t \in \mathbb{R}_+\}$$

Sei $(\Omega, \mathcal{A})$ ein Meßraum (d.h. $\Omega \neq \emptyset$ beliebig, $\mathcal{A}$ $\sigma-$Algebra auf Ω).

Definition 1.7. *Eine Abbildung $P : \mathcal{A} \longrightarrow [0, \infty]$ heißt* **Maß** *(auf dem Meßraum $(\Omega, \mathcal{A})$), falls:*

(*i*) $P(\emptyset) = 0$

(*ii*) $P\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ *für alle $A_i \in \mathcal{A}$ mit $A_i \cap A_j = \emptyset$ ($i, j \in \mathbb{N}$, $i \neq j$)*

„σ-Additivität“

$(\Omega, \mathcal{A}, P)$ heißt dann **Maßraum***.*
P heißt **W-Maß** *oder* **Verteilung***, falls zusätzlich*

(*iii*) $P(\Omega) = 1$

$(\Omega, \mathcal{A}, P)$ heißt dann **Wahrscheinlichkeitsraum***.*
Für ein Ereignis $A \in \mathcal{A}$ bezeichnet $P(A)$ die „Wahrscheinlichkeit, daß A eintritt“.

Bemerkungen:

a) Für ein W-Maß P gilt: $P(A^c) = 1 - P(A) \; \forall\, A \in \mathcal{A}$
(Allgemein: Falls $P(\Omega) < \infty$, dann folgt
$P(\emptyset) = 0)$ und $P(A^c) = P(\Omega) - P(A)$
bereits aus (ii) und (iii))

b) $\forall\, A, B \in \mathcal{A}$:
$\boxed{A \subset B \quad \Rightarrow \quad P(A) \leq P(B)}$ „Monotonie"
(denn $P(B) = P(A) + P(B \backslash A)$)

c) $\forall\, A, B \in \mathcal{A}: \quad P(A \cup B) + P(A \cap B) = P(A) + P(B)$
Mit Induktion nach n folgt:

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{1 \leq i \leq n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) \\ &\quad + \sum_{1 \leq i<j<k \leq n} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right) \end{aligned}$$

d) Insbesondere folgt $\forall A, B \in \mathcal{A}$
$\boxed{P(A \cup B) \leq P(A) + P(B)}$
und $\forall\, A_i \in \mathcal{A}$ $(i = 1, \cdots, n)$:

$$\begin{array}{lllll} (i) & P\left(\bigcup\limits_{i=1}^{n} A_i\right) & \leq & \sum\limits_{1 \leq i \leq n} P(A_i) & \\ (ii) & P\left(\bigcup\limits_{i=1}^{n} A_i\right) & \geq & \sum\limits_{1 \leq i \leq n} P(A_i) - \sum\limits_{1 \leq i<j \leq n} P(A_i \cap A_j) & \\ (iii) & P\left(\bigcup\limits_{i=1}^{n} A_i\right) & \leq & \sum\limits_{1 \leq i \leq n} P(A_i) - \sum\limits_{1 \leq i<j \leq n} P(A_i \cap A_j) + \sum\limits_{1 \leq i<j<k \leq n} P(A_i \cap A_j \cap A_k) & \\ (iv) & \quad \cdots & & & \end{array}$$

(*„Bonferroni-Ungleichungen"*)

Lemma 1.8. *Sei* $(\Omega, \mathcal{A})$ *ein Meßraum und* $P : \mathcal{A} \to [0, \infty]$ *eine Abbildung mit* $P(\emptyset) = 0$ *und*

$$P(A \cup B) = P(A) + P(B) \qquad \text{„\underline{(endliche) Additivität}“}$$

für alle $A, B \in \mathcal{A}$ *mit* $A \cap B = \emptyset$.

Dann sind äquivalent:

(i) P *ist ein Maß auf* $(\Omega, \mathcal{A})$.
(ii) $\forall\, A_i \in \mathcal{A}$ *mit* $A_i \subset A_{i+1}$ $(i \in \mathbb{N})$:

$$P\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \lim_{i \to \infty} P(A_i) \qquad \text{„\underline{Stetigkeit von unten}“}$$

Ist $P(\Omega) < \infty$, so sind dazu ferner äquivalent:

(iii) $\forall\, A_i \in \mathcal{A}$ *mit* $A_i \supset A_{i+1}$ $(i \in \mathbb{N})$:
$$P\left(\bigcap_{i\in\mathbb{N}} A_i\right) = \lim_{i\to\infty} P(A_i) \qquad \text{„\underline{Stetigkeit von oben}“}$$

(iv) $\forall\, A_i \in \mathcal{A}$ *mit* $A_i \supset A_{i+1}$ $(i \in \mathbb{N})$ *und* $\bigcap A_i = \emptyset$:
$$\lim_{i\to\infty} P(A_i) = 0 \qquad \text{„\underline{$\emptyset$ – Stetigkeit}“}$$

Bemerkung: Die Bedingung $P(\Omega) < \infty$ ist i.a. nicht überflüssig.
Beispiel: $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$, $P = \lambda =$ Lebesgue-Maß auf $\mathbb{R}$.

$$A_i =]i, \infty[\quad \Rightarrow \quad \bigcap_i A_i = \emptyset,$$

aber

$$P(A_i) = \infty \quad (\forall i) \quad \Rightarrow \quad \lim_{i\to\infty} P(A_i) = \infty.$$

Beweis des Lemmas:
$(i) \Rightarrow ii)$: Sei $B_1 = A_1$ und $B_{i+1} = A_{i+1} \setminus A_i$ $(i \in \mathbb{N})$

$$\Rightarrow \quad A_i = \bigcup_{j=1}^{i} B_j, \quad B_i \cap B_j = \emptyset \quad (i \neq j)$$

$$P\left(\bigcup A_i\right) = P\left(\bigcup B_j\right) \overset{\sigma-\text{Add.}}{=} \sum_j P(B_j) =$$

$$= \lim_{i\to\infty} \sum_{j=1}^{i} P(B_j) \overset{\text{Add.}}{=} \lim_{i\to\infty} P\left(\bigcup_{j=1}^{i} B_j\right) = \lim_{i\to\infty} P(A_i)$$

$(ii) \Rightarrow (i)$: Gegeben sind disjunkte B_j. Definiere $A_i = \bigcup\limits_{j=1}^{i} B_j$.
Dann gilt analog
$\sum P(B_j) = \cdots\cdots = \lim\limits_{i\to\infty} P(A_i) = P\left(\bigcup A_i\right) = P\left(\bigcup B_i\right)$.

$(ii) \Leftrightarrow (iii)$: wegen $P\left(\bigcap A_i\right) = P(\Omega) - P\left(\bigcup A_i^c\right)$

$(iii) \Rightarrow (iv)$: trivial, $(iv) \Rightarrow (iii)$: mit $B_i = A_i \setminus \bigcap\limits_{j=1}^{i-1} A_j$. $\square$

Korollar 1.9. $\forall A_i \in \mathcal{A}$ $(i \in \mathbb{N})$ *gilt:*

$$P\left(\bigcup_{i\in\mathbb{N}} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i) \qquad \text{„\textit{Subadditivität}“}$$

Beweis: c) und Stetigkeit von unten. □

Beispiel 1.10. „Ziege oder Auto“

- Sie sind Mitspieler in einer Fernsehshow. Der Showmaster zeigt Ihnen 3 verschlossene Türen. Hinter einer ist ein Auto, hinter den beiden anderen eine Ziege.
  Sie dürfen sich für eine der 3 Türen entscheiden und, wenn sie geöffnet wird, gewinnen Sie, was dahinter ist.
- Bevor die Tür Ihrer Wahl allerdings geöffnet wird, öffnet der Showmaster eine der beiden anderen Türen und zeigt Ihnen die dahinter stehende Ziege.
- Nun bietet Ihnen der Showmaster an, Ihre Entscheidung zu revidieren und die andere der beiden ungeöffneten Türen zu wählen.
- Ist es für Sie von Vorteil zu wechseln?

Beispiele von W-Räumen

Beispiele 1.11.

ad 1) (Münzwurf): $\Omega = \{K, Z\}$, $\mathcal{A} = \mathcal{P}(\Omega) = \{\emptyset, \{K\}, \{Z\}, \{K, Z\}\}$
Ideale („faire“) Münze: $P(\{K\}) = P(\{Z\}) = \frac{1}{2}, P(\emptyset) = 0, P(\Omega) = 1.$
ad 2) (Würfeln): $\Omega = \{1, \cdots, 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$
Idealer Würfel: $P(\{1\}) = P(\{2\}) = \cdots = P(\{6\}) = \frac{1}{6}$
$\Rightarrow P(A) = \frac{1}{6} \cdot |A|$
Hinweis: Es wäre nicht überraschend, wenn gilt:
$P(\{1\}) < P(\{2\}) < \cdots < P(\{6\})$,
denn die Seite mit 6 Augen enthält 6 Löcher und ist damit leichter als die gegenüberliegende Seite mit einem Loch.
ad 5) (Zufallszahl):
$\Omega = [0, 1], \quad \mathcal{A} = \mathcal{B}([0, 1]) = \sigma\left(\{]a, b[: \ 0 \leq a < b \leq 1\}\right),$
$P = \lambda|_{[0,1]}$ = Lebesgue-Maß (auf $\mathbb{R}$) eingeschränkt auf $[0, 1]$

= „Gleichverteilung auf $[0,1]$“
Wichtig: Die Gleichverteilung kann nicht mit $\mathcal{A} = \mathcal{P}([0,1])$ modelliert werden.
Hierzu braucht man eine echt kleinere $\sigma-$Algebra.

Exotisches

Beispiel 1.12. Ω überabzählbar, $\mathcal{A} = \{A \subset \Omega : A \text{ oder } A^c \text{ abzählbar}\}$

$$P(A) = \begin{cases} 0, & A \quad \text{abzählbar} \\ 1, & A^c \quad \text{abzählbar} \end{cases}$$

Weiteres wichtiges

Beispiel 1.13.
$\Omega \neq \emptyset$ beliebig, $\mathcal{A}$ beliebige $\sigma-$Algebra auf Ω (z. B. $\mathcal{P}(\Omega)$), $\omega \in \Omega$ beliebig
$\Rightarrow P = \delta_\omega$ =Dirac-Maß in ω, ist W-Maß auf $(\Omega, \mathcal{A})$ definiert durch

$$P(A) = \delta_\omega(A) = 1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \text{sonst} \end{cases}$$

Offenbar ist $P(\Omega) = 1$ und

$$P\left(\bigcup_{i\in\mathbb{N}} A_i\right) = 1_{\bigcup\limits_{i\in\mathbb{N}} A_i}(\omega) = \sum_{i\in\mathbb{N}} 1_{A_i}(\omega) = \sum_{i\in\mathbb{N}} P(A_i)$$

(falls $A_i \cap A_j = \emptyset \ (\forall i \neq j)$)bei

<u>Allgemeiner:</u> $\omega_k \in \Omega$ und $p_k \in [0,1]$ $(k \in \mathbb{N})$ mit

$$\sum_{k=1}^{\infty} p_k = 1$$

$\Rightarrow P = \sum\limits_{k=1}^{\infty} p_k \cdot \delta_{\omega_k}$ ist W-Maß auf $(\Omega, \mathcal{A})$
<u>Beweis:</u> $P(\Omega) = \sum\limits_k p_k = 1$ und

$$\begin{aligned} \mathrm{P}\left(\bigcup_i A_i\right) &= \sum_k p_k \cdot \delta_{\omega_k}\left(\bigcup_i A_i\right) = \sum_k p_k \cdot \left(\sum_i \delta_{\omega_k}(A_i)\right) \\ &= \sum_i \left(\sum_k p_k \cdot \delta_{\omega_k}(A_i)\right) = \sum_i P(A_i) \end{aligned}$$

Definition 1.14. *Ein W-Raum* $(\Omega, \mathcal{A}, P)$ *heißt* **diskret**, *falls* Ω *abzählbar ist und* $\mathcal{A} = \mathcal{P}(\Omega)$.

Sei $p_\omega = P(\{\omega\}) \Rightarrow P = \sum_{\omega\in\Omega} p_\omega \cdot \delta_\omega$, d.h. $P(A) = \sum_{\omega\in A} p_\omega \qquad \forall\, A \subset \Omega$.

Definition 1.15. *Ein W-Raum* $(\Omega, \mathcal{A}, P)$ *heißt* **Laplace-Raum**, *falls* Ω *endlich ist,* $\mathcal{A} = \mathcal{P}(\Omega)$ *und*

$$P(A) = \frac{|A|}{|\Omega|}$$

d.h. $P(A)$ ist das Verhältnis der „Anzahl der für A günstigen Fälle“ zur „Anzahl der möglichen Fälle“.

Berechnung von $P(A)$ erfolgt mittels Kombinatorik.

Beispiel 1.16. „Ziehen mit Zurücklegen“
n Kugeln (s schwarze, $n-s$ weiße) in Urne,
m werden gezogen <u>mit</u> Zurücklegen.
Mit welcher Wahrscheinlichkeit sind darunter k schwarze?

$$\Omega = \{(a_1, \cdots, a_m) :\ a_i \in \Omega_0\} = \Omega_0^m$$

mit $\Omega_0 = \{1, \ldots, n\}$, wobei man sich die Kugeln mit $1, \cdots, n$ nummeriert denkt und a_i den $i-$ten Zug beschreibt, d.h.

$$|\Omega| = n^m$$

$A_i =$„beim i-ten Zug wird eine schwarze Kugel gezogen“

$$\Rightarrow P(A_i) = \frac{s}{n}, \quad P(A_i^c) = \frac{n-s}{n}$$

$A_i \cap A_j =$„beim i-ten und j-ten Zug schwarz“ $(i \neq j)$

$$\Rightarrow P(A_i \cap A_j) = \left(\frac{s}{n}\right)^2,$$

$A_i \cap A_j^c =$„beim i-ten Zug schwarz, beim j-ten Zug weiß“

$$\Rightarrow P(A_i \cap A_j^c) = \frac{s}{n} \cdot \frac{n-s}{n},$$

$B_{i_1,\cdots,i_k} =$„beim i_1-ten, i_2-ten, … und i_k-ten Zug schwarz, sonst weiß“
$= \bigcup_{l\in\{i_1,\ldots,i_k\}} A_l \cap \bigcap_{m\notin\{i_1,\ldots,i_k\}} A_m^c$

$$\begin{aligned}\Rightarrow P(B_{i_1,\cdots,i_k}) &= \prod_{l\in\{i_1,\ldots,i_k\}} P(A_l) \cdot \prod_{m\notin\{i_1,\ldots,i_k\}} P(A_m^c) \\ &= \left(\frac{s}{n}\right)^k \cdot \left(\frac{n-s}{n}\right)^{m-k}\end{aligned}$$

$B_{[k]}$ =„genau k schwarze Kugeln werden gezogen“
$= \bigcup\limits_{\{i_1,\ldots,i_k\}\subset\{1,\ldots,m\}} B_{i_1,\ldots,i_k}$

$$\begin{aligned}
\Rightarrow P(B_{[k]}) &= \sum_{\{i_1,\ldots,i_k\}\subset\{1,\ldots,m\}} P(B_{i_1,\ldots,i_k}) \\
&= \binom{m}{k} \cdot \left(\frac{s}{n}\right)^k \cdot \left(\frac{n-s}{n}\right)^{m-k} \\
&= \binom{m}{k} \cdot p^k \cdot (1-p)^{m-k} \text{ mit } p = \frac{s}{n} \text{ und } \binom{m}{k} = \frac{m!}{k!(m-k)!} = \text{„}k \text{ aus } m\text{“}
\end{aligned}$$

Dies nennt man die **Binomialverteilung**

.

Beispiel 1.17. „Ziehen ohne Zurücklegen“
n Kugeln, s schwarze,
m gezogen <u>ohne</u> Zurücklegen, $m \leq \inf\{s, n-s\}$.
$\Omega = \{\{a_1, \ldots, a_m\} : a_i \in \Omega_0$, alle a_i verschieden $\}$
$\Rightarrow |\Omega| = \binom{n}{m}$,
A_k=„genau k schwarze Kugeln werden gezogen“,
$|A_k| = \binom{s}{k}\binom{n-s}{m-k} \Rightarrow \mathrm{P}(A_k) = \frac{\binom{s}{k}\binom{n-s}{m-k}}{\binom{n}{m}}$, für $k = 0, 1, \ldots, m$.
Dies nennt man die **hypergeometrische Verteilung**.
Für $n \to \infty$ und $s \to \infty$ mit $p = \frac{s}{n}$ konstant, und m fest gilt

$$\mathrm{P}(A_k) \to \binom{m}{k} p^k (1-p)^{m-k},$$

d.h. die hypergeometrische Verteilung konvergiert in obigem Sinne gegen die Binomialverteilung. Es ist also für große n egal, ob mit oder ohne Zurücklegen gezogen wird.

Kapitel 2

Meßbare Abbildungen, Bildmaße, Verteilungsfunktionen

Seien $(\Omega, \mathcal{A})$ und $(\Omega', \mathcal{A}')$ Meßräume.

Definition 2.1. *Eine Abbildung $f : \Omega \to \Omega'$ heißt* **meßbar** *(genauer: $\mathcal{A}/\mathcal{A}'$ -* **meßbar)***, falls*

$$f^{-1}(A') \in \mathcal{A} \quad \text{für alle } A' \in \mathcal{A}'$$

Dabei ist $f^{-1}(A') = \{\omega \in \Omega : f(\omega) \in A'\}$.

Bemerkungen:

a) $f : \Omega_0 \to \Omega_1 \quad \mathcal{A}_0/\mathcal{A}_1$-meßbar und $g : \Omega_1 \to \Omega_2 \quad \mathcal{A}_1/\mathcal{A}_2$-meßbar
$\Rightarrow g \circ f : \Omega_0 \to \Omega_2$ ist $\mathcal{A}_0/\mathcal{A}_2$-meßbar

b) $\mathcal{A} = \mathcal{P}(\Omega) \Rightarrow$ alle f: $\Omega \to \Omega'$ sind meßbar.

c) Ist $\mathcal{A}' = \sigma(\mathcal{A}'_0)$ und $\mathrm{f}^{-1}(A') \in \mathcal{A}\,(\forall A' \in \mathcal{A}'_0)$, so ist f meßbar.
(Denn: $\{A' : f^{-1}(A') \in \mathcal{A}\}$ ist σ-Algebra (siehe Übung) auf Ω' und enthält $\mathcal{A}'_0$, also auch $\sigma(\mathcal{A}'_0)$)

d) Seien Ω, Ω' topologische Räume, $\mathcal{A}$, $\mathcal{A}'$ deren Borelsche σ-Algebren:
f: $\Omega \to \Omega'$ stetig $\Rightarrow$ f meßbar.

e) Sei Ω beliebig, $\neq \emptyset$, $(\Omega', \mathcal{A}')$ Meßraum, f: $\Omega \to \Omega'$
$\Rightarrow \sigma(\mathrm{f}) = \{f^{-1}(A') : A' \in \mathcal{A}'\}$ ist σ-Algebra auf Ω (Übung!)
(„von f erzeugte σ-Algebra“).
Es gilt:f ist $\sigma(\mathrm{f})/\mathcal{A}'$-meßbar!
$\sigma(\mathrm{f})$ ist die kleinste σ-Algebra $\mathcal{A}$ auf Ω, für die f $\mathcal{A}/\mathcal{A}'$-meßbar ist.

Satz 2.2. *Sei* $(\Omega, \mathcal{A}, P)$ *ein W-Raum,* $(\Omega', \mathcal{A}')$ *ein Meßraum und* $f : \Omega \to \Omega'$ *meßbar. Dann definiert*

$$P'(A') = P(f^{-1}(A')) \qquad (\forall A' \in \mathcal{A}')$$

ein W-Maß auf $(\Omega', \mathcal{A}')$.
Genannt wird es „das Bild von P unter f" oder „die Verteilung von f unter P".
In Zeichen: $P' = P \circ f^{-1} = f(P) = P_f = f_* P$

Beweis:

- $P'(\emptyset) = P(\mathrm{f}^{-1}(\emptyset)) = P(\emptyset) = 0$,
  $P'(\Omega') = P(\mathrm{f}^{-1}(\Omega')) = P(\Omega) = 1$
- $A'_i \in \mathcal{A}'$ $(i \in \mathbb{N})$ disjunkt $\Rightarrow A_i := \mathrm{f}^{-1}(A'_i) \in \mathcal{A}$ disjunkt,
  $P'(\bigcup A'_i) = P\left(f^{-1}(\bigcup A'_i)\right) = P\left(\bigcup \mathrm{f}^{-1}(A'_i)\right) = P\left(\bigcup A_i\right)$
  $= \sum P(A_i) = \sum P'(A'_i)$. □

Definition 2.3. *Sei P ein W-Maß auf* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. *Die Funktion* $F : \mathbb{R} \longrightarrow [0,1]$ *definiert durch*

$$F(x) = P(]-\infty, x]) \quad (*)$$

heißt **Verteilungsfunktion** *von* P.

Satz 2.4.

a) *Die Verteilungsfunktion F eines W-Maßes ist*

   1. *monoton* (d.h. $x \le y \Rightarrow F(x) \le F(y)$),
   2. *rechtsseitig stetig* (d.h. $F(x) = \lim_{\epsilon \searrow 0} F(x+\epsilon)$),
   3. *normiert* (d.h. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to +\infty} F(x) = 1$).

b) *Zu jeder solchen Funktion* $F : \mathbb{R} \longrightarrow [0,1]$ *gibt es ein (sogar genau ein) W-Maß P auf* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *mit* $(*)$.

Beweis:

a) Einfach! Monotonie ist offensichtlich.
   Rechtsstetigkeit

$$\begin{aligned} F(x) = P(]-\infty, x]) = P(\bigcap_n]-\infty, x+\frac{1}{n}]) &= \lim_n P(]-\infty, x+\frac{1}{n}]) \\ &= \lim_n F(x+\frac{1}{n}) \end{aligned}$$

Normierung
$\lim\limits_{n\to\infty} F(n) = \lim\limits_{n\to\infty} P(]-\infty, n]) = P(]-\infty, +\infty[) = 1$
und $\lim\limits_{n\to\infty} F(n) = \ldots = P(\emptyset) = 0$

b) Sei $\lambda = \lambda|_{[0,1]}$ das Lebesgue-Maß auf $([0,1], \mathcal{B}([0,1]))$ und $f : [0,1] \longrightarrow \mathbb{R}$ mit

$$f(t) := \sup\{x \in \mathbb{R} : F(x) < t\}$$

(Grob gesagt: f ist Inverse von F.)
Genauer:

$$f^{-1}(]-\infty, x]) = \{t \in [0,1] : f(t) \leq x\} = \{t \in [0,1] : F(x) \geq t\} = [0, F(x)] \; (**)$$

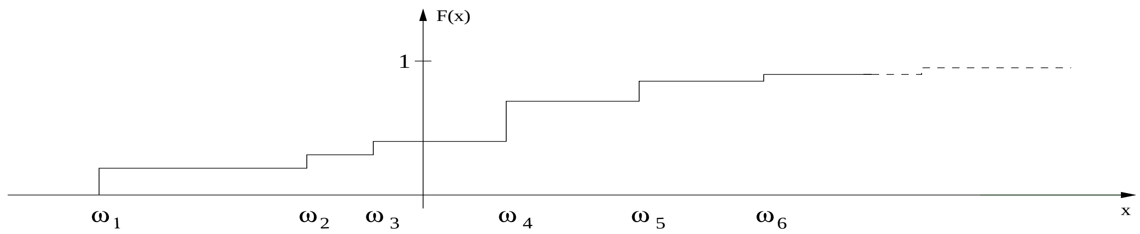
(Denn: aus $F(x) \geq t$ folgt $f(t) \leq x$, wegen $x \notin \{y : F(y) < t\}$;
aus $F(x) < t$ folgt, wegen Rechtsstetigkeit von F:
$F(x + \epsilon) < t$ und $f(t) \geq x + \epsilon > x$.)
Wegen $(**)$: f ist meßbar!
Betrachte $P = f(\lambda) = \ldots = \lambda \circ f^{-1}$ das Bild von λ unter f. P ist W-Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit

$$P(]-\infty, x]) = \lambda(f^{-1}(]-\infty, x])) = \lambda([0, F(x)]) = F(x) \quad \square$$

Bemerkung: Das Maß P ist durch $(*)$ sogar eindeutig festgelegt! *(Später!)*

Bemerkungen:
1. Jede Verteilungsfunktion F hat höchstens abzählbar viele Unstetigkeitsstellen. (da es maximal k Stellen mit $F(x) - F(x_-) \geq \frac{1}{k}$ gibt.($\forall k \in \mathbb{N}$)).
2. Ist $\Omega \subset \mathbb{R}$ und $(\Omega, \mathcal{A}, \mathcal{P})$ diskret, so ist

$$F(x) = \sum_{\substack{\omega \in \Omega \\ \omega \leq x}} P(\{\omega\})$$



Beispiele 2.5.

a.) Binomialverteilung: $F(x) = \sum_{\substack{0 \le k \le m \\ k \le x}} \binom{m}{k} \cdot p^k \cdot (1-p)^{m-k}$

b.) Hypergeometrische Verteilung: $F(x) = \frac{1}{\binom{n}{m}} \cdot \sum_{\substack{0 \le k \le m \\ k \le x}} \binom{s}{k} \cdot \binom{n-s}{m-k}$

c.) Geometrische Verteilung: $F(x) = p \cdot \sum_{1 \le k \le x} (1-p)^{k-1}$

d.) Poisson-Verteilung: $F(x) = e^{-\lambda} \cdot \sum_{0 \le k \le x} \frac{\lambda^k}{k!}$

3. Stets gilt: $P(\{x\}) = F(x) - F(x_-)$.
   Folglich: F stetig in $x \Longleftrightarrow P(\{x\}) = 0$
4. F heißt **absolut-stetig**, falls eine Borel-meßbare Funktion $f : \mathbb{R} \longrightarrow [0, \infty]$ existiert mit

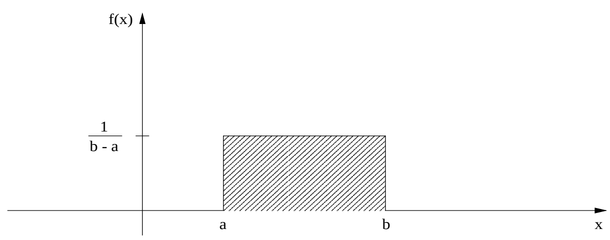
$$F(x) = \int_{-\infty}^{x} f(t)\, dt \quad (\forall x \in \mathbb{R})$$

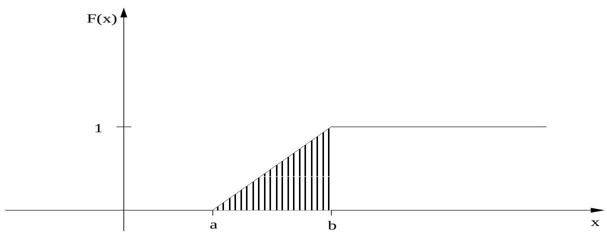
   Jede absolut-stetige Funktion ist stetig, aber nicht umgekehrt
   (siehe *Gleichverteilung auf Cantor-Menge*).
   f heißt **Dichte** von F (bzw. von P), falls F Verteilungsfunktion von P.

Beispiele 2.6. *:*

a.) Gleichverteilung auf $[a, b]$

$$f(x) = \begin{cases} \frac{1}{b-a} & , \quad x \in [a,b] \\ 0 & , \quad \text{sonst} \end{cases}$$



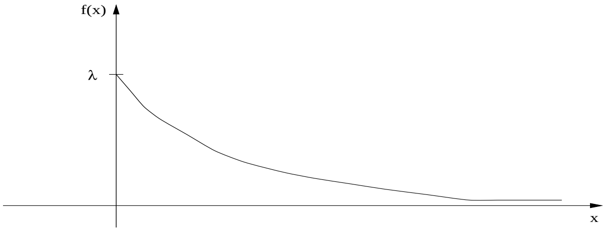


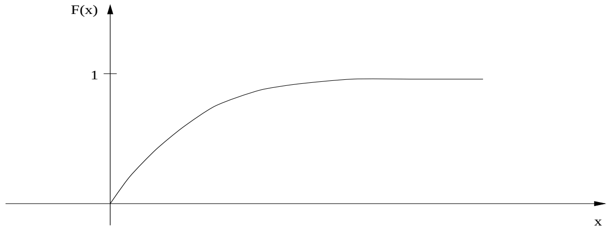
b.) Exponentialverteilung mit Parameter $\lambda > 0$

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & , \quad x \geq 0 \\ 0 & , \quad \text{sonst} \end{cases}$$

stetiges Analogon zur geometrischen Verteilung

$$(\text{Für } p_n = \tfrac{\lambda}{n} \text{ und } n \to \infty : \sum_{n(x-1)<x\leq nx} p_n(1-p_n)^{k_n-1} \longrightarrow \lambda \int_{x-1}^{x} e^{-\lambda \cdot t}\,dt)$$





c.) Gamma-Verteilung mit Parametern $\alpha, \lambda > 0$

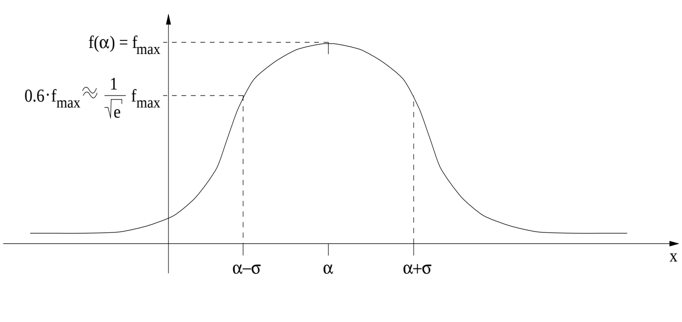
$$f(x) = \begin{cases} \lambda^{\alpha} \cdot x^{\alpha-1} e^{-\lambda x} \cdot \frac{1}{\Gamma(\alpha)} & , \quad x \geq 0 \\ 0 & , \quad x < 0 \end{cases}$$

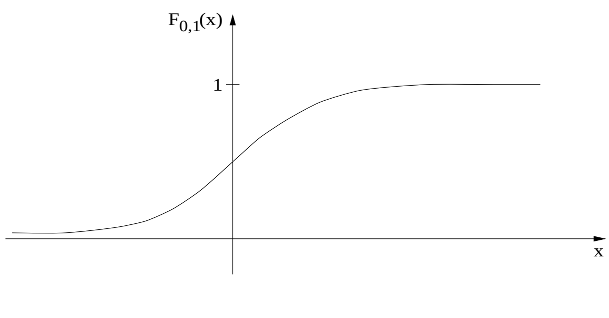
$$\text{mit} \quad \Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx.$$

Für $\alpha = 1$: Exponentialverteilung mit Parameter λ.

d.) Normalverteilung $N(\alpha, \sigma^2)$ mit Mittelwert $\alpha \in \mathbb{R}$ und Varianz $\sigma^2 > 0$ (bzw. Standard-Normalverteilung $N(0,1)$)

$$\begin{aligned} f_{\alpha,\sigma^2}(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{|x-\alpha|^2}{2\sigma^2}\right) = \frac{1}{\sigma} \cdot f_{0,1}\left(\frac{x-\alpha}{\sigma}\right) \\ F_{\alpha,\sigma^2}(x) &= \int_{-\infty}^{x} f_{\alpha,\sigma^2}(t)dt = F_{0,1}\left(\frac{x-\alpha}{\sigma}\right) \\ & \text{mit} \quad F_{0,1} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \end{aligned}$$





keine geschlossene Darstellung; Abschätzungen, Tabellen.

Lemma 2.7. *Für $x > 0$ gilt:*

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) \cdot e^{-\frac{x^2}{2}} \leq \int_x^\infty e^{-\frac{t^2}{2}} dt \leq \frac{1}{x} \cdot e^{-\frac{x^2}{2}}$$

Damit insbesondere für $x \to \infty$

$$\int_x^\infty e^{-\frac{t^2}{2}} dt \sim \frac{1}{x} \cdot e^{-\frac{x^2}{2}}$$

Beweis: Übung!

Zufallsvariablen

Gegeben ist ein W-Raum $(\Omega, \mathcal{A}, P)$.

Definition 2.8.

1. *Eine* **Zufallsvariable (ZV)** *auf Ω ist eine $\mathcal{A}/\mathcal{B}(\mathbb{R})$-meßbare Abb. $X : \Omega \longrightarrow \mathbb{R}$*
   (bzw. gelegentlich: eine $\mathcal{A}/\mathcal{B}(\overline{\mathbb{R}})$-meßbare Abb. $X : \Omega \longrightarrow \overline{\mathbb{R}} = [-\infty, +\infty]$.
   $\overline{\mathbb{R}}$ ist ein topologischer Raum.)
2. *Das W-Maß P_X auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ heißt* **Verteilung** *von X.*
   Es gilt: $P_X(A) = P(X \in A) = P(\{\omega : X(\omega) \in A\})$
3. *Die* **Verteilungsfunktion** *von X (genauer: Verteilungsfunktion von P_X) ist $F_X : x \longmapsto F_X(x) = P(X \leq x)$.*
4. *X definiert auf $(\Omega, \mathcal{A}, P)$ und X' definiert auf $(\Omega, \mathcal{A}, P)$ heißen* **gleich in Verteilung**, *falls $P_X = P_{X'}$ (bzw. $F_X = F_{X'}$)*
   Kurz: $X \stackrel{\mathrm{d}}{=} X'$.
   Nicht $(\Omega, \mathcal{A}, P)$ ist relevant bzw. "real", sondern $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$.

Bemerkungen:

1. Hinreichend für Meßbarkeit einer Abb. $X : \Omega \longrightarrow \mathbb{R}$ ist
$$\{X \leq \alpha\} \in \mathcal{A} \quad (\forall \alpha \in \mathbb{Q})$$
(bzw. $\{X < \alpha\} \in \mathcal{A}$ oder $\{X \geq \alpha\} \in \mathcal{A}$ oder $\{X > \alpha\} \in \mathcal{A}$).
<u>Denn</u> daraus folgt: $\{X < \beta\} = \bigcup_{\substack{\alpha \in \mathbb{Q} \\ \alpha < \beta}} \{X \leq \alpha\} \in \mathcal{A}$ für $\beta \in \mathbb{R}$
$$\Big(= \bigcup_{\substack{\alpha \in \mathbb{Q} \\ \alpha < \beta}} \{X < \alpha\} \in \mathcal{A}\Big)$$
Aber: $\sigma\{]-\infty, \beta[: \beta \in \mathbb{R}\} = \mathcal{B}(\mathbb{R})$.

2. Indikatorfunktion $\mathbb{1}_A : \omega \longmapsto \begin{cases} 0 & , \ \omega \in A^c \\ 1 & , \ \omega \in A \end{cases}$ eines Ereignisses $A \in \mathcal{A}$ ist ZV.
<u>Denn:</u> $\{\mathbb{1}_A \leq \alpha\} = \begin{cases} \emptyset & , \ \alpha < 0 \\ A^c & , \ 0 \leq \alpha < 1 \\ \Omega & , \ \alpha \geq 1 \end{cases}$

3. Ist Ω topologischer Raum und $\mathcal{A} = \mathcal{B}(\Omega)$, so ist jede nach unten (bzw. nach oben) halbstetige Funktion $X : \Omega \longrightarrow \mathbb{R}$ meßbar.
<u>Denn</u>: $\{X > \alpha\}$ ist offen! (bzw. $\{X < \alpha\}$ ist offen!)

4. Ist X ZV auf Ω und $f : \mathbb{R} \longrightarrow \mathbb{R}$ meßbar (z. B. stetig, halbstetig), so ist $f(X)$ ZV:
z. B. X^2, $|X|$, e^X, $X_+ = \sup\{X, 0\}$, $\alpha \cdot X$

Satz 2.9. *Seien $X_n : \Omega \longrightarrow \mathbb{R}$ ZV $(n \in \mathbb{N})$. Dann sind auch die folgenden Funktionen ZV:*

(i) $S_k = \sum_{n=1}^{k} X_n \quad (k \in \mathbb{N})$

(ii) $S_\infty = \sum_{n=1}^{\infty} X_n$, *falls konvergent für alle* ω

(iii) $\inf_{n \in \mathbb{N}} X_n, \quad \sup_{n \in \mathbb{N}} X_n$

(iv) $\liminf_{n \to \infty} X_n, \quad \limsup_{n \to \infty} X_n$

Beweis:
(i) Induktion nach k:
$$\{S_{k+1} < \alpha\} = \bigcup_{\substack{\beta, \gamma \in \mathbb{Q} \\ \beta + \gamma < \alpha}} \{S_k < \beta\} \cap \{X_{k+1} < \gamma\} \in \mathcal{A};$$

(iii) $\{\inf_{n\in\mathbb{N}} X_n < \alpha\} = \bigcup_{n\in\mathbb{N}} \{X_n < \alpha\} \in \mathcal{A}$;

(iv) wegen $\liminf_{n\to\infty} X_n = \sup_{n\in\mathbb{N}} \left(\inf_{k\geq n} X_k\right)$ und (iii);

(ii) wegen (iv). $\square$

Korollar 2.10. *Die Mengen*

$$\Omega_0 = \{\lim_{n\to\infty} X_n \textit{ existiert }\} = \{\limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n\}$$

und

$$\Omega_1 = \{S_\infty \textit{ existiert }\}$$

sind meßbar.

Definition 2.11. *X heißt* **elementare ZV**, *falls*

$$X = \sum_{k=1}^{n} \alpha_k \cdot \mathbb{1}_{A_k} \qquad \textit{mit } \alpha_k \in \mathbb{R}, A_k \in \mathcal{A}.$$

Satz 2.12 (Approximation von ZV durch elementare ZV). *Für jede ZV $X \geq 0$ existieren elementare ZV X_n mit $0 \leq X_1 \leq X_2 \leq \ldots$ und*

$$X = \lim_{n\to\infty} X_n.$$

Beweis:

$$X_n = \sum_{k=0}^{2^n\cdot n} k \cdot 2^{-n} \cdot \mathbb{1}_{\underbrace{\{k\cdot 2^{-n} \leq X < (k+1)\cdot 2^{-n}\}}_{=A_k\in\mathcal{A}}} \nearrow X. \qquad \square$$

Bemerkung: Jede ZV X ist Differenz $X_+ - X_-$ von ZV $X_\pm \geq 0$.

Kapitel 3

Existenz und Eindeutigkeit von Maßfortsetzungen

Problem 1: Geg. Maße μ_1, μ_2 (bzw. W-Maße P_1, P_2) auf Meßraum $(\Omega, \mathcal{A})$ mit $\mathcal{A} = \sigma(\mathcal{A}_0)$ und

$$\mu_1(A) = \mu_2(A) \qquad \forall A \in \mathcal{A}_0.$$

Frage: Folgt daraus bereits $\mu_1 = \mu_2$?

Antwort: nein!

Frage: Welche Bedingung muß $\mathcal{A}_0$ erfüllen?

Problem 2: Geg. eine Mengenfunktion $\mu_0 : \mathcal{A}_0 \longrightarrow [0, \infty]$.

Frage: Welche Bedingung muß $\mathcal{A}_0$ und/oder μ_0 erfüllen, damit: $\exists$ Maß μ auf $(\Omega, \mathcal{A})$ mit $\mu(A) = \mu_0(A)$ $\forall A \in \mathcal{A}_0$.

Sei $\Omega \neq \emptyset$.

Definition 3.1.

a) Ein Mengensystem $\mathcal{D} \subseteq \mathcal{P}(\Omega)$ heißt **Dynkin-System**, *falls*

- $\Omega \in \mathcal{D}$
- $A \in \mathcal{D} \Longrightarrow A^c \in \mathcal{D}$
- $A_i \in \mathcal{D}$ $(i \in \mathbb{N})$ *paarweise disjunkt* $\Longrightarrow \bigcup\limits_{i \in \mathbb{N}} A_i \in \mathcal{D}$.

b) Ein Mengensystem $\mathcal{A}_0 \subset \mathcal{P}(\Omega)$ heißt $\bigcap$-stabil, falls

- $A, B \in \mathcal{A}_0 \Longrightarrow A \cap B \in \mathcal{A}_0$

Beispiel 3.2. P_1, P_2 W-Maße auf $(\Omega, \mathcal{A}) \Rightarrow \mathcal{D} = \{A \in \mathcal{A} : P_1(A) = P_2(A)\}$ ist ein Dynkin-System

Beispiel 3.3. $\mathcal{A}_0 = \{]a, b[: a, b \in \mathbb{R}\}$ ist $\bigcap$-stabil.

Bemerkung: $A, B \in \mathcal{D}$ und $A \subseteq B \Rightarrow B \setminus A = (B^c \dot{\bigcup} A)^c \in \mathcal{D}$

Proposition 3.4. *Für $\mathcal{A} \subset \mathcal{P}(\Omega)$ sind äquivalent:*

(i) $\mathcal{A}$ ist σ-Algebra

(ii) $\mathcal{A}$ ist $\bigcap$- stabiles Dynkin-System.

Beweis:
(i)$\Rightarrow$(ii): klar;
(ii)$\Rightarrow$(i): Wegen $\bigcap$-Stabilität gilt: $A, B \in \mathcal{A} \Rightarrow A \bigcup B = (A^c \bigcap B^c)^c \in \mathcal{A}$.
Weiterhin: $A \setminus B \in \mathcal{A}$.
Für $A_i \in \mathcal{A}$ setze $B_i = \bigcup_{k \leq i} A_k \in \mathcal{A}$ und $B_0 = \emptyset$
$\Rightarrow \bigcup_{i \in \mathbb{N}} A_i = \bigcup_{i \in \mathbb{N}} B_i = \dot{\bigcup}_{i \in \mathbb{N}} \underbrace{(B_i \setminus B_{i-1})}_{\in \mathcal{A}} \in \mathcal{A}$. □

Proposition 3.5. *Ist $\mathcal{A}_0$ $\bigcap$-stabil, so ist*

$$\mathcal{D}(\mathcal{A}_0) = \sigma(\mathcal{A}_0).$$

Hierbei ist

$$\mathcal{D}(\mathcal{A}_0) = \bigcap_{\substack{\mathcal{D} \supset \mathcal{A}_0 \\ \mathcal{D}\,\text{Dynkin-System}}} \mathcal{D}$$

das von $\mathcal{A}_0$ erzeugte Dynkin-System und $\sigma(\mathcal{A}_0)$ die von $\mathcal{A}_0$ erzeugte σ-Algebra.

Beweis:

- $\mathcal{D}(\mathcal{A}_0) \subset \sigma(\mathcal{A}_0)$ klar (denn: $\bigcap_{Dynkin-System} \subset \bigcap_{\sigma-Algebra}$)

- $\mathcal{D}(\mathcal{A}_0) \supset \sigma(\mathcal{A}_0)$: zu zeigen (wegen Prop. 3.4): $\mathcal{D}(\mathcal{A}_0)$ ist $\bigcap$-stabil
  Seien $A, B \in \mathcal{D}(\mathcal{A}_0)$.
  Definiere $\mathcal{D}_A = \{B_0 \subset \Omega : A \bigcap B_0 \in \mathcal{D}(\mathcal{A}_0)\}$
  Offenbar gilt: $\mathcal{D}_A \supset \mathcal{A}_0$ und $\mathcal{D}_A$ Dynkin-System $\Rightarrow \mathcal{D}_A \supset \mathcal{D}(\mathcal{A}_0)$
  $\Rightarrow \mathcal{D}_A \ni B \Rightarrow A \bigcap B \in \mathcal{D}(\mathcal{A}_0)$ nach Def. von $\mathcal{D}_A$. $\square$

Satz 3.6 (Eindeutigkeitssatz). *μ_1, μ_2 seien Maße auf $(\Omega, \mathcal{A})$ und $\mathcal{A}_0$ sei ein $\bigcap$-stabiler Erzeuger von $\mathcal{A}$ mit: $\exists\, \Omega_n \in \mathcal{A}_0$ $(n \in \mathbb{N})$ mit $\Omega_n \nearrow \Omega$ und $\mu_i(\Omega_n) < \infty$. Gilt dann*

$$\mu_1(A) = \mu_2(A) \qquad\qquad (\forall A \in \mathcal{A}_0),$$

so ist $\mu_1 = \mu_2$.

Bemerkung: Für endliche Maße (insbes. W-Maße) μ_i kann man stets $\Omega_n = \Omega$ wählen.

Beweis des Satzes 3.6:
Für endliche Maße gilt:
$\mathcal{D}_0 = \{A \in \mathcal{A} : \mu_1(A) = \mu_2(A)\}$ ist Dynkin-System und
$\sigma(\mathcal{A}_0) = \mathcal{A} \supset \mathcal{D}_0 \supset \mathcal{A}_0 \overset{\text{Prop.3.5}}{\Longrightarrow} \sigma(\mathcal{A}_0) = \mathcal{D}(\mathcal{A}_0) \subset \mathcal{D}_0$
Allgemeiner Beweis: Betrachte $\mathcal{D}_n = \{A \in \mathcal{A} : \mu_1(A \bigcap \Omega_n) = \mu_2(A \bigcap \Omega_n)\} \Rightarrow \mathcal{D}_n = \mathcal{A}$
Also: $\forall A \in \mathcal{A} : \mu_1(A) = \lim_{n\to\infty} \mu_1(A \bigcap \Omega_n) = \lim_{n\to\infty} \mu_2(A \bigcap \Omega_n) = \mu_2(A)$. $\square$

Korollar 3.7. *Ein W-Maß P auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ist eindeutig festgelegt durch seine Verteilungsfunktion F.*

Beweis: $P(]a, b]) = P(]-\infty, b]) - P(]-\infty, a]) = F(b) - F(a)$ $(\forall a < b)$ und $\{]a, b] : a, b \in \mathbb{R}, a < b\}$ ist $\bigcap$-stabiler Erzeuger von $\mathcal{B}(\mathbb{R})$. $\square$

Definition 3.8.

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

- *$\emptyset \in \mathcal{A}$ (bzw. $\mathcal{A}$ nichtleer)*
- *$A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$*
- *$A, B \in \mathcal{A} \Rightarrow A \bigcup B \in \mathcal{A}$.*

b) Eine Abbildung $\mu : \mathcal{A} \longrightarrow [0,\infty]$ *heißt* **Prämaß** *auf* $\mathcal{A}$*, falls*

- $\mu(\emptyset) = 0$

- $A_i \in \mathcal{A}$ *disjunkt* $(i \in \mathbb{N})$ *<u>und</u>* $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{A}$
  $\Longrightarrow \mu(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} \mu(A_i)$.

c) μ *heißt* σ**-endlich**, *falls* $\exists \Omega_n \in \mathcal{A}\,(n \in \mathbb{N})$ *mit* $\Omega_n \nearrow \Omega$ *und* $\mu(\Omega_n) < \infty$.

Beispiel 3.9. $\{\bigcup_{i=1}^{k}]a_i, b_i] : k \in \mathbb{N}, a_i, b_i \in \overline{\mathbb{R}}, a_i < b_i\}$ ist Algebra auf $\mathbb{R}$.

Satz 3.10 (Fortsetzungssatz von Carathèodory). *Sei* μ_0 *ein* σ*-endliches Prämaß auf einer Algebra* $\mathcal{A}$.
Dann existiert genau ein Maß μ *auf* $\sigma(\mathcal{A})$ *mit*

$$\mu(A) = \mu_0(A) \qquad \forall A \in \mathcal{A}.$$

Beweis: Die Eindeutigkeit ist bereits gezeigt, da Algebra $\bigcap$-stabil.

Beweisskizze für die Existenz:
a) Für $E \subset \Omega$ setze

$$\mu^*(E) = \inf\{\sum_{i\in\mathbb{N}} \mu(A_i) : (A_i)_{i\in\mathbb{N}} \text{ Folge in } \mathcal{A} \text{ mit } \bigcup_{i\in\mathbb{N}} A_i \supset E\}.$$

Dann gilt für die Abbildung $\mu^* : \mathcal{P}(\Omega) \longrightarrow [0,\infty]$

- $\mu^*(\emptyset) = 0$

- $\mu^*(E_1) \leq \mu^*(E_2)$ falls $E_1 \subseteq E_2$

- $\mu^*(\bigcup_{n\in\mathbb{N}} E_n) \leq \sum_{n\in\mathbb{N}} \mu^*(E_n)$

μ^* heißt **äußeres Maß**.
Eine Menge $F \subset \Omega$ heißt μ^***-meßbar**, falls

$$\mu^*(E) = \mu^*(E \cap F) + \mu^*(E \cap F^c) \qquad (\forall E \subset \Omega).$$

(Hinweis: Stets gilt " $\leq$ ".) $\boxed{\text{a)}}$

b) *Jedes* $A \in \mathcal{A}$ *ist* μ^**-meßbar und* $\mu^*(A) = \mu(A)$.

Klar ist $\mu^*(A) = \mu(A)$.

Zur μ^*-Meßbarkeit: o.B.d.A. $\mu^*(E) < \infty$.

$\Rightarrow \exists\, (A_i)_i$ in $\mathcal{A}$ mit $E \subset \bigcup A_i$ und $\sum \mu(A_i) \leq \mu^*(E) + \epsilon$

$\Rightarrow \mu(A_i) = \mu^*(A_i \cap A) + \mu^*(A_i \cap A^c)$

$\Rightarrow \mu^*(E) + \epsilon \geq \sum \mu^*(A_i \cap A) + \sum \mu^*(A_i \cap A^c) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c)$ $\boxed{\text{b)}}$

c) Die Menge $\mathcal{A}^*$ der μ^*-meßbaren Mengen ist eine σ-Algebra und die Einschränkung von μ^* auf $\mathcal{A}^*$ ist ein Maß.

Beweis:

α) $E \in \mathcal{A}^* \Rightarrow E^c \in \mathcal{A}^*$ trivial

β) $E_1, E_2 \in \mathcal{A}^* \Rightarrow E_1 \bigcup E_2 \in \mathcal{A}^*$

denn:

$$\begin{aligned}
&\mu^*(G \cap (E_1 \cup E_2)) + \mu^*(G \cap (E_1^c \cup E_2^c)) \\
&\leq \mu^*(G \cap E_1) + \mu^*(G \cap E_1^c \cap E_2) + \mu^*(G \cap E_1^c \cap E_2^c) \\
&= \mu^*(G \cap E_1) + \mu^*(G \cap E_1^c) = \mu^*(G) \qquad\qquad (\forall G \subset \Omega)
\end{aligned}$$

$\nwarrow$ μ^*-Meßbarkeit von E_2 $\qquad$ $\uparrow$ μ^*-Meßbarkeit von E_1

γ) $E_i \in \mathcal{A}^*\, (i \in \mathbb{N}) \Rightarrow \bigcup\limits_{i \in \mathbb{N}} E_i \in \mathcal{A}^*$ (nicht trivial, siehe [Bauer] ,[Durrrett])

δ) $E_i \in \mathcal{A}^*$ disjunkt $\Rightarrow \mu^*(\bigcup\limits_{i \in \mathbb{N}} E_i) = \sum\limits_{i \in \mathbb{N}} \mu^*(E_i)$

denn sei $F_k = \bigcup\limits_{i \leq k} E_i$

$$\begin{aligned}
\mu^*(F_k) &= \mu^*(F_k \cap E_k) + \mu^*(F_k \cap E_k^c) \\
&= \mu^*(E_k) + \mu^*(F_{k-1})
\end{aligned}$$

$\Rightarrow$ (mit Induktion über k): $\mu^*(\bigcup\limits_{i \leq k} E_i) = \sum\limits_{i \leq k} \mu^*(E_i)$

$\Rightarrow$ (mit Monotonie von μ^*): $\mu^*(\bigcup\limits_{i \in \mathbb{N}} E_i) \geq \sum\limits_{i \in \mathbb{N}} \mu^*(E_i)$ $\boxed{\text{c)}}$

d) $\mathcal{A}^* \supset \sigma(\mathcal{A})$. Das gesuchte μ ist die Einschränkung von μ^* auf $\sigma(\mathcal{A})$ $\boxed{\text{d)}}$.

Korollar 3.11. *Auf* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *existiert genau ein Maß* λ *mit* $\lambda(]a, b]) = b - a\, (\forall a < b)$.

Beweis: $\{]a, b] : a < b\}$ ist $\bigcap$-stabiler Erzeuger von $\mathcal{D}(\overline{\mathbb{R}})$. $\Rightarrow$ Eindeutigkeit; und $\{\bigcup\limits_{i=1}^{k}]a_i, b_i] : k \in \mathbb{N}, a_i < b_i\}$ ist Algebra. $\Rightarrow$ Existenz. $\square$

Anwendung: Produktmaße

Seien $(\Omega_i, \mathcal{A}_i, \mu_i)$ Maßräume $(i = 1, \ldots, k)$,

$\Omega = \Omega_1 \times \ldots \times \Omega_k$ "Produktraum"

$\mathcal{A} = \sigma(\{A_1 \times \ldots \times A_k : A_i \in \mathcal{A}_i\})$

$\quad = \mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_k$ "Produkt-σ-Algebra"

Satz 3.12. *Sind alle μ_i σ-endlich, so existiert genau ein Maß μ auf $(\Omega, \mathcal{A})$ mit*

$$\mu(A_1 \times \ldots \times A_k) = \prod_{i=1}^{k} \mu_i(A_i) \qquad (\forall A_i \in \mathcal{A}_i)$$

μ heißt "Produkt-Maß".
In Zeichen: $\mu = \mu_1 \otimes \ldots \otimes \mu_k$.

Beweis: Sei $\mathcal{A}_0 = \{A_1 \times \ldots \times A_k : A_i \in \mathcal{A}_i\}$
und $\mathcal{A}_1 = \left\{ \bigcup_{i=1}^{N} B_i : N \in \mathbb{N}, B_i \in \mathcal{A}_0 \right\}$.
Dann gilt:

- $\mathcal{A}_0$ ist $\bigcap$-stabil $\Rightarrow$ Eindeutigkeit von μ

- $\mathcal{A}_1$ ist Algebra,

$$\forall A \in \mathcal{A}_1 : A = \dot{\bigcup_{j=1}^{N}} B_j, \quad B_j \in \mathcal{A}_0$$

$$\text{setze } \mu(A) = \sum_{j=1}^{N} \mu(B_j) \text{ mit } \mu(B_j) = \prod_{i=1}^{k} \mu_i(A_i)$$

$$\text{falls } B_j = A_1 \times \ldots \times A_k \in \mathcal{A}_0.$$

- Beachte: Die Definition von μ auf $\mathcal{A}_1$ ist konsistent (insbesondere unabhängig von $(B_j)_{j \in J}$). μ ist Prämaß auf $\mathcal{A}_1$
  $\Rightarrow$ μ besitzt eine eindeutige Fortsetzung auf $\sigma(\mathcal{A}_1)$.

- $\sigma(\mathcal{A}_1) = \mathcal{A}$. $\square$

Beispiel 3.13. λ^d das Lebesgue-Maß auf $\mathbb{R}^d$ ist eindeutig definiert durch

$$\lambda^d(]a_1, b_1] \times \ldots \times]a_d, b_d]) = \prod_{i=1}^{d} (b_i - a_i)$$

$\Rightarrow$ $\exists$ genau ein translationsinvariantes Maß λ^d auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ mit $\lambda^d(]0, 1]^d) = 1$.

Unendliche Produkte von W-Räumen

Seien $(\Omega_i, \mathcal{A}_i, P_i)$ W-Räume $\forall i \in I, I \neq \emptyset$ beliebige Indexmenge.

Beispiel 3.14. $I = \mathbb{N}$, $(\Omega_i, \mathcal{A}_i, P_i)$ beschreibt i-tes Zufallsspiel,

$A_i \in \mathcal{A}_i$	„im i-ten Spiel tritt Ereignis A_i ein“
$A_1 \times A_2 \times \times \ldots \times A_k \times \Omega_{k+1} \times \Omega_{k+2} \times \ldots$	„im 1-ten Spiel tritt A_1 ein <u>und</u> im 2-ten Spiel A_2 <u>und</u> ... <u>und</u> im k-ten Spiel A_k“

Intiutiv klar

$$P(A_1 \times A_2 \times \ldots \times A_k \times \Omega_{k+1} \times \ldots) = P_1(A_1) \cdot P_2(A_2) \cdot \ldots \cdot P_k(A_k)$$

Dies definiert tatsächlich ein W-Maß.

<u>Genauer</u>:Für $J \subset I$ sei $\Omega_J = \prod_{i \in J} \Omega_i$, $\Omega = \Omega_I$
und $\pi_J : \Omega \longrightarrow \Omega_J$, $\omega = (\omega_i)_{i \in I} \mapsto (\omega_i)_{i \in J}$.
Insbesondere $\pi_i = \pi(\{i\}) : \Omega \longrightarrow \Omega_i$, $\omega \mapsto \omega_i$ “Projektionen“

$\mathcal{A} = \bigotimes_{i \in I} \mathcal{A}_i = \sigma(\pi_i : i \in I)$ bezeichne die kleinste σ-Algebra auf Ω, bzgl. derer alle Projektionen π_i meßbar sind.
“Produkt-σ-Algebra“

Bemerkung: Dann sind auch alle π_J mit <u>endlichem</u> $J \subset I$ meßbar!

Satz 3.15. *Auf* $(\Omega, \mathcal{A})$ *existiert genau ein W-Maß*

$$P = \bigotimes_{i \in I} P_i \textit{ mit } \pi_J(P) = P_J \ \forall \textit{ endlichen } J \subset I, J \neq \emptyset$$

wobei $P_J = \bigotimes_{i \in J} P_i$ *das endliche Produkt der* P_i, $i \in J$ *bezeichnet.*

Beweisskizze: :
Für endliche J sei $\mathcal{A}_J = \bigotimes_{i \in J} \mathcal{A}_i$ und $\mathcal{Z}_J = \pi_J^{-1}(\mathcal{A}_J)$.

(<u>Beispiel:</u> $I = \mathbb{N} = \{1, 2, 3, \ldots\}$, $J = \{2, 4, 5\}$
$\Longrightarrow Z \in \mathcal{Z}_J \Leftrightarrow Z = \{\omega \in \Omega : (\omega_2, \omega_4, \omega_5) \in A\}$ mit $A \in \mathcal{A}_2 \otimes \mathcal{A}_4 \otimes \mathcal{A}_5$)

$$\mathcal{Z} = \bigcup_{\substack{J \subset I \\ J \text{ endl.}}} \mathcal{Z}_J \qquad = \text{" Menge der Zylindermegen in } \Omega \text{ "};$$

$\mathcal{Z}$ ist Algebra auf Ω und erzeugt $\mathcal{A}$. (nachrechnen!)
Für $Z = \pi_J^{-1}(A) \in \mathcal{Z}$ definiere $P(Z) = P(\pi_J^{-1}(A)) = \pi_J(P)(A) = P_J(A)$.
Diese Definition ist konsistent (unabhängig von J). P ist ein Prämaß auf $\mathcal{Z}$. (nachrechnen!)
$\Rightarrow$ P läßt sich eindeutig auf $\mathcal{A}$ fortsetzen.
Ferner: $P(\Omega) = P_J(\Omega_J) = 1$, d.h. P ist W-Maß. $\square$

Beispiel 3.16. Sei $(\Omega_i, \mathcal{A}_i, P_i) = (\Omega_0, \mathcal{A}_0, P_0)$ $\forall i \in I$

$$\Rightarrow \Omega = \prod_{i \in I} \Omega_0 = \Omega_0^I = \{f : I \longrightarrow \Omega_0 \text{ beliebige Abbildung }\}$$

Sei Ω_0 ein topologischer Raum und $\mathcal{A}_0 = \mathcal{B}(\Omega_0)$.
Dann ist auch $\Omega = \Omega_0^I$ ein topol. Raum (versehen mit der kleinsten Topologie, bzgl. derer alle Projektionen $\pi_i : \Omega \longrightarrow \Omega_0$, $i \in I$, stetig sind).

Beachte

a) Ist I abzählbar, so ist $\mathcal{B}(\Omega_0)^I = \mathcal{B}(\Omega_0^I)$.

b) Ist I überabzählbar und $|\Omega_0| \geq 2$, so ist $\mathcal{B}(\Omega_0)^I \subsetneqq \mathcal{B}(\Omega_0^I)$.
z. B. liegt $A = \{f : I \longrightarrow \Omega_0,\ f \text{ konstant auf } I\}$ nicht in $\mathcal{B}(\Omega_0)^I$.

Beispiele 3.17. ∞-oft Würfeln
$(\Omega_i, \mathcal{A}_i, P_i) = (\Omega_1, \mathcal{A}_1, P_1)$ $\forall i \in \mathbb{N}$ mit $\Omega_1 = \{1, \ldots, 6\}, \mathcal{A}_1 = \mathcal{P}(\Omega_1)$,
Sei $p = P_1(\{6\}) > 0$.
• Sei $A =$ "niemals die 6" $= A_1^{\mathbb{N}}$ mit $A_1 = \{1, 2, \ldots, 5\} \in \mathcal{A}_1$

$$\begin{aligned}
\Rightarrow A &= \bigcap_{k \in \mathbb{N}} \underbrace{A_1 \times \ldots \times A_1}_{\substack{k\text{-mal} \\ =: A_1^{(k)}}} \times \Omega_1 \times \Omega_1 \times \ldots \in \mathcal{A} \\
\Rightarrow P(A) &= P\left(\bigcap_{k \in \mathbb{N}} A_1^{(k)}\right) = \lim_{k \to \infty} P(A_1^{(k)}) = \lim_{k \to \infty} (P_1(A_1))^k \\
&= \lim_{k \to \infty} (1-p)^k = 0
\end{aligned}$$

- Sei B_k = "beim k-ten Wurf erstmals eine 6"

$$
\begin{array}{llll}
\Rightarrow & P(B_k) & = p \cdot (1-p)^{k-1} & \text{(geometrische Verteilung)} \\
& = P(T = k) & & \\
& = P_T(\{k\}) & \text{mit ZV} & T : \Omega \longrightarrow \mathbb{N} \\
& & & T(\omega) = \text{Nummer des ersten Wurfs, bei dem eine 6 auftritt} \\
& & & = \inf\{n : X_n = 6\} \\
& & & \text{ mit } X_n : \Omega \longrightarrow \Omega_0 \text{ "Ausgang des n-ten Spiels".} \\
& & & X_n \text{ ist ZV.}
\end{array}
$$

P_T ist diskretes W-Maß auf $\mathbb{N}$

T ist geometrisch verteilt.

Teil II

Integrationstheorie

Kapitel 4

Integral und Erwartungswert

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum.
Ziel: Wir wollen das Integral $\int f d\mu$ (genauer: $\int_\Omega f(x) d\mu(x)$ oder $\int_\Omega f(x)\mu(dx)$) von f bzgl. μ für eine möglichst große Klasse von $\mathcal{A}/\mathcal{B}(\overline{\mathbb{R}})$-meßbaren Funktionen $f : \Omega \longrightarrow \overline{\mathbb{R}}$ definieren.

Konstruktion in 4 Schritten.

Schritt 1 (Indikatorfunktion)

Sei $f = \mathbb{1}_A$ mit $A \in \mathcal{A}$.

Definiere: $\displaystyle\int f d\mu := \mu(A)$

Schritt 2 (nicht-negative Elementarfunktion)

$$f = \sum_{i=1}^{k} \alpha_i \cdot \mathbb{1}_{A_i}, \qquad A_i \in \mathcal{A}, k \in \mathbb{N}, \alpha_i \in [0, \infty[$$

Bemerkungen:
a) o.B.d.A. $(A_i)_{i\in I}$ paarweise disjunkt und $\bigcup\limits_{i\in I} A_i = \Omega$

b) Die Darstellung ist nicht eindeutig, aber wenn

$$f = \sum_{i=1}^{k} \alpha_i \cdot \mathbb{1}_{A_i} = \sum_{j=1}^{l} \beta_j \cdot \mathbb{1}_{B_j} \quad \begin{array}{l} \text{mit } (A_i)_{i\in I} \text{ paarweise disjunkt} \\ \text{mit } (B_j)_{j\in J} \text{ paarweise disjunkt} \\ \text{und } \Omega = \bigcup A_i = \bigcup B_j \end{array}$$

so folgt: $\sum\limits_{i=1}^{k} \alpha_i \cdot \mu(A_i) = \sum\limits_{j=1}^{l} \beta_j \cdot \mu(B_j)$

denn: $\mu(A_i) = \sum\limits_{j=1}^{l} \mu(A_i \cap B_j)$ und $\mu(B_j) = \sum\limits_{i=1}^{k} \mu(B_j \cap A_i)$

$$\Rightarrow \sum_i \alpha_i \cdot \mu(A_i) = \sum_{i,j} \alpha_i \cdot \mu(A_i \cap B_j) = \sum_{i,j} \beta_j \cdot \mu(A_i \cap B_j) = \sum_j \mu(B_j)$$

wobei $\alpha_i = \beta_j$, falls $\mu(A_i \cap B_j) > 0$ (denn: $(A_i)_i$ und $(B_j)_j$ sind jeweils paarweise disjunkt).

Definiere : $\displaystyle\int f d\mu := \sum_{i=1}^{k} \alpha_i \cdot \mu(A_i) \qquad \in [0, \infty]$ mit $(A_i)_{i \in I}$ disjunkt

Lemma 4.1. *Für Elementarfunktionen $f, g \geq 0$ und reelles $\alpha \geq 0$ gilt:*

(i) $\displaystyle\int (f+g) d\mu = \int f d\mu + \int g d\mu$

(ii) $\displaystyle\int (\alpha f) d\mu = \alpha \cdot \int f d\mu$

(iii) $\displaystyle f \leq g \Rightarrow \int f d\mu \leq \int g d\mu$

Beweis:

(i) Sei $f = \sum\limits_{i=1}^{k} \alpha_i \cdot \mathbb{1}_{A_i}$ und $g = \sum\limits_{j=1}^{l} \beta_j \cdot \mathbb{1}_{B_j}$ mit $(A_i)_i, (B_j)_j$ jeweils disjunkt und

$$\Omega = \bigcup A_i = \bigcup B_j$$

$$\Rightarrow f = \sum_{i,j} \alpha_i \cdot \mathbb{1}_{A_i \cap B_j},\ g = \sum_{i,j} \beta_j \cdot \mathbb{1}_{A_i \cap B_j}$$

$$\Rightarrow f + g = \sum_{i,j} (\alpha_i + \beta_j) \cdot \mathbb{1}_{A_i \cap B_j}$$

$$\begin{aligned} \Rightarrow \textstyle\int (f+g) d\mu &= \sum_{i,j} (\alpha_i + \beta_j) \cdot \mu(A_i \cap B_j) \\ &= \sum_{i,j} \alpha_i \cdot \mu(A_i \cap B_j) + \sum_{i,j} \beta_j \cdot \mu(A_i \cap B_j) \\ &= \textstyle\int f d\mu + \int g d\mu \end{aligned}$$

(ii) trivial,

(iii) Seien f und g dargestellt wie in (i). Dann gilt $\alpha_i \leq \beta_j$, falls $\mu(A_i \bigcap B_j) > 0$.
$\Rightarrow$ Beh. $\square$

Schritt 3 (Nicht-negative meßbare Funktionen):

$f : \Omega \longrightarrow [0, \infty]$ $\mathcal{A}/\mathcal{B}(\bar{\mathbb{R}})$meßbar
$\Rightarrow f = \sup_{n \in \mathbb{N}} f_n$ mit $(f_n)_n$ isotone Folge von Elementarfunktionen mit $f_n \geq 0$

Lemma 4.2. *Sind $(u_n)_n$ und $(v_m)_m$ isotone Folgen nicht-negativer Elementarfunktionen mit*

$$\sup_n u_n = \sup_m v_m \quad \text{, so ist}$$

$$\sup_n \int u_n d\mu = \sup_m \int v_m d\mu.$$

Beweis:
Sei $v_m = \sum_{j=1}^{k^{(m)}} \alpha_j^{(m)} \cdot \mathbb{1}_{A_j^{(m)}}$ und $B_{n,m}^{\alpha} = \{u_n \geq \alpha \cdot v_m\} \in \mathcal{A}$ $(\forall \alpha \in \mathbb{R})$.
Für $\alpha \in]0, 1[$ folgt wegen $\sup_n u_n \geq v_m$

$$\bigcup_n B_{n,m}^{\alpha} = \Omega$$

und damit $\bigcup_n B_{n,m}^{\alpha} \bigcap A_j^{(m)} = A_j^{(m)}$.
Offenbar ist $u_n \geq \alpha \cdot v_m \cdot \mathbb{1}_{B_{n,m}^{\alpha}}$ =Elementarfunktion.

$$\Rightarrow \quad \int u_n d\mu \geq \alpha \int (v_m \cdot \mathbb{1}_{B_{n,m}^{\alpha}}) d\mu = \alpha \cdot \sum_{j=1}^{k^{(m)}} \alpha_j^{(m)} \cdot \mu(A_j^{(m)} \cap B_{n,m}^{\alpha})$$

$$\Rightarrow \sup_n \int u_n d\mu \geq \alpha \cdot \sup_n \sum_j \alpha_j^{(m)} \cdot \mu(A_j^{(m)} \cap B_{n,m}^{\alpha})$$

$$= \alpha \cdot \sum_j \alpha_j^{(m)} \cdot \mu(A_j^{(m)}) = \alpha \cdot \int v_m d\mu$$

Für $\alpha \to 1$: $\quad \sup_n \int u_n d\mu \geq \sup_m \int v_m d\mu$. Und umgekehrt! $\square$

Definition 4.3. *Für $f \geq 0$, $\mathcal{A}/\mathcal{B}(\overline{\mathbb{R}})$-meßbar ist*

$$\int f d\mu := \sup_n \int f_n d\mu \qquad \in [0, \infty]$$

mit $(f_n)_n$ beliebige isotone Folge von Elementarfunktionen mit $f_n \nearrow f$.

Lemma 4.4. *Für meßbare $f, g > 0$ gilt:*

(i) $\int (f+g)d\mu = \int f d\mu + \int g d\mu$

(ii) $\int (\alpha f)d\mu = \alpha \cdot \int f d\mu$

(iii) $f \leq g \Rightarrow \int f d\mu \leq \int g d\mu$

Beweis: klar □

Schritt 4 (integrierbare Funktionen):

$$f : \Omega \longrightarrow [-\infty, +\infty] \quad \text{meßbar}$$

und sei $f_+ = \sup\{f, 0\}$, $f_- = \sup\{-f, 0\}$.

Definition 4.5.

a) $f : \Omega \longrightarrow \overline{\mathbb{R}}$ *heißt* **integrierbar** *bzgl* μ *oder* μ-**integrierbar**, *falls* f $\mathcal{A}/\mathcal{B}(\overline{\mathbb{R}})$-*meßbar ist und* $\int |f| d\mu < \infty$.

b) Falls $\int f_+ d\mu < \infty$ <u>*oder*</u> $\int f_- d\mu < \infty$, *so setzt man*

$$\int f d\mu := \int f_+ d\mu - \int f_- d\mu \qquad \in [-\infty, +\infty]$$

und $\int f d\mu$ *heißt* μ-**Integral von** f.

c) $\int\limits_A f d\mu := \int 1\!\!1_A \cdot f d\mu$, $(\forall A \in \mathcal{A})$, $\int\limits_A f d\mu$ *heißt* μ-**Integral von** f **auf** A.

Bemerkungen:

a) Für eine $\mathcal{A}/\mathcal{B}(\overline{\mathbb{R}})$-meßbare Funktion f sind äquivalent:
- f ist integrierbar,
- f_+ <u>und</u> f_- sind integrierbar,
- $f = u - v$ mit integrierbaren $u, v \geq 0$,
- $|f|$ ist integrierbar,
- $|f| \leq w$ mit integrierbaren w

b) Ist $f = u - v$ mit integrierbaren u, v, so ist
$$\int f d\mu = \int u d\mu - \int v d\mu$$

c) Ist f integrierbar, so ist $\mu(\{x : f(x) = +\infty \text{ oder } f(x) = -\infty\}) = 0$.

d) Sind f, g integrierbar mit $\mu(\{x : f(x) \neq g(x)\}) = 0$, so ist
$\int f d\mu = \int g d\mu$

f) f, g integrierbar $\Rightarrow \mu(\{x : f(x) + g(x) \text{ nicht definiert }\}) = 0$ und
$\int (f+g) d\mu = \int f d\mu + \int g d\mu$

Definition 4.6. $\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ *=Menge aller reellwertigen μ-integrierbaren Funktionen*

Proposition 4.7. $\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ *ist ein reeller Vektorraum und die Integration $f \mapsto \int f d\mu$ bzgl. μ ist eine positive Linearform auf $\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$.*

(D.h. es gilt (i)-(iii) von oben $\forall f, g \in \mathcal{L}^1,\ \forall \alpha \in \mathbb{R}$.)

Beweis: Klar! □

e) Zur Definition von $\int f d\mu$ genügt es, wenn $f : \Omega_0 \longrightarrow \overline{\mathbb{R}}$ (meßbar) gegeben ist mit $\mu(\Omega \setminus \Omega_0) = 0$
$\Rightarrow f$ läßt sich fortsetzen zu meßbarem $\tilde{f} : \Omega \longrightarrow \overline{\mathbb{R}}$ mit $f = \tilde{f}$ auf Ω_0.

Bemerkung: Man sagt, eine Eigenschaft E gilt μ**-fast überall** (μ-f.ü.) auf Ω, falls

$$\mu(\{x : E(x) \text{ gilt nicht }\}) = 0$$

Beispiele 4.8.
$f \leq g$ μ-f.ü. $\Rightarrow \int f d\mu \leq \int g d\mu$
$f = g$ μ-f.ü. $\Rightarrow \int f d\mu = \int g d\mu$
$\int |f| d\mu < \infty \Rightarrow |f| < \infty$ μ-f.ü.

Spezielle Fälle:

(i) Ist $(\Omega, \mathcal{A}, P)$ ein W-Raum und X eine ZV auf Ω mit $\int |X| dP < \infty$, so heißt

$$E(X) := \int X dP$$

Erwartungswert von X (= Prognosewert, mittlerer Wert, im Schnitt zu erwartender Wert von X).
Für $\int_A X dP$ schreibt man $E(X; A)$.
Statt P-fast überall, sagt man: P**-fast sicher** (P-f.s.).

(ii) Ist $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$, so schreibt man

$\int f(x)dx$ für $\int f d\lambda^d$ (**Lebesgue-Integral**).

Im Falle $d = 1$ und $A =]a,b[$ oder $= [a,b]$ schreibt man

$\int\limits_a^b f(x)dx$ für $\int\limits_A f d\lambda^1$

(iii) Ist $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ mit $\mu(]a,b]) = F(b) - F(a)$ mit isotoner, rechtsseitig stetiger Funktion $F : \mathbb{R} \longrightarrow \mathbb{R}$ (=Verteilungsfunktion), so schreibt man

$\int f(x)dF(x)$ für $\int f d\mu$ (**Stieltjes-Integral**).

Ist $F \in \mathcal{C}^1(\mathbb{R})$ mit $F' = \varphi$, so gilt:

$\int f(x)dF(x) = \int f(x)\varphi(x)dx$.

(iv) Ist Ω abzählbar, $\mathcal{A} = \mathcal{P}(\Omega)$ und $\mu = \sum\limits_{k\in\Omega} \delta_k$ (**Zählmaß**), so ist

$$\int f d\mu = \sum_{k\in\Omega} f(k)$$

Beispiel 4.9. Würfeln mit fairem Würfel

$\Omega = \Omega_1^{\mathbb{N}}, \mathcal{A} = \mathcal{A}_1^{\mathbb{N}}, P = P_1^{\mathbb{N}}$ mit $\Omega_1 = \{1, \ldots, 6\}, \mathcal{A}_1 = \mathcal{P}(\Omega_1)$

$P_1(k) = \frac{1}{6}\ \forall k \in \Omega_1$

- $X_n : \omega \mapsto \omega_n$ ZV "Ausgang des n-ten Spiels"

$$\begin{aligned} E(X_n) = \int X_n dP = \int X_n dP_n &= \sum_{\omega_n\in\Omega_1} \omega_n \cdot P_n(\{\omega_n\}) \\ &= \sum_{k=1}^{6} k \cdot P_1(\{k\}) = \frac{1}{6}\sum_{k=1}^{6} k = 3,5 \end{aligned}$$

="im Schnitt zu erwartende Augenzahl"

- $S_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ ZV "durchschnittliche Augenzahl in den ersten n Spielen"

$$\Rightarrow E(S_n) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) \underset{\substack{\uparrow \\ \text{Linearität}}}{=} \frac{1}{n}\sum_{i=1}^{n} E(X_i) = 3,5$$

Beispiel 4.10. Beliebiger Würfel, $P_1(\{6\}) = p > 0$.

$T(\omega) = \min\{n : X_n(\omega) = 6\}$ ZV "Wartezeit bis zur ersten 6"

$\Rightarrow P(\{T = k\}) = p \cdot (1-p)^{k-1}$

$$\begin{aligned} \Rightarrow\ E(T) &= \int_\Omega T(\omega)dP(\omega) \\ &= \sum_{k\in\mathbb{N}} \int_{B_k} T(\omega)dP(\omega) = \sum_{k\in\mathbb{N}} k \cdot P(B_k) \\ &= \sum_{k\in\mathbb{N}} k \cdot p \cdot (1-p)^{k-1} = \frac{1}{p} \end{aligned}$$

Beispiel 4.11. n Kugeln werden zufällig in r Urnen $F_1, \ldots, F_r$ gelegt.

$B_k = \{$ beim k-ten Wurf erstmals eine 6 $\}$

$p_k(r,n)$ = "Wahrscheinlichkeit dafür, daß k Urnen leer bleiben" = ...

(s. Übung)

Sei K die Anzahl der leeren Urnen $\Rightarrow E(K) = \sum\limits_{k=0}^{r} k \cdot p_k(r,n) = ?$

Anderer Weg:

Mit $A_i = \{\omega : \text{ Urne } F_i \text{ ist leer } \}$ gilt:

$K = \sum\limits_{i=1}^{r} \mathbb{1}_{A_i} \Rightarrow E(K) = \sum\limits_{i=1}^{r} P(A_i) = r \cdot \left(1 - \frac{1}{r}\right)^n$

Sei nun $\frac{n}{r} = c$ und $K = K_r$. Für $r \to \infty$ gilt:

$$\frac{E(K_r)}{r} = \left(1 - \frac{1}{r}\right)^{r \cdot c} \xrightarrow{(\leq)} e^{-c}.$$

Also: 1 Million Kugeln in 1 Million Urnen $\Rightarrow \frac{1}{e} \cdot 1$ Million Urnen leer.

Kapitel 5

Integration bzgl. eines Bildmaßes, Produktmaßes oder Maßes mit Dichte

Satz 5.1 (Transformationssatz). *Gegeben* $(\Omega, \mathcal{A}, \mu)$ *und* $(\Omega', \mathcal{A}')$ *sowie* $\mathcal{A}/\mathcal{A}'$*-meßbare Abbildung* $T : \Omega \longrightarrow \Omega'$. *Dann gilt für jede* $\mathcal{A}'/\mathcal{B}(\overline{\mathbb{R}})$*-meßbare Funktion* $f' : \Omega' \longrightarrow \overline{\mathbb{R}}$:

(a) $\displaystyle\int_{\Omega'} f'_{\pm} dT(\mu) = \int_{\Omega} f'_{\pm} \circ T d\mu$

(b) f' $T(\mu)$*-integrierbar* $\Longleftrightarrow$ $f' \circ T$ μ*-integrierbar*

Beweis: (b) folgt aus (a),
(a) wird in 3 Schritten bewiesen:

1. $f' = \mathbb{1}_{A'}$ mit $A' \in \mathcal{A}'$
   $\Rightarrow \int f' dT(\mu) = T(\mu)(A') \underset{\substack{\uparrow \\ \text{Def. vom Bildmaß}}}{=} \mu(T^{-1}(A')) = \int f d\mu$ mit $f = \mathbb{1}_{T^{-1}(A')} = \mathbb{1}_{A'} \circ T$

2. Linke und rechte Seite in (a) ist linear in f'
   $\Rightarrow$ (mit 1.) Behauptung für $f' = \sum \alpha_i \cdot \mathbb{1}_{A'_i}$ = nicht-negative Elementarfunktion.

3. Übergang zu isotonen Limiten $\Rightarrow$ Behauptung für meßbare $f' \geq 0$. □

Korollar 5.2. *Sei X eine reelle ZV auf $(\Omega, \mathcal{A}, P)$ und $f : \mathbb{R} \longrightarrow \overline{\mathbb{R}}$ meßbar mit $f(X)$ P-integrierbar oder ≥ 0. Dann gilt:*

$$E(f(X)) = \int_{\mathbb{R}} f(t) dP_X(t)$$

$$\textit{Insbesondere gilt:} \quad E(X) = \int_{\mathbb{R}} t dP_X(t)$$
$$\textit{und} \quad E(X^2) = \int_{\mathbb{R}} t^2 dP_X(t)$$

Falls X nur Werte in $\mathbb{N}$ annimmt, gilt

$$E(X) = \sum_{k \in \mathbb{N}} k \cdot P(X = k)$$
$$\textit{und} \quad E(X^2) = \sum_{k \in \mathbb{N}} k^2 \cdot P(X = k)$$

Beispiel 5.3 (Würfeln, Bewerben, ... mit Erfolgswahrscheinlichkeit $p > 0$).

T = Anzahl der "Spiele" bis zum ersten Erfolg (einschließlich)

$$\begin{aligned}
\Rightarrow E(T) &= \sum_{k \in \mathbb{N}} k \cdot P(T = k) = p \cdot \sum_{k \in \mathbb{N}} k \cdot (1-p)^{k-1} \\
&= \tfrac{p}{1-p} \cdot \sum_{k \geq 1} \sum_{l \geq 1} \mathbb{1}_{\{l \leq k\}} \cdot (1-p)^k \underset{\substack{\uparrow \\ \text{Fubini}}}{=} \tfrac{p}{1-p} \cdot \sum_{l \geq 1} \sum_{k \geq l} (1-p)^k \\
&= \tfrac{p}{1-p} \cdot \sum_{l \geq 1} \tfrac{(1-p)^l}{p} = \tfrac{1}{p}.
\end{aligned}$$

Bemerkung: $E(X)$ und $E(X^2)$ etc. sind also durch die Verteilung P_X auf $(\mathbb{R}, \mathcal{B}(\overline{\mathbb{R}}))$ bestimmt. ($\Rightarrow$ "$(\Omega, \mathcal{A}, P)$ ist irrelevant")

Satz 5.4 (von Fubini). *Seien $(\Omega_i, \mathcal{A}_i, \mu_i)$ für $i = 1, 2$ zwei σ-endliche Maßräume und sei $f : \Omega_1 \times \Omega_2 \longrightarrow \overline{\mathbb{R}}$ eine $\mathcal{A}_1 \otimes \mathcal{A}_2 / \mathcal{B}(\overline{\mathbb{R}})$-meßbare Funktion.*

a) Ist $f \geq 0$, so ist $f(x, \cdot) : y \mapsto f(x, y)$ für jedes $x \in \Omega_1$ $\mathcal{A}_2 / \mathcal{B}(\overline{\mathbb{R}})$-meßbar und $f_1 : x \mapsto \int f(x, y) \mu_2(dy)$ $\mathcal{A}_1 / \mathcal{B}(\overline{\mathbb{R}})$-meßbar.
Entsprechend ist $f(\cdot, y) : x \mapsto f(x, y)$ für jedes $y \in \Omega_2$ $\mathcal{A}_1 / \mathcal{B}(\overline{\mathbb{R}})$-meßbar und $f_2 : y \mapsto \int f(x, y) \mu_1(dx)$ $\mathcal{A}_2 / \mathcal{B}(\overline{\mathbb{R}})$-meßbar.

b) Ist $f \in \mathcal{L}^1(\mu_1 \otimes \mu_2)$, so ist $f(x, \cdot) \in \mathcal{L}^1(\mu_2)$ für μ_1-f.a. x und $f_1 \in \mathcal{L}^1(\mu_1)$.
Entsprechend $f(\cdot, y) \in \mathcal{L}^1(\mu_1)$ für μ_2-f.a. y und $f_2 \in \mathcal{L}^1(\mu_2)$.

c) Ist $f \geq 0$ oder $f \in \mathcal{L}^1(\mu_1 \otimes \mu_2)$, so gilt:

$$\int_{\Omega_1}\int_{\Omega_2} f(x,y)\mu_2(dy)\mu_1(dx) = \int_{\Omega_1\times\Omega_2} f d\mu_1\otimes\mu_2 = \int_{\Omega_2}\int_{\Omega_1} f(x,y)\mu_1(dx)\mu_2(dy)$$

Beweis: durch maßtheoretische Induktion.

Sei also (o.B.d.A.?!) $f = \mathbb{1}_A$ mit $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$.

Der x-Schnitt von A ist die Menge $A_x = \{y : (x,y) \in A\} \subset \Omega_2$.

(i) Beh.: $A_x \in \mathcal{A}_2$ für jedes x.

Denn: $(A_x)^c = (A^c)_x$ und $\bigcup(A_i)_x = (\bigcup A_i)_x$

$\Rightarrow \tilde{\mathcal{A}} := \{A \subset \Omega_1 \times \Omega_2 : A_x \in \mathcal{A}_2\}$ ist σ-Algebra

und $\tilde{\mathcal{A}}$ enthält $\{A_1 \times A_2 : A_i \in \mathcal{A}_i\} \Rightarrow \tilde{\mathcal{A}} \supset \mathcal{A}_1 \otimes \mathcal{A}_2$.

(ii) Beh.: $x \mapsto \mu_2(A_x)$ ist $\mathcal{A}_1/\mathcal{B}(\overline{\mathbb{R}})$-meßbar und

$$\int \mu_2(A_x)d\mu_1(x) = \mu(A) \qquad \text{mit } \mu = \mu_1 \otimes \mu_2.$$

Denn: o.B.d.A. μ_1 und μ_2 endlich.

Dann ist $\mathcal{D} = \{A \subset \Omega_1 \times \Omega_2 :$ Beh.(ii) gilt $\}$ ein Dynkin-System.

Nämlich,

- $\Omega \in \mathcal{D}$
- $A \in \mathcal{D} \Rightarrow$
  $\mu_2((A^c)_x) = \mu_2((A_x)^c) = \mu_2(\Omega_2) - \mu_2(A_x) =$ meßbar als Funktion von x
  und $\int \mu_2((A^c)_x)d\mu_1(x) = \mu_1(\Omega_1)\cdot\mu_2(\Omega_2) - \mu(A) = \mu(A^c)$.
- A_n disjunkt $\in \mathcal{D} \Rightarrow \mu_2\left((\bigcup A_n)_x\right) = \mu_2\left(\bigcup(A_n)_x\right) = \sum\limits_{n\in\mathbb{N}} \mu_2((A_n)_x)$ meßbar
  als Funktion von x
  und $\int \mu_2((\bigcup A_n)_x)d\mu_1(x) = \sum\int \mu_2((A_n)_x)d\mu_1(x) = \sum \mu((A_n)) = \mu(\bigcup(A_n)) = \mu(A)$.

Ferner gilt: $\mathcal{D}$ enthält das $\bigcap$-stabile System $\{A_1 \times A_2 : A_i \in \mathcal{A}_i\}$

$\Rightarrow \mathcal{D} \supset \mathcal{A}_1 \otimes \mathcal{A}_2$. □

Korollar 5.5. *Sei* $(\Omega, \mathcal{A}, P)$ *W-Raum und* X_n *ZV mit* $X_n \geq 0$ $(\forall n \in \mathbb{N})$ *oder* $\sum\limits_n E(|X_n|) < \infty$. *Dann:*

$$E(\sum_n X_n) = \sum_n E(X_n).$$

Beweis: Fubini mit $\mu_1 = P$ und $\mu_2 = \sum\limits_{n\in\mathbb{N}} \delta_n$ ("Zählmaß"). □

Beispiel 5.6. Sei $f(x,y) = \frac{x^2-y^2}{(x^2+y^2)^2}$ in $]0,1[^2 \subset \mathbb{R}^2$.

Dann ist $\int\limits_0^1\int\limits_0^1 f(x,y)dydx = \frac{\pi}{4}$

und $\int\limits_0^1\int\limits_0^1 f(x,y)dxdy = -\frac{\pi}{4}$, da gilt: $-\int\limits_0^1 \frac{x^2-y^2}{(x^2+y^2)^2} dx = -\frac{x}{x^2+y^2}\Big|_0^1 = -\frac{1}{1+y^2}$

<u>und</u> $-\int\limits_0^1 \frac{1}{1+y^2} dy = -\arctan y|_0^1 = -\frac{\pi}{4}$

Beispiel 5.7. Sei $\mu_1 = \lambda^1$ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und $\mu_2 = \sum\limits_{x \in \mathbb{R}} \delta_x$ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

Für $f(x,y) = \begin{cases} 1 & , \quad x = y \\ 0 & , \quad \text{sonst} \end{cases}$ gilt

$$\int \int f(x,y)\mu_1(dx)\mu_2(dy) = 0$$

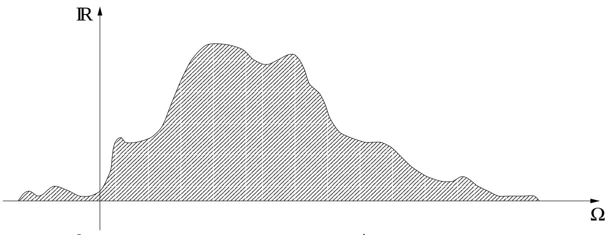
und $$\int \int f(x,y)\mu_2(dy)\mu_1(dx) = \infty.$$

(Daraus folgt notwendig, daß μ_2 nicht σ-endlich ist.)

Anwendungen von "Fubini"

a) Vertauschen von $\sum$ mit $\int$: Korollar 7.4.

b) Interpretation von $\int\limits_\Omega f(x)\mu(dx)$ für $f \geq 0$ als Fläche unter dem Graphen



d.h. $\int\limits_\Omega f(x)\mu(dx) = (\mu \otimes \lambda^1)(F)$ mit $F = \{(x,t) \in \Omega \times \mathbb{R} : 0 \leq t \leq f(x)\}$

Sei $(\Omega, \mathcal{A}, \mu)$ Maßraum und $\varphi \geq 0$ $\mathcal{A}/\mathcal{B}(\overline{\mathbb{R}})$-meßbar.

Satz 5.8. *Durch* $\nu : A \mapsto \int\limits_A \varphi d\mu$ *wird ein Maß* ν *auf* $(\Omega, \mathcal{A})$ *definiert. Für jedes* $\mathcal{A}/\mathcal{B}(\overline{\mathbb{R}})$*-meßbare* $f \geq 0$ *gilt*

$$\int f d\nu = \int f \cdot \varphi d\mu.$$

Beweis: Seien (A_n) disjunkt. Dann gilt
$\nu\left(\bigcup A_n\right) = \int \sum \left(\varphi \cdot \mathbb{1}_{A_n}\right) d\mu = \sum \int \varphi \cdot \mathbb{1}_{A_n} d\mu = \sum \nu(A_n)$.
$\Rightarrow \nu$ ist Maß
Rest folgt mit maßtheoretischer Induktion. $\square$

Schreibweise: $\nu = \varphi \cdot \mu$ bzw. $\nu(dx) = \varphi(x) \cdot \mu(dx)$
(=“Maß mit Dichte φ bzgl μ“)

Definition 5.9. *Seien μ, ν Maße auf $(\Omega, \mathcal{A})$.*
a) ν heißt **absolut-stetig** *bzgl.μ (kurz: $\nu \ll \mu$), falls*

$$\forall A \in \mathcal{A} : \mu(A) = 0 \Longrightarrow \nu(A) = 0$$

b) ν heißt **singulär** *bzgl. μ (kurz: $\nu \perp \mu$) oder ν und μ sind* **singulär zueinander**, *falls*

$$\exists A \in \mathcal{A} : \mu(A) = 0 \text{ und } \nu(A^c) = 0$$

Satz 5.10 (von Radon-Nikodym). *Sei μ σ-endlich. Dann gilt:*

$$\nu \ll \mu \Longleftrightarrow \exists \mathcal{A}/\mathcal{B}(\overline{\mathbb{R}})\text{-meßbare Funktion } \varphi \geq 0 \text{ mit } \nu = \varphi \cdot \mu$$

φ ist dabei μ-f.ü. eindeutig bestimmt.

Schreibweise: $\varphi = \frac{d\nu}{d\mu}$ “Radon-Nikodym-Dichte“ bzw. “Radon-Nikodym-Ableitung“

Beweis:
“$\Leftarrow$“ einfach: $\forall A \in \mathcal{A}$ mit $\mu(A) = 0$ gilt: $\mathbb{1}_A \cdot \varphi = 0$ μ-f.ü.
$\Rightarrow \nu(A) = \int \mathbb{1}_A \cdot \varphi d\mu = 0$
“$\Rightarrow$“ s. [Bauer-Maßtheorie]

Idee: Betrachte $\Phi = \{\tilde{\varphi} \geq 0 \text{ meßbar mit } \int_A \tilde{\varphi} d\mu \leq \nu(A) \;\; \forall A \in \mathcal{A}\}$
$\Rightarrow \varphi$ ist sup-stabil (d.h. mit $\tilde{\varphi}_1$ und $\tilde{\varphi}_2$ in φ, so auch $\sup(\ldots)$)
und $\exists \varphi \in \Phi$ mit $\int \varphi d\mu = \sup_{\tilde{\varphi} \in \Phi} \int \tilde{\varphi} d\mu$
$\Rightarrow \nu = \varphi \cdot \mu$. $\square$

μ, ν heißen **äquivalent** ($\mu \sim \nu$), falls $\nu \ll \mu$ und $\nu \ll \mu$.
In diesem Fall gilt: $\frac{d\nu}{d\mu} = \left(\frac{d\mu}{d\nu}\right)^{-1}$ f.ü. (bzgl. μ oder/und ν).

Allgemein: $\mu \ll \nu$ und $\nu \ll \rho$ (σ-endlich angenommen) $\Rightarrow \mu \ll \rho$ und $\frac{d\mu}{d\rho} = \frac{d\mu}{d\nu} \cdot \frac{d\nu}{d\rho}$ ρ-f.ü..

Beispiel 5.11. $(\Omega, \mathcal{A})$ sei diskret.
$\Rightarrow \frac{d\nu}{d\mu}(x) = \frac{\nu(\{x\})}{\mu(\{x\})}$ $\forall x$ mit $\mu(\{x\}) > 0$.

Korollar 5.12 (Lebesgue-Zerlegung). *Seien ν, μ σ-endliche Maße auf $(\Omega, \mathcal{A})$. Dann gibt es eindeutig bestimmte Maße ν_s und ν_a mit $\nu_s \perp \mu, \nu_a \ll \mu$ und*

$$\nu = \nu_s + \nu_a.$$

ν_s heißt **singulärer Anteil** *von ν und ν_a* **absolut-stetiger Anteil**.

Beweis:
Betrachte $\rho = \nu + \mu$. Sei $\varphi = \frac{d\nu}{d\rho}$ und $\psi = \frac{d\mu}{d\rho} = 1 - \varphi$.
Sei $\Phi = \{\varphi < 1\} \Rightarrow \nu_a(A) = \nu(A \cap \Phi)$ und $\nu_s(A) = \nu(A \cap \Phi^c)$

Dann:• $\nu_a + \nu_s = \nu$

•$\nu_a \ll \mu$: Wegen $\nu_a(A) = \int\limits_{\{\varphi<1\}} \mathbb{1}_A \cdot \varphi d\rho = \int\limits_{\{\varphi<1\}} \mathbb{1}_A \cdot \frac{\varphi}{1-\varphi} d\mu = \int\limits_{A\cap\{\varphi<1\}} \frac{\varphi}{1-\varphi} d\mu$
ist $\nu_a(A) = 0$, falls $\mu(A) = 0$.

• $\nu_s \perp \mu$ wegen
$\nu_s(\Phi) = 0$ und $\mu(\Phi^c) = \int\limits_{\Phi^c} d\mu = \int\limits_{\Phi^c} (1-\varphi) d\rho = 0$.

Damit folgt die Existenz der Zerlegung. Die Zerlegung ist eindeutig, da $\tilde{\nu} \ll \mu$ und $\tilde{\nu} \perp \mu \Rightarrow \tilde{\nu} = 0$.
Wäre also $\nu = \tilde{\nu}_s + \tilde{\nu}_a$, so bilde $(\tilde{\nu}_{s/a})_a(A) = \tilde{\nu}_{s/a}(A \cap \Phi)$ und $(\tilde{\nu}_{s/a})_s(A) = \tilde{\nu}_{s/a}(A \cap \Phi^c)$
$\Rightarrow \tilde{\nu}_{aa} \leq \nu_a, \tilde{\nu}_{ss} \leq \nu_s$ und $\tilde{\nu}_{sa} = \tilde{\nu}_{as} = 0$. $\square$

Bemerkung: Für ein endliches Maß ν auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ gibt es eine (i.a. von der Lebesgue-Zerlegung verschiedene) eindeutige Zerlegung $\nu = \nu_c + \nu_d$ in einen "stetigen Anteil" ν_c, d.h. mit stetiger Verteilungsfunktion $x \mapsto \nu_c(]-\infty, x])$, und einen "unstetigen Anteil"$\nu_d = \sum\limits_{i\in\mathbb{N}} \alpha_i \cdot \delta_{x_i}$.
Ist $\nu = \nu_s + \nu_a$ die Lebesgue-Zerlegung von ν bzgl. λ^1, so gilt:
$\nu_c \geq \nu_a$ und $\nu_d \leq \nu_s$.
I.a. aber <u>keine</u> Gleichheit! (Siehe Gleichverteilung auf Cantor-Menge.)
Offenbar $\nu_a = (\nu_c)_a = (\nu_a)_c$ und $\nu_d = (\nu_d)_s = (\nu_s)_d$.
Daher: $\nu = \nu_a + \nu_{sc} + \nu_d$ mit $\nu_{sc} = (\nu_c)_s = (\nu_s)_c$ "singulär-stetiger Anteil".
Anwendung in der Spektraltheorie, Physik.

Kapitel 6

L^p- Räume, Ungleichungen

Gegeben $(\Omega, \mathcal{A}, P)$ W-Raum.

Satz 6.1 (Jensensche Ungleichung). *Sei* $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ *konvex (d.h.* $\lambda\varphi(x) + (1-\lambda)\varphi(y) \geq \varphi(\lambda x + (1-\lambda)y)$ $\forall x, y \in \mathbb{R}, \forall \lambda \in [0,1]$*) und* X *ZV auf* Ω*, mit* X *und* $\varphi \circ X$ *integrierbar. Dann gilt:*

$$\varphi(E(X)) \leq E(\varphi \circ X)$$

Wichtig: Gilt nur für normierte Maße!

Beweis: $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ konvex $\Longleftrightarrow$
$\forall x_0 \in \mathbb{R} :$ $\exists$ lineare Funktion $\ell : \mathbb{R} \longrightarrow \mathbb{R}$, $x \mapsto ax + b$
mit $\ell(x_0) = \varphi(x_0)$ und $\ell(x) \leq \varphi(x)$ $(\forall x)$ "Stützebene" .
Wähle $x_0 = E(X)$

$$\Rightarrow \varphi(E(X)) = \ell(E(X)) = a \cdot E(X) + b \underset{\substack{\uparrow \\ P \text{ normiert}}}{=} E(aX+b) = E(\ell(X)) \underset{\substack{\uparrow \\ \text{Monotonie}}}{\leq} E(\varphi(X)).$$

$\square$

Beispiele 6.2.

(1) $|E(X)| \leq E(|X|)$ (mit $\varphi(x) = |x|$)

(2) $\exp(\pm E(X)) \leq E(\exp(\pm X))$ (mit $\varphi(x) = e^{\pm x}$)

(3) $E(X)^2 \leq E(X^2)$ (mit $\varphi(x) = x^2$)

$$\operatorname{var}(X) := E(X^2) - E(X)^2 = E[(X - E(X))^2] \in [0, \infty]$$

heißt **Varianz** von X.
$\text{var}(X) =$ "mittlerer quadratischer Prognosefehler"
Kennzahl für Fluktuationen um $E(X)$, Charakterisierung des Risikos;

$$\text{var}(X) = 0 \Leftrightarrow X \text{ f.s. konstant, und zwar } = E(X)$$

d.h X verhält sich "deterministisch".
Offenbar $\text{var}(aX + b) = a^2 \cdot \text{var}(X)$.
$\sigma(X) = \sqrt{\text{var}(X)}$ heißt **Standard-Abweichung.**
(Bitte nicht mit von X erzeugte σ-Algebra verwechseln!)

(4) $|E(X)|^p \leq E(|X|^p),\ 1 \leq p < \infty$ (mit $\varphi(x) = |x|^p$)

Sei

$$\|X\|_p = E(|X|^p)^{\frac{1}{p}}$$

die "L^p-Norm" und

$$\mathcal{L}^p(\Omega, \mathcal{A}, P) = \{\text{reelle ZV } X \text{ auf } (\Omega, \mathcal{A}) \text{ mit } \|X\|_p < \infty\} = \{X : |X|^p \in \mathcal{L}^1(\Omega, \mathcal{A}, P)\}$$

der Raum der p-fach integrierbaren ZV auf Ω.
Aus (4) folgt: $\|X\|_p \leq \|X\|_q$ und damit $\mathcal{L}^p \supset \mathcal{L}^q$, falls $p \leq q$ mit $p, q \in [1, \infty[$.
(Achtung: Das gilt nicht für allgem. Maß-Räume!)
Es gilt: $X \in \mathcal{L}^2 \Leftrightarrow X \in \mathcal{L}^1$ und $\text{var}X < \infty$.
Ferner sei

$$\|X\|_\infty = \inf\{\alpha \in \overline{\mathbb{R}} : |X| \leq \alpha\ P\text{-f.s.}\}$$

und

$$\mathcal{L}^\infty(\Omega, \mathcal{A}, P) = \{ \text{ ZV } X \text{ auf } (\Omega, \mathcal{A}) \text{ mit } \|X\|_\infty < \infty\}.$$

Offenbar $\mathcal{L}^p \supset \mathcal{L}^\infty\ \forall p$.

Satz 6.3 (Höldersche Ungleichung). *Seien $p, q \in [1, \infty]$ mit $\frac{1}{p} + \frac{1}{q} = 1$ (z.B. $p = q = 2$ oder $p = 1, q = \infty$).*
Dann gilt $\forall$ ZV X, Y auf Ω:

$$\|XY\|_1 \leq \|X\|_p \cdot \|X\|_q$$

Bemerkungen:
a) Gilt auch für allgem. Maßräume.

b) Allgem.: $\frac{1}{p} + \frac{1}{q} = \frac{1}{r} \Rightarrow \|X \cdot Y\|_r \leq \|X\|_p \cdot \|Y\|_q$

c) Speziell: $p = q = 2 \Rightarrow$ Cauchy-Schwarz-Ungleichung

Beweis:
Falls $\|X\|_p$ oder $\|Y\|_q = 0$, so folgt $|X \cdot Y| = 0$ P-f.s..
Also o.B.d.A $\|X\|_p$ und $\|Y\|_q > 0$ und auch $< \infty$. Durch Ersetzen von X durch $X/\|X\|_p$ und Y durch $Y/\|Y\|_q$ kann man $\|X\|_p = 1 = \|Y\|_q$ erzwingen.
Sei zunächst $p, q \in]1, \infty[$. Die Behauptung folgt dann aus

$$(*) \qquad x \cdot y \leq \frac{x^p}{p} + \frac{y^q}{q} \qquad (x, y \geq 0)$$

Nämlich, $E|X \cdot Y| \leq E(\frac{|X|^p}{p} + \frac{|Y|^q}{q}) = \frac{1}{p} + \frac{1}{q} = 1$.

Zum Beweis von $(*)$: Sei $\varphi(x) = \frac{x^p}{p} + \frac{y^q}{q} - x \cdot y$
$\Rightarrow \varphi$ hat Min. an der Stelle $x_0 = y^{\frac{1}{p}-1}$ und $\varphi(x_0) = 0$.
Sei nun $p = 1$ und $q = \infty$. (Umgekehrt analog!)

Aus $\|Y\|_q = 1$ folgt $|Y| \leq 1$ P-f.s. $\Rightarrow |X \cdot Y| \leq |X|$ P-f.s.
$\Rightarrow E(|X \cdot Y|) \leq E(|X|)$. □

Bemerkung: $X \in \mathcal{L}^p,\ Y \in \mathcal{L}^1 \Rightarrow X \cdot Y \in \mathcal{L}^1$.
Insbesondere $X, Y \in \mathcal{L}^2 \Rightarrow X \cdot Y \in \mathcal{L}^1$.

Satz 6.4 (Minkowski'sche Ungleichung). *Für $p \in [1, \infty]$ und ZV X, Y auf Ω mit (P-f.s.) wohldefiniertem $X + Y$ gilt*

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Beweis: $p = 1$: bereits bewiesen

$p = \infty$: klar (denn: $|X| \leq \alpha$ P-f.s. und $|Y| \leq \beta$ P-f.s. $\Rightarrow |X + Y| \leq \alpha + \beta$ P-f.s.)
$1 < p < \infty$: wähle q mit $\frac{1}{p} + \frac{1}{q} = 1$,
o.B.d.A. $\|X\|_p < \infty$, $\|Y\|_p < \infty \Rightarrow \|X + Y\|_p < \infty$
(denn: $|X + Y|^p \leq 2^p \cdot (|X|^p + |Y|^p)$)

$$\begin{aligned} \Rightarrow \quad \|X + Y\|_p^p &= E(|X + Y|^{p-1} \cdot |X|) + E(|X + Y|^{p-1} \cdot |Y|) \\ &\underset{\substack{\uparrow \\ \text{Hölder}}}{\leq} \||X + Y|^{p-1}\|_q \cdot \|X\|_p + \||X + Y|^{p-1}\|_q \cdot \|Y\|_p \\ &= \|X + Y\|_p^{p-1} \cdot (\|X\|_p + \|Y\|_p). \end{aligned}$$

□

Korollar 6.5. *$\mathcal{L}^p(\Omega, \mathcal{A}, P)$ ist ein Vektorraum* ($\forall p \in [1, \infty]$).

$\|\cdot\|_p$ *ist eine Halbnorm:* $\|\alpha \cdot X\|_p = |\alpha| \cdot \|X\|_p$,
$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$
(aber <u>nicht</u> $\|X\|_p = 0 \Rightarrow X = 0$*)*

Sei $\mathcal{N} = \{X \in \mathcal{L}^p(\Omega, \mathcal{A}, P) : \|X\|_p = 0\}$
$= \{X$ reelle ZV mit $X = 0$ P-f.s. $\}$ unabhängig von $p \in [1, \infty]$.

Definition 6.6. $L^p(\Omega, \mathcal{A}, P) = \mathcal{L}^p(\Omega, \mathcal{A}, P)/\mathcal{N}$
= Menge der Äquivalenzklassen in $\mathcal{L}^p$ unter der Äquivalenzrelation

$$X \sim Y \Leftrightarrow X = Y \ P\text{-}f.s.$$

Korollar 6.7. *$L^p(\Omega, \mathcal{A}, P)$ ist ein normierter Raum mit der Norm $\|\cdot\|_p$ und $L^2(\Omega, \mathcal{A}, P)$ ist ein euklidischer Vektorraum.*

Bemerkung: L^p ist sogar Banachraum (L^2 ist Hilbert-Raum).

Bemerkung:
Die Räume $\mathcal{L}^p$ lassen sich auch für $p \in]0, 1[$ (und sogar für $p \in]-\infty, 0[$) definieren.
Hölder- und Minkowsky-Ungleichung gelten nicht mehr!

Satz 6.8 (Tschebyscheff-Markoff-Ungleichung = Chebyshev-Markov-Ungleichung).

$\forall p > 0, \alpha > 0:$ $$P(|X| \geq \alpha) \leq \frac{1}{\alpha^p} E(|X|^p)$$

Beweis: Sei $A_\alpha = \{|X| \geq \alpha\}$. Dann ist $E(|X|^p) \geq \int\limits_{A_\alpha} |X|^p dP \geq \alpha^p \cdot P(A_\alpha)$ □

Korollar 6.9 (Tschebyscheff-Ungleichung = Chebyshev-Ungleichung).

$\forall \alpha > 0:$ $$P(|X - E(X)| \geq \alpha) \leq \frac{1}{\alpha^2} \cdot varX.$$

Beweis: obige Ungleichung mit $p = 2$ und $X - E(X)$ statt X. □

<u>Also</u>: Die Varianz kontrolliert die Abweichung vom Erwartungswert.
Große Abweichungen sind unwahrscheinlich!

Beispiel 6.10 (Bernoullische Versuchsfolge).

$(\Omega, \mathcal{A}, P) = (\Omega_1, \mathcal{A}_1, P_1)^{\mathbb{N}}$ mit $\mathcal{A}_1 = \{\emptyset, A, A^c, \Omega\}$
$A =$ "Erfolg", $A^c =$ "Mißerfolg" und $P_1(A) = p$.
<u>Typisch</u>: $\Omega_1 = \{0, 1\}$, $A = \{1\}$

$$X_i(\omega) = \mathbb{1}_A(\omega_i) = \begin{cases} 1 & \text{bei Erfolg im i-ten Spiel} \\ 0 & \text{sonst} \end{cases} \qquad (i \in \mathbb{N})$$

$\Rightarrow E(X_i) = p$
$\text{var}(X_i) = E(X_i^2) - p^2 = p - p^2 = p \cdot (1 - p)$

Sei $S_n = \sum\limits_{i=1}^{n} X_i$ "Anzahl der Erfolge in den ersten n Spielen"

$\Rightarrow P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$ Binomial-Verteilung

$$\begin{aligned}
\Rightarrow E(S_n) &= \sum_{k=0}^{n} k \cdot P(S_n = k) \\
&= \sum_{k=1}^{n} k \cdot \binom{n}{k} p^k \cdot (1-p)^{n-k} \\
&= n \cdot p \cdot \sum_{k=1}^{n} \binom{n-1}{k-1} \cdot p^{k-1} \cdot (1-p)^{(n-1)-(k-1)} = n \cdot p \\
\operatorname{var}(S_n) &= \sum_{k=0}^{n} k^2 \cdot P(S_n = k) - (n \cdot p)^2 \\
&= \sum_{k=0}^{n} k \cdot (k-1) \cdot P(S_n = k) + np - (np)^2 \\
&= n \cdot (n-1) \cdot p^2 + np - (np)^2 = n \cdot p(1-p) \\
\sigma(S_n) &\sim \sqrt{n}
\end{aligned}$$

Sei $Y_n = \frac{1}{n} S_n$ "Erfolgshäufigkeit in den ersten n Spielen"

$\Rightarrow \quad E(Y_n) = \frac{1}{n} E(S_n) = p$

$\operatorname{var}(Y_n) = \frac{1}{n^2} \operatorname{var}(S_n) = \frac{1}{n} \cdot p(1-p)$ ($\to 0$ für $n \to \infty$! Also irgendwie deterministisch.)

$\sigma(Y_n) \sim \frac{1}{\sqrt{n}}$

Also $\operatorname{var}(Y_n) \longrightarrow 0$ für $n \to \infty$ und damit $Y_n \longrightarrow p$ in einem gewissen Sinne.
(Offenbar nicht für alle $\omega \in \Omega$. Denn $\omega = (1, 1, \ldots)$ liefert $Y_n(\omega) \equiv 1$ und $\omega = (0, 0, \ldots)$ liefert $Y_n(\omega) \equiv 0$.)

Kapitel 7

Konvergenzsätze

$(\Omega, \mathcal{A}, P)$ W-Raum, X, X_n $(n \in \mathbb{N})$ ZV auf Ω, $1 \leq p < \infty$

a) Konvergenzbegriffe in der Stochastik

Definition 7.1.

(i) $X_n \longrightarrow X$ *P-f.s. bzw.* $X = \lim\limits_{n\to\infty} X_n$ *P-f.s., falls* $\lim\limits_{n\to\infty} X_n$ *P-f.s. existiert und mit* X *übereinstimmt. ("fast sichere Konvergenz").*

(ii) $X_n \longrightarrow X$ *P-stochastisch bzw.* $X = \underset{n \to \infty}{P\text{-}\lim} X_n$*, falls* $\forall \alpha > 0$:

$P(|X_n - X| \geq \alpha) \longrightarrow 0$ *für* $n \to \infty$
("stochastische Konvergenz", "Konvergenz in Wahrscheinlichkeit").

(iii) $X_n \longrightarrow X$ *in* L^p *bzw.* $X = L^p\text{-}\lim X_n$*, falls* $\|X_n - X\|_p \longrightarrow 0$ *für* $n \to \infty$
("L^p-Konvergenz", "Konvergenz im p-ten Mittel" bzw. "Konvergenz im Mittel", falls p=1)

(iv) Weiterer Konvergenzbegriff: $X_n \xrightarrow{\mathrm{d}} X$*, falls* $P_{X_n} \longrightarrow P_X$ *schwach ("Konvergenz in Verteilung", "schwache Konvergenz der Verteilungen").*

Bemerkungen:
- $X_n \longrightarrow X$ in L^p (für ein $p \in [1, \infty[$) $\Rightarrow X_n \longrightarrow X$ in L^1

- P-stochastisch (und P-f.s. und L^p-) Limiten sind P-f.s. eindeutig bestimmt!

b) Punktweise Konvergenz und Konvergenz der Integrale

Satz 7.2 (von der monotonen Konvergenz (Beppo Levi)). *Sei $(X_n)_n$ isotone Folge von ZV $X_n \geq 0$. Dann*

$$E(\sup_n X_n) = \sup_n E(X_n).$$

Bemerkung: Alle $\sup\limits_n$ sind $\lim\limits_{n\to\infty}$.

Beweis: Meßbarkeit bereits gezeigt. Ferner gilt $\forall n$:

$X_n = \sup\limits_m X_{n,m}$ mit isotoner Folge von Elementarfunktionen $X_{n,m} \geq 0$

Setze $X = \sup\limits_n X_n = \sup\limits_{n,m} X_{n.m} \Rightarrow$ (Def. von $E(.)$):

$$E(X_n) \underset{\substack{\uparrow \\ \text{wegen } X_n \leq X}}{\leq} E(X) = \sup_{n,m} E(X_{n,m}) \underset{\substack{\uparrow \\ \text{wegen } X_{n,m} \leq X_n}}{\leq} \sup_n E(X_n).$$

□

Lemma 7.3 (von Fatou). *Sei $(X_n)_n$ beliebige Folge von ZV $X_n \geq 0$. Dann*

$$E(\liminf_{n\to\infty} X_n) \leq \liminf_{n\to\infty} E(X_n).$$

Beweis: Betrachte $Y_n = \inf\limits_{k\geq n} X_k \Rightarrow (Y_n)_n$ isoton $\Rightarrow$

$$E(\liminf X_n) = E(\sup Y_n) = \sup_n E(Y_n) \underset{\substack{\uparrow \\ \text{wegen } X_n \leq X_k \ \forall k \geq n}}{\leq} \sup_n \inf_{k\geq n} E(X_k).$$

□

Im Lemma von Fatou gilt i.a. keine Gleichheit!

Beispiel 7.4. $(\Omega, \mathcal{A}, P) = ([0,1], \mathcal{B}([0,1]), P)$ mit $P(dx) = (2\pi)^{-\frac{1}{2}} \cdot e^{-\frac{x^2}{2}} \cdot \lambda^1(dx)$

$X_n = n \cdot \mathbb{1}_{]n,\infty]} \cdot e^{-\frac{n^2}{2}} \Rightarrow \liminf X_n \equiv 0$

$E(X_n) \geq (2\pi)^{-\frac{1}{2}} \cdot \left(1 - \frac{1}{n^2}\right)$, $E(\liminf X_n) = 0$

$X_n = \infty \cdot \mathbb{1}_{]n,\infty]} \Rightarrow E(X_n) = \infty$, $E(\liminf X_n) = 0$

Beachte: Hier ist $(X_n)_n$ antiton.

Kein Analogon zum Satz von Beppo Levi für antitone Folgen ≥ 0
(bzw. isotone Folgen ≤ 0)!

Satz 7.5 (von der majorisierten Konvergenz (Lebesgue)). *Sei* $p \in [1, \infty[$. *Sei* $X_n \longrightarrow X$ *P-f.s. und*

$$\|X_n\| \leq Y \qquad\qquad (\forall n \in \mathbb{N})$$

für ein Y *mit* $E(|Y|^p) < \infty$.
Dann gilt: $X_n \longrightarrow X$ *in* L^p.

Beweis: Wegen $|X_n| \leq Y$ $(\forall n \in \mathbb{N})$ folgt $|X| \leq Y$ P-f.s. und damit $E(|X|^p) < \infty$.
Sei $Z_n = |X_n - X|^p$.
Zu zeigen: $EZ_n \longrightarrow 0$.

Es gilt: $0 \leq Z_n \leq (|X| + Y)^p =: Z$ mit $E(Z) < \infty$.
Lemma von Fatou für $Z - Z_n \geq 0 \Rightarrow$
$E(\liminf(Z - Z_n)) \leq \liminf E(Z - Z_n) = E(Z) - \limsup E(Z_n)$.
Aus $X_n \longrightarrow X$ P-f.s. folgt $Z_n \longrightarrow 0$ P-f.s.. Also $E(\liminf(Z - Z_n)) = E(Z)$.
$\Rightarrow \limsup E(Z_n) \leq 0 \Rightarrow$ (wegen $Z_n \geq 0$): $\lim E(Z_n) = 0$. $\square$

Bemerkungen:

(1) Falls die X_n gleichmäßig dominiert werden in L^p, so gilt:

$$P\text{-f.s. - Konvergenz} \Rightarrow L^p\text{-Konvergenz}$$

(2) Im Satz von Lebesgue genügt: X_n konvergiert P-f.s. $\ldots \Rightarrow \exists X \in \mathcal{L}^p$ mit $X_n \longrightarrow X$ P-f.s. und $\ldots$.

(3) Aus $X_n \longrightarrow X$ in L^p folgt $\|X_n\|_p \longrightarrow \|X\|_p$.

c) Konvergenz in L^p

Satz 7.6. *Sei* $(X_n)_n$ *Cauchy-Folge in* $\mathcal{L}^p$ $(1 \leq p < \infty)$. *Dann* $\exists X \in \mathcal{L}^p$ *mit:*

(i) $X_n \longrightarrow X$ *in* L^p

(ii) $X_n \longrightarrow X$ *P-stochastisch*

(iii) $X_{n_k} \longrightarrow X$ *P-f.s. für geeignete Teilfolge* $(n_k)_k$

Beweis:

zu (iii): $(X_n)_n$ Cauchy in $\mathcal{L}^p \Rightarrow \exists (n_k)_k$ mit $\|X_{n_{k+1}} - X_{n_k}\|_p \le 2^{-k}$.

Sei $Y_k = X_{n_{k+1}} - X_{n_k}$ und $Y = \sum_{k=1}^{\infty} |Y_k|$

$$\Rightarrow \|Y\|_p \underset{\substack{\uparrow \\ \text{monotone Konvergenz}}}{=} \sup_l \| \sum_{k=1}^{l} |Y_k| \|_p \le \sum_{k=1}^{\infty} \|Y_k\|_p \le 1$$

$\Rightarrow Y < \infty$ P-f.ü., $\sum_{k=1}^{l} Y_k$ konv. P-f.s.

$(X_{n_l})_l$ konv. P-f.s.

Ferner: $|X_{n_{l+1}}| \le Y + |X_{n_1}| \in \mathcal{L}^p$

Majorisierte Konvergenz $\Rightarrow \exists X \in \mathcal{L}^p$ mit $X_{n_l} \longrightarrow X$ P-f.s. und in L^p (also (iii)).

zu (i): $(X_n)_n$ Cauchy in $\mathcal{L}^p \Rightarrow \forall \epsilon > 0 : \exists m_\epsilon : \|X_m - X_n\| \le \epsilon$ $(\forall n, m \ge m_\epsilon)$

$X_{n_k} \longrightarrow X$ in $\mathcal{L}^p \Rightarrow \exists k : \|X_{n_k} - X\|_p \le \epsilon$ und $n_k \ge m_\epsilon$.

Beides zusammen: $\|X_n - X\|_p \le \|X_n - X_{n_k}\|_p + \|X_{n_k} - X\|_p \le 2\epsilon$.

zu (ii): Chebyshev-Markov: $\forall \alpha > 0$:

$P(|X_n - X| \ge \alpha) \le \alpha^{-p} \cdot \|X_n - X\|_p^p \longrightarrow 0$ für $n \to \infty$. $\square$

Korollar 7.7 (Riesz-Fischer). $\mathcal{L}^p(\Omega, \mathcal{A}, P)$ *ist vollständig* $(1 \le p < \infty)$. *(Ebenso* $L^p(\Omega, \mathcal{A}, P)$.*)*

Beachte: (iii) in obigen Satz gilt i.a. nur für Teilfolgen (s.Übung)!

d) Fast sichere Konvergenz und Stochastische Konvergenz

Satz 7.8.

(i) $X_n \longrightarrow X$ *P-f.s.* $\Longrightarrow X_n \longrightarrow X$ *P-stoch.*

(ii) $X_n \longrightarrow X$ *P-stoch.* $\Longrightarrow \exists$ *Teilfolge* $X_{n_k} \longrightarrow X$ *P-f.s.*

Beweis:

(i) $X_n \longrightarrow X$ P-f.s.

$\Leftrightarrow 1 = P(\{\omega : X_n(\omega) \longrightarrow X(\omega)\}) = P(\bigcap_k \bigcup_m \bigcap_{n \ge m} \{\omega : |X_n(\omega) - X(\omega)| \le \frac{1}{k}\})$

$\Leftrightarrow 1 = P(\bigcup_m \ldots) \qquad \forall k$

$$\Rightarrow \forall k: \ 1 = \lim_{m\to\infty} \underbrace{P(\bigcap_{n \ge m} \{\omega : |X_n(\omega) - X(\omega)| \le \frac{1}{k}\})}_{\le}$$

$$= \lim_{m\to\infty} \overbrace{P(\{\omega : |X_m(\omega) - X(\omega)| \le \frac{1}{k}\})} \Rightarrow \text{ stoch. Konv.}$$

(ii) $X_n \longrightarrow X$ P-stoch. $\Leftrightarrow \forall \epsilon > 0 : P(|X_n - X|) \geq \epsilon) \longrightarrow 0$
Insbesondere: $\forall k \in \mathbb{N} : \exists n_k > n_{k-1} : P(|X_{n_k} - X| \geq \frac{1}{k}) \leq 2^{-k}$
$\Rightarrow \sum\limits_{k=1}^{\infty} P(|X_{n_k} - X| \geq \frac{1}{k}) < \infty$
Mit Borel-Cantelli-Lemma (s. unten):
$\Rightarrow P(\{\omega : |X_{n_k}(\omega) - X(\omega)| \geq \frac{1}{k}$ für unendlich viele $k \in \mathbb{N}\}) = 0$
$\Rightarrow P(\{\omega : X_{n_k}(\omega) \longrightarrow X(\omega)$ für $k \to \infty\}) = 1$. □

Korollar 7.9. *$X_n \longrightarrow X$ P-stoch. $\Longleftrightarrow \forall$ Teilfolgen $(X_{n_k}) : \exists$ Teilfolgen $(X_{n_{k_l}})$ mit $X_{n_{k_l}} \longrightarrow X$ P-f.s.*

Zur Formulierung des Borel-Cantelli-Lemmas betrachte $A_n \in \mathcal{A}$ und definiere

$$\limsup_{n\to\infty} A_n := \bigcap_m \bigcup_{n \geq m} A_n = \{\omega : \omega \in A_n \text{ für unendlich viele } n \in \mathbb{N}\}$$

Schreibweise: $P(\limsup A_n) = P(A_n$ unendlich oft$)$.

Lemma 7.10 (Borel-Cantelli I). *$A_n \in \mathcal{A}$. Dann*

$$\sum_{n=1}^{\infty} P(A_n) < \infty \Longrightarrow P(\limsup_{n\to\infty} A_n) = 0$$

Beweis:
Sei $N(\omega) = \sum\limits_n \mathbb{1}_{A_n}(\omega)$ Anzahl der $n \in \mathbb{N}$ mit $\omega \in A_n$,
also $\limsup\limits_{n\to\infty} A_n = \{\omega : N(\omega) = \infty\}$.
Nach Fubini: $E(N) = \sum\limits_n P(A_n) < \infty$.
Also $N < \infty$ P-f.s.. □

Beispiel 7.11. $n = c \cdot r$ Kugeln zufällig in r Urnen $F_1, F_2, \ldots, F_r$ legen
K_r Anzahl der leeren Urnen
$\Rightarrow E(K_r) = r \cdot \left(1 - \frac{1}{r}\right)^{c\cdot r}$ und $E\left(\frac{K_r}{r}\right) \longrightarrow e^{-c}$ für $r \to \infty$.
Beh.: $\frac{K_r}{r} \longrightarrow e^{-c}$ P-stochastisch

Beweis: Berechne $\mathrm{var}(K_r) = E(K_r^2) - (EK_r)^2$
Sei wieder $A_i = \{$Urne F_i ist leer$\}$.

$$
\begin{aligned}
\text{Nun ist}\quad E(K_r^2) &= E\left[\left(\sum_{i=1}^r \mathbb{1}_{A_i}\right)^2\right] = \sum_{i,j} P(A_i \cap A_j)\\
&= \sum_i P(A_i) + \sum_{i\neq j} P(A_i \cap A_j)\\
&= r\cdot\left(1-\tfrac{1}{r}\right)^n + r\cdot(r-1)\cdot\left(1-\tfrac{2}{r}\right)^n
\end{aligned}
$$

$$
\operatorname{var}\left(\tfrac{K_r}{r}\right) = \tfrac{1}{r}\cdot\left(1-\tfrac{1}{r}\right)^n + \tfrac{r-1}{r}\cdot\left(1-\tfrac{2}{r}\right)^{r\cdot c} - \left(1-\tfrac{1}{r}\right)^{2\cdot r\cdot c}
\longrightarrow 0 + 1\cdot e^{-2c} - e^{-2c} = 0
$$

Also mit Chebyshev:

$$
\begin{aligned}
P\left(\left|\tfrac{K_r}{r} - e^{-c}\right| \geq \epsilon\right) &\leq \tfrac{1}{\epsilon^2}\cdot E\left(\left|\tfrac{K_r}{r} - e^{-c}\right|^2\right)\\
&\leq \tfrac{2}{\epsilon^2}(\underbrace{E\left(\left|\frac{K_r}{r} - E\left(\frac{K_r}{r}\right)\right|^2\right)}_{=\operatorname{var}\frac{K_r}{r}\longrightarrow 0} + \underbrace{\left|E\left(\frac{K_r}{r}\right) - e^{-c}\right|^2}_{\longrightarrow 0})\\
&\longrightarrow 0.
\end{aligned}
$$

□

e) Gleichgradige Integrierbarkeit und L^p-Konvergenz

Definition 7.12. *$(X_n)_n$ heißt* **gleichgradig integrierbar**, *falls*

$$
\underbrace{\sup_n E(|X_i|; |X_i| > M)}_{=\overbrace{\int_{\{|X_i|>M\}} |X_i| dP}} \longrightarrow 0 \text{ für } M \to \infty.
$$

Proposition 7.13. *Aus jeder der folgenden Bedingungen folgt die gleichgradige Integrierbarkeit von $(X_n)_n$:*

(i) $\forall n : |X_n| \leq Y$ *mit* $Y \in \mathcal{L}^1$ *(Majoranten-Bed.)*

(ii) $\sup_n E(|X_n| \cdot \log_+ |X_n|) < \infty$ *(Entropie-Bed.)*

(iii) $\sup_n E(|X_n|^p) < \infty$ *für ein* $p > 1$ *(Glm L^p-Beschränktheit)*

(iv) $\sup_n E(g(|X_n|)) < \infty$ *für ein* $g : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ *mit* $\frac{g(x)}{x} \longrightarrow \infty$ *für* $x \to \infty$.

Beweis:

(i) $\sup_n \int_{\{|X_n|>M\}} |X_n| dP \leq \int_{\{Y>M\}} Y dP = \int \underbrace{Y \cdot \mathbb{1}_{\{Y>M\}}}_{\substack{\longrightarrow 0\ P\text{-f.s.}\\ \text{für } M\to\infty}} dP \overset{\text{Lebesgue mit Majorante } Y}{\longrightarrow} 0.$

(ii),(iii) $\Leftarrow$ (iv).

(iv) Wegen $\frac{g(x)}{x} \longrightarrow \infty$ für $x \to \infty$ gilt:
$\forall \epsilon > 0 : \exists M = M_\epsilon : \forall x \geq M : \frac{x}{g(x)} \leq \epsilon.$

$$\Rightarrow \int\limits_{\{|X_n| \geq M\}} |X_n| dP = \int\limits_{\{|X_n| \geq M\}} g(|X_n|) \cdot \frac{|X_n|}{g(|X_n|)} dP \leq \epsilon \cdot \int\limits_{\{|X_n| \geq M\}} g(|X_n|) dP \leq \epsilon \cdot \underset{\substack{\uparrow \\ \text{gleichmäßig in } n \text{ (!)}}}{G}$$

mit $G = \sup_n \int g(|X_n|) dP < \infty$. □

Proposition 7.14. *Äquivalent sind:*
(i) $(X_n)_n$ *gleichgradig integrierbar*

(ii) $\sup_n E(|X_n|) < \infty$ *__und__* $\forall \epsilon > 0 : \exists \delta > 0 : \forall A \in \mathcal{A}$*:*

$$P(A) < \delta \Longrightarrow \int_A |X_n| dP < \epsilon \qquad (\forall n)$$

Beweis: (Übung). □

Satz 7.15. *Sei* $(X_n)_n \subset \mathcal{L}^p$ *und* $X_n \longrightarrow X$ *P-stoch.* $(1 \leq p < \infty)$.
Dann sind äquivalent:
(i) $(X_n^p)_n$ *gleichgradig integrierbar*

(ii) $X_n \longrightarrow X$ *in* L^p

(iii) $E(|X_n|^p) \longrightarrow E(|X|^p) < \infty.$!

Beweis:

(i)$\Rightarrow$(ii): $X_n \longrightarrow X$ P-stoch. $\Rightarrow$ $\exists$ Teilfolge mit $X_{n_k} \longrightarrow X$ P-f.s.
Nach Fatou: $E(|X|^p) = E(\lim_k |X_{n_k}|^p) \leq \liminf_k E(|X_{n_k}|^p)$
$\leq \sup_n E(|X_n|^p) < \infty$ nach Proposition
Aus $|X_n - X|^p \leq 2^p \cdot (|X_n|^p + |X|^p)$ folgt die gleichgradige Integrierbarkeit von $(|X_n - X|^p)_n$.
Denn: Für $\epsilon > 0$ gilt:

$$\begin{aligned}
\int |X_n - X|^p dP &= \int\limits_{\{|X_n-X|>\epsilon\}} |X_n - X|^p dP + \int\limits_{\{|X_n-X|\leq\epsilon\}} |X_n - X|^p dP \\
&\leq \int\limits_{\{|X_n-X|>\epsilon\}} |X_n - X|^p dP + \epsilon^p.
\end{aligned}$$

Wegen $X_n \longrightarrow X$ P-stochastisch, ist $P(|X_n - X| > \epsilon) \longrightarrow 0$ für $n \to \infty$ und damit (wegen (ii) in obiger Proposition):

$$\int |X_n - X|^p dP \overset{n\to\infty}{\longrightarrow} 0.$$

(ii)$\Rightarrow$(iii): (Minkowski-Ungleichung) trivial

(iii)$\Rightarrow$(i): Sei $\psi_M \in \mathcal{C}(\mathbb{R}_+)$

mit $\psi_M(x) = \begin{cases} x, & \text{in } [0, M-1] \\ 0, & \text{in } [M, \infty] \\ \text{linear}, & \text{in } [M-1, M]. \end{cases}$

Für $\epsilon > 0$ und M hinreichend groß ist $E|X|^p - E\psi_M(|X|^p) \leq \epsilon$.

Aus dem Lebesgue-Konvergenzsatz (mit Majorante M) folgt

$$E\psi_M(|X_n|^p) \longrightarrow E\psi_M(|X|^p).$$

Mit (iii) folgt daraus

$$\begin{aligned}
E(|X_n|^p; |X_n| \geq M) &\leq E(|X_n|^p) - E\psi_M(|X_n|^p) \\
&\longrightarrow E(|X|^p) - E\psi_M(|X|^p) \leq \epsilon
\end{aligned}$$

d.h. $\exists n_0 : \sup\limits_{n\geq n_0} E(\ldots) \leq 2\epsilon$

Wähle M^* mit $\sup\limits_{n<n_0} E(|X_n|^p; |X_n| \geq M^*) \leq 2\epsilon$ und $\bar{M} = \sup\{M, M^*\}$.

$\Rightarrow \sup\limits_n E(|X_n|^p; |X_n| \geq \bar{M}) \leq 2\epsilon.\square$

f) Schwache Konvergenz

Geg.: E top. Raum mit $\mathcal{B} = \mathcal{B}(E)$ Borel-σ-Algebra, W-Maße ν_n ($n \in \mathbb{N}$), ν auf $(E, \mathcal{B})$.

Ges.: Vernünftiger Begriff für $\nu_n \to \nu$

Beispiel: $\Leftrightarrow \nu_n(A) \to \nu(A)$ ($\forall A \in \mathcal{B}$) (Punktweise, stark)

Zu restriktiv! z.B. $x_n \to x \not\Rightarrow \delta_{x_n}(A) \to \delta_x(A)$ ($\forall A \in \mathcal{B}$)

Definition 7.16. *$(\nu_n)_n$ konvergiert* **schwach** *gegen ν (in Zeichen: $\nu_n \to \nu$ schwach oder $\nu_n \overset{w}{\to} \nu$; gelegentlich $\nu_n \rightharpoonup \nu$ oder $\nu_n \Rightarrow \nu$), falls*

$$\int f d\nu_n \to \int f d\nu \quad (\forall f \in \mathcal{C}_b(E)).$$

Satz 7.17 (Portmonteaū). *Sei E metrisch. Dann sind die folgenden Aussagen äquivalent:*

(i) $\nu_n \overset{w}{\to} \nu$

(ii) $\int f d\nu_n \to \int f d\nu \quad \forall f$ *beschränkt, gleichmäßig stetig*

(iii) $\limsup \nu_n(A) \le \nu(A) \quad \forall A$ *abgeschlossen*

(iv) $\liminf \nu_n(A) \ge \nu(A) \quad \forall A$ *offen*

(v) $\lim \nu_n(A) = \nu(A) \quad \forall A$ *mit* $\nu(\partial A) = 0$ *("ν-randlos")*

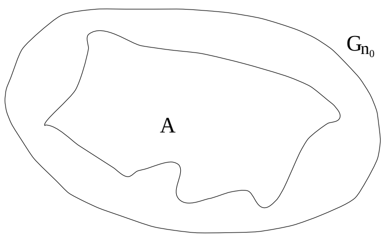
(vi) $\int f d\nu_n \to \int f d\nu \quad \forall f$ *beschränkt, meßbar und ν-f.ü. stetig*
(d.h. $\nu(\{x : f$ *unstetig in* $x\}) = 0$*)*

Beweis:

(i)$\Rightarrow$(ii) klar, ebenso (vi)$\Rightarrow$(i)

(iv)$\Rightarrow$(i) Ersetze f durch $f + C \Rightarrow$ o.B.d.A. $f \ge 0$.
$\liminf \int f d\nu_n = \liminf \int \nu_n\{f > \lambda\} d\lambda \overset{\text{Fatou}}{\ge} \int \liminf \nu_n\{f > \lambda\} d\lambda \ge \int \nu(f > \lambda) d\lambda = \int f d\nu$
Analog: $\liminf \int (-f) d\nu_n \ge \int (-f) d\nu \Rightarrow$ Behauptung

(ii)$\Rightarrow$(iii) Sei A abgeschlossen.
$\Rightarrow A = \bigcap\limits_n G_n$ mit $G_n = \{x : d(x, A) < \frac{1}{n}\}$ offen
$\Rightarrow \nu(A) = \lim\limits_n \nu(G_n)$, d.h. $\forall \epsilon > 0 : \exists n_0 : \nu(G_{n_0}) \le \nu(A) + \epsilon$.
Wähle f beschränkt, ≥ 0 und gleichmäßig stetig mit $f = 0$ auf $G^c_{n_0}$ und $f = 1$ auf A.



$$\Rightarrow \limsup_n \nu_n(A) \le \limsup \int f d\nu_n = \int f d\nu \le \nu(G_{n_0}) \le \nu(A) + \epsilon.$$

(iii)$\Leftrightarrow$(iv) Komplementbildung.

(iv)$\Rightarrow$(v) $\nu(A) = \nu(\mathring{A}) \le \liminf \nu_n(\mathring{A}) \le \liminf \nu_n(A)$ und
$\nu(A) = \nu(\bar{A}) \ge \limsup \nu_n(\bar{A}) \ge \limsup \nu_n(A)$.

(v)⇒(vi) f beschränkt $\Rightarrow \forall\epsilon > 0 : \exists c_0 < c_1 < \ldots < c_l$ mit $c_0 \le f \le c_l$ auf Ω, $|c_k - c_{k-1}| \le \epsilon$ und

$$\nu(\{x : f(x) = c_k\}) = 0$$

f ν-f.ü. stetig $\Rightarrow A_k := \{x : c_k < f(x) < c_{k+1}\}$ ist ν-randlos.

Sei nun $g = \sum_{k=0}^{l-1} c_k \cdot \mathbb{1}_{A_k} \Rightarrow |g - f| \le \epsilon$,

$$\begin{aligned} \left|\int f d\nu_n - \int f d\nu\right| &\le 2\epsilon + \left|\int g d\nu_n - \int g d\nu\right| \\ &\le 2\epsilon + \sum_{k=0}^{l-1} c_k \cdot \big(\underbrace{\nu_n(A_k) - \nu(A_k)}_{\to 0 \text{ für } n \to \infty}\big) \\ &\to 2\epsilon \end{aligned}$$

□

Korollar 7.18. *Für W-Maße auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ sind äquivalent*
(i) $\nu_n \to \nu$ schwach

(ii) $F_n(x) \to F(x)$ in allen Stetigkeitsstellen x von F (wobei $F_n(x) = \nu_n(]-\infty, x])$ und $F(x) = \nu(]-\infty, x])$).

Beweis:

(i)⇒(ii) Ist F stetig in x, so ist $]-\infty, x]$ eine ν-randlose Menge. Also: Beh. folgt aus (i)⇒(v) im Satz.

(ii)⇒(i) Sei $X_n(\omega) = \sup\{x : F_n(x) < \omega\}$ und $X(\omega) = \sup\{x : F(x) < \omega\}$
$\Rightarrow \lambda^1_{X_n} = \nu_n, \lambda^1_X = \nu$ mit $\lambda^1 = \lambda^1_{|]0,1[}$
(denn $\lambda^1_{X_n}(]-\infty, x]) = \lambda^1(X_n^{-1}(]-\infty, x])) = \lambda^1([0, F(x)]) = F(x)$)
X_n, X sind ZV auf $([0,1], \mathcal{B}([0,1]))$.

a) Behauptung: $X_n(\omega) \to X(\omega)$ für alle $\omega \in [0,1] \setminus N$ mit

$$\begin{aligned} N &= \{\omega :]a_\omega, b_\omega[\neq \emptyset\} \text{ “Unstetigkeitsstellen von } X\text{“} \\ a_\omega &= \sup\{x : F(x) < \omega\} = X(\omega) \\ b_\omega &= \inf\{x : F(x) > \omega\} = X(\omega_+) \end{aligned}$$

N ist abzählbar, denn die $]a_\omega, b_\omega[$ sind disjunkt.

z.z. (1) $\qquad \liminf_{n\to\infty} X_n(\omega) \ge X(\omega) \quad \forall \omega \notin N$

(denn: für $x < X(\omega)$ mit F stetig in x gilt: $F(x) < \omega$
und damit für hinreichend große n: $F_n(x) < \omega$.

D.h. $X_n(\omega) \geq x$ und damit (wegen x beliebig) $X_n(\omega) \geq X(\omega)$.)

$$\text{z.z. (2)} \qquad \limsup X_n(\omega) \leq X(\omega) \quad \forall \omega \notin N$$

(denn für $x > X(\omega)$ mit F stetig in x gilt: $F(x) > \omega$.
$\Rightarrow F_n(x) > \omega \Rightarrow X_n(\omega) \leq x \Rightarrow X_n(\omega) \leq X(\omega)$.)

b) Behauptung: $X_n \to X$ λ^1-f.s., denn $\lambda^1(N) = 0$.

c) Behauptung: $\int g(X_n) d\lambda^1 \to \int g(X) d\lambda^1$ $\forall g \in \mathcal{C}_b(\mathbb{R})$,
denn $g(X_n) \to g(X)$ λ^1-f.s. und $g(X_n) \leq M \Rightarrow$ Behauptung wegen Satz von majorisierten Konvergenz (für $\lambda^1 = \lambda_{|]0,1[}$)

d) Behauptung: $\nu_n \to \nu$ schwach, denn

$$\int g d\nu_n = \int g d\left(\lambda^1_{X_n}\right) = \int g(X_n) d\lambda^1 \to \int g(X) d\lambda^1 = \int g d\nu \quad \forall g \in \mathcal{C}_b(\mathbb{R}). \square$$

g) Konvergenz in Verteilung

Geg. $(\Omega, \mathcal{A}, P)$, ZV $(X_n)_n, X$.

Definition 7.19. *$X_n \to X$* ***in Verteilung****, falls $P_{X_n} \to P_X$ schwach. (oder $X_n \Rightarrow X$, $X_n \stackrel{\mathrm{d}}{\to} X$)*

Achtung: Aus $X_n \stackrel{\mathrm{d}}{\to} X$ und $Y_n \stackrel{\mathrm{d}}{\to} Y$ folgt nicht $X_n + Y_n \to X + Y$.

Satz 7.20. *$X_n \to X$ P-stochastisch $\Rightarrow X_n \to X$ in Verteilung. (Falls X konstant ist, gilt auch die Umkehrung.)*

Beweis: Sei f beschränkt, gleichmäßig stetig

$$\begin{aligned} \Rightarrow \left| \int f dP_{X_n} - \int f dP_X \right| &= |E(f(X_n) - f(X))| \\ &\leq E(|f(X_n) - f(X)|; d(X_n, X) \leq \delta) + E(|\ldots|; d(X_n, X) > \delta) \\ &= (*) \end{aligned}$$

f gleichmäßig stetig $\Rightarrow \forall \epsilon : \exists \delta : |f(x) - f(y)| \leq \epsilon$, falls $d(x,y) \leq \delta$
$\Rightarrow (*) \leq \epsilon + 2 \cdot \|f\|_\infty \cdot \underbrace{P(d(X_n, X) > \delta)}_{\to 0 \text{ für } n \to \infty \text{ (stochastische Konvergenz)}}$ Umkehrung

siehe unten!

Bemerkung: $X_n \to X$ in Verteilung $\Rightarrow$

$$E(f(X_n)) \to E(f(X)) \qquad \forall f \in \mathcal{C}_b(\mathbb{R}).$$

Zum Beweis von " $\Leftarrow$ " in obigem Satz:
Sei $P_{X_n} \to P_X = \delta_\alpha$ mit $\alpha \in \mathbb{R}$.
Für $\epsilon > 0$ wähle $f, g \in \mathcal{C}_b(\mathbb{R})$ mit $f \leq \mathbb{1}_{]\alpha-\epsilon,\alpha+\epsilon[} \leq g$ und $f(\alpha) = 1 = g(\alpha)$. Dann ist

$$\begin{array}{ccc} \int f dP_{X_n} \leq & P(X_n \in]\alpha-\epsilon, \alpha+\epsilon[) & \leq \int g dP_{X_n} \\ \downarrow & & \downarrow \\ \int f dP_X = & f(\alpha) = 1 = g(\alpha) & = \int g dP_X \end{array}$$

d.h. $P(X_n \in]\alpha-\epsilon, \alpha+\epsilon[) \to 1$ für $n \to \infty$.
$\quad = P(|X_n - \alpha| < \epsilon) = P(|X_n - X| < \epsilon)$
m.a.W. $P(|X_n - X| \geq \epsilon) \to 0$ für $n \to \infty$ $(\forall \epsilon > 0)$. □

Beispiel 7.21 (Zufällige Permutation).
etwa vertauschen von Hüten, Männern (im Tanzkurs), ...
$\Omega = \{\text{Permutation von } \{1, \ldots, N\}\}$
K_N = Anzahl der Fixpunkte $\Rightarrow$ (Übung; Serie 1, Aufgabe 2):

$$P(K_N = k) = \binom{N}{k} N^{-k} \cdot \sum_{m=0}^{N-k} \frac{(-1)^m}{m!} \to \frac{1}{k!} \cdot e^{-1} \quad \text{für } N \to \infty$$

("Poisson-Verteilung" mit Parameter 1).

<u>Beachte</u>: $E(K_N) = 1 \; \forall N$,
denn sei $A_i = \{i \text{ ist Fixpunkt }\} \Rightarrow P(A_i) = \frac{1}{N}$
und $E(K_N) = E\left(\sum_{i=1}^{N} \mathbb{1}_{A_i}\right) = N \cdot P(A_i) = 1.$
Unabhängig von der Größe der Verwirrung!

Beispiel 7.22 (Warten auf seltene Ereignisse).
$(\Omega, \mathcal{A}, P)$ Bernoulli-Versuchsfolge mit Erfolgswahrscheinlichkeit $p > 0$
T = Wartezeit bis zum ersten Erfolg = Anzahl der Spiele = K
$\Rightarrow P(T = k) = p(1-p)^{k-1} \quad (k \in \mathbb{N})$
$\Rightarrow P(T \geq x) = (1-p)^{[x]-1} \qquad (x \in \mathbb{R})$ ("geometrisch verteilt")

Statt 1 Versuch pro Zeiteinheit mit Erfolgswahrscheinlichkeit p machen wir nun n Versuche pro Zeiteinheit mit Erfolgswahrscheinlichkeit $p_n = \frac{p}{n}$.
Sei T_n = Wartezeit bis zum 1.Erfolg

$= \frac{1}{n}$ (Anzahl der Spiele bis zum 1.Erfolg) $= \frac{1}{n} \cdot K_n$ bei Spiel mit Erfolgswahrscheinlichkeit $p_n = \frac{p}{n}$

$\Rightarrow P(T_n \geq x) = P(K_n \geq x \cdot n) = \left(1 - \frac{p}{n}\right)^{[x \cdot n]-1} \to e^{-p \cdot x}$ für $n \to \infty$
("exponential-verteilt")
D.h. die Verteilung von T_n konv. schwach gegen die Exponentialverteilung mit Parameter p.
Konkret:

- T = Anzahl der jährlichen Zahnarztbesuche bis zum 1. Entdecken eines Loches
  K_{12} = Anzahl der monatlichen ..., K_{365} = Anzahl der täglichen ...

- T = Anzahl der Messungen bis ein radioaktiver Zerfall festgestellt wird

Beispiel 7.23 (Trefferanzahl bei geringer Trefferwahrscheinlichkeit). Sei X_n = Anzahl der Erfolge in den ersten n Spielen bei Spiel mit Erfolgsw. $p_n = \frac{\lambda}{n}$ ($\Rightarrow E(X_n) = n \cdot p_n = p$, d.h. insbesondere X_n ist "klein")

$$\begin{aligned}
\Rightarrow P(X_n = k) &= \binom{n}{k} \cdot p_n^k \cdot (1-p_n)^{n-k} \\
&= \frac{n!}{k!(n-k)!} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \cdot \underbrace{\frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{n^k}}_{\to 1} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\to 1} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{n}}_{\to e^{-p}} \\
&\to \frac{\lambda^k}{k!} \cdot e^{-\lambda} \qquad \text{("Poisson-Verteilung mit Parameter } \lambda\text{")}
\end{aligned}$$

"Poisson-Approximation der Binomialverteilung"
D.h. Binomialverteilung mit Erfolgswahrscheinlichkeit $\frac{\lambda}{n}$ konvergiert schwach gegen Poisson-Verteilung mit Parameter λ.

z.B.

- Tote durch Hufschlag in Kavallerie-Regiment (n = Anzahl der Hufschläge)

- Anzahl der Kunden, die zwischen 13^{40} Uhr und 13^{55} Uhr bei einer Bank eintreffen
  (n = Anzahl der Hufschläge)

- Anzahl der Babies, die am 22.11. geboren werden (n = Frauen)

- Anzahl der Verkehrsunfälle an einem Tag ($n =$ Autofahrer)

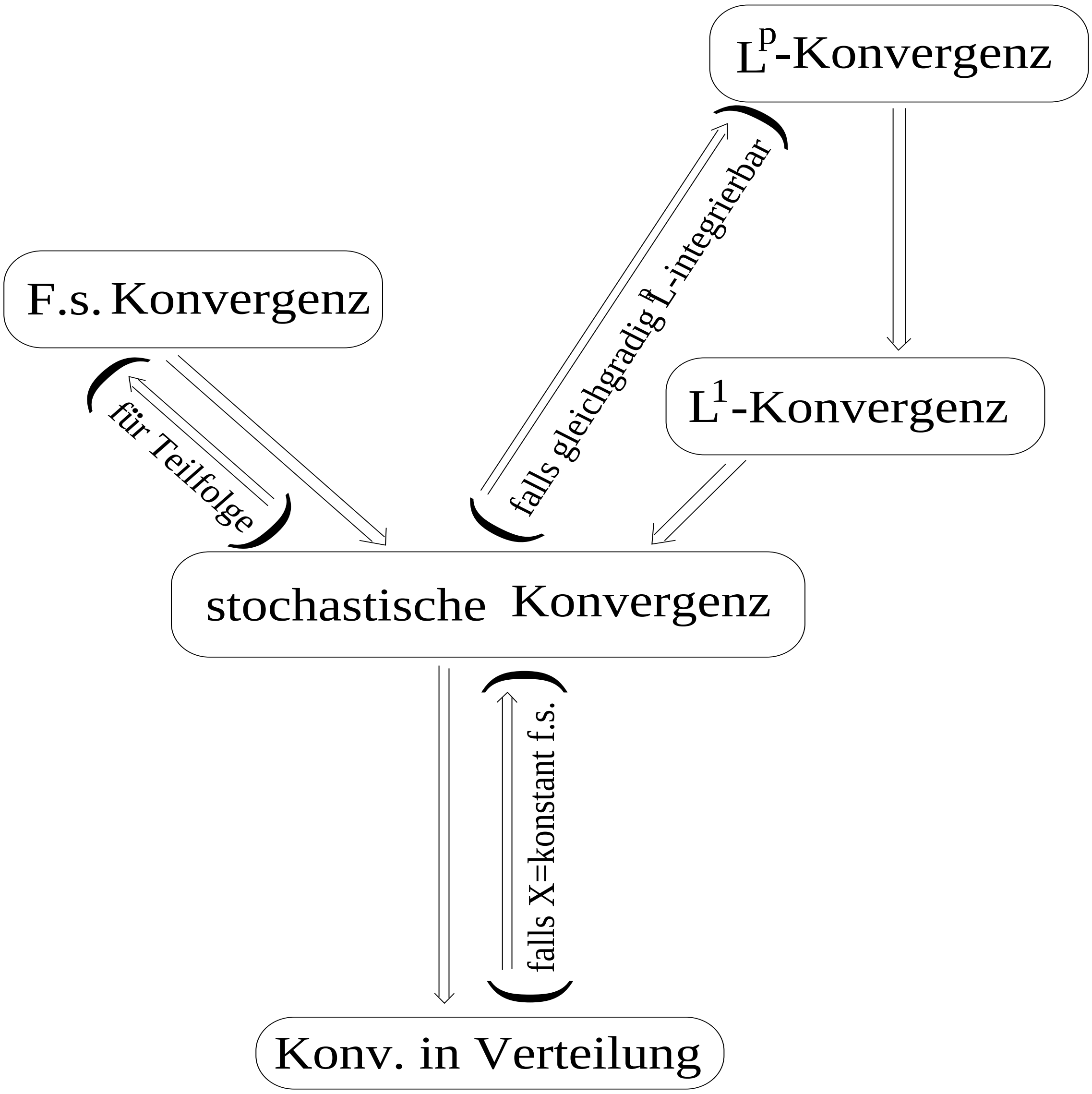
Beispiele 7.24 (für schwache Konvergenz / Konvergenz in Verteilung).

1. $T_n =$ Wartezeit bis zum 1.Erfolg bei Spiel mit Erfolgswahrscheinlichkeit $p_n = \frac{p}{n}$.
   Geometrische Verteilung:
   $$P(T_n \geq x) = \left(1 - \frac{p}{n}\right)^{[x \cdot n]-1}$$
   $F_n(x) = P_{T_n}(]-\infty, x]) = 1 - P(T_n > x) = \dots$
   $\stackrel{n\to\infty}{\longrightarrow} F(x) = 1 - e^{-px}$ Exponentialverteilung

2. $X_n =$ Anzahl der Erfolge in den ersten n Spielen bei Spiel mit Erfolgswahrscheinlichkeit $p_n = \frac{p}{n}$
   Binomialverteilung
   $P_{X_n}(\{k\}) = P(X_n = k) = \binom{n}{k}\left(\frac{p}{n}\right)^k \cdot \left(1 - \frac{p}{n}\right)^{n-k} \qquad (k = 0, 1, \dots, n)$
   $\stackrel{n\to\infty}{\longrightarrow} P(\{k\}) = \frac{p^k}{k!}e^{-p} \qquad (k \in \mathbb{N}_0)$
   Poisson-Verteilung

3. $X_n =$ Anzahl der Erfolge in den ersten n Spielen bei Erfolgswahrscheinlichkeit $\frac{1}{2}$
   Binomialverteilung
   $$P(X_n = k) = \binom{n}{k} 2^{-n}$$
   Standardisierung
   $$Y_n = \left(X_n - \frac{n}{2}\right)\frac{1}{\sqrt{\frac{n}{2}}}$$
   De Moivre / Laplace
   $$P_{Y_n} \longrightarrow N(0,1) = \text{ Standard-Normalverteilung.}$$

h) Zusammenfassung



Teil III

Unabhängigkeit

Kapitel 8

Unabhängigkeit

Fundamentaler Begriff der W-Theorie
("W-Theorie=Maßtheorie+Unabhängigkeit")

Gegeben stets $(\Omega, \mathcal{A}, P)$ W-Raum

Definition 8.1. *Ereignisse $A, B \in \mathcal{A}$ heißen* ***unabhängig voneinander****, falls*

$$P(A \cap B) = P(A) \cdot P(B).$$

X, Y heißen unabhängig voneinander, falls $\forall A, B \in \mathcal{B}(\mathbb{R})$:

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B).$$

Beispiele 8.2.

- $B \in \mathcal{A}$ mit $P(B) = 0$:
  jedes $A \in \mathcal{A}$ ist unabhängig von B
  (denn $P(B) = 0 \Rightarrow P(A \cap B) = 0 \Rightarrow P(A \cap B) = P(A) \cdot P(B)$)

- $B \in \mathcal{A}$ mit $P(B) > 0$:
  $A \in \mathcal{A}$ ist unabhängig von $B \Leftrightarrow P(A) = P(A|B)$
  (denn $\frac{P(A \cap B)}{P(A) \cdot P(B)} = \frac{1}{P(A)} \cdot P(A|B)$.)

- A und B unabhängig $\Leftrightarrow$ A und B^c unabhängig $\Leftrightarrow \ldots$
  (denn $P(A \cap B^c) = P(A) - P(A \cap B) \underset{\substack{\uparrow \\ \text{unabhängig}}}{=} P(A) - P(A) \cdot P(B) = P(A) \cdot P(B^c)$.)

- $\Omega = \Omega_1 \times \Omega_2, \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2, P = P_1 \otimes P_2 \Rightarrow \forall A_1 \in \mathcal{A}_1$ und $A_2 \in \mathcal{A}_2$ sind die Mengen $A = A_1 \times \Omega_2$ und $B = \Omega_1 \times A_2$ unabhängig
  (denn $P(A) = P_1(A_1), P(B) = P_2(A_2)$ und $P(A \cap B) = P(A_1 \times A_2) = P_1(A_1) \cdot P_2(A_2)$)

- Konkret: $2\times$ Würfeln $\Omega = \{1, \ldots, 6\}^2$
  $A =$"beim 1.Wurf eine 6"
  $B =$"beim 2.Wurf eine 6"
  $P(A) = \frac{1}{6}, P(B) = \frac{1}{6}, P(A \cap B) = \frac{1}{36}$.
  Bemerkung: Nicht unabhängig sind die Ereignisse
  $A =$"es gibt viele Störche"
  $B =$"es gibt viele Babys"

- A unabhängig von $A \Leftrightarrow P(A) = 0$ oder $P(A) = 1$

Definition 8.3.

(i) Eine Familie $(A_i)_{i\in I}$ von Ereignissen $A_i \in \mathcal{A}$ (mit I beliebig) heißt (stochastisch) **unabhängig** *(bzgl. P), falls $\forall J \subset I$ endlich:*

$$P\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} P(A_i).$$

(ii) Eine Familie $(\mathcal{B}_i)_{i\in I}$ von Mengensystemen $\mathcal{B}_i \subset \mathcal{A}$ heißt **unabhängig**, *falls für jede Wahl $A_i \in \mathcal{B}_i$ $(i \in I)$ gilt:*
$(A_i)_{i\in I}$ ist unabhängig.

Bemerkungen:

- (i) ist Spezialfall von (ii) mit $B_i = \{A_i\}$

- $(A_i)_{i\in I}$ unabhängig $\Leftrightarrow \forall$ endlichen $J \subset I : (A_i)_{i\in J}$ unabhängig
  $\Rightarrow \forall i, j \in I, i \neq j : A_i$ und A_j unabhängig ("paarweise Unabhängigkeit")

Achtung: i.a. gilt nicht: paarweise Unabhängigkeit $\Rightarrow$ Unabhängigkeit

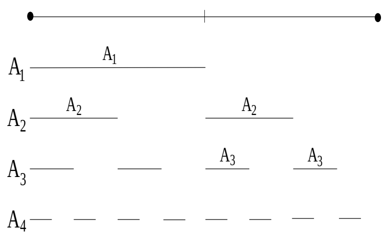
Beispiel 8.4. $2\times$ Würfeln
$A_1 =$ "Augenzahl beim 1.Wurf ungerade"

$A_2 =$ "Augenzahl beim 2.Wurf ungerade"
$A_3 =$ "Summe der Augenzahlen ungerade"
$\Rightarrow A_1$ und A_2 unabhängig, A_2 und A_3 unabhängig, A_3 und A_1 unabhängig
<u>aber</u> A_1, A_2 und A_3 <u>nicht</u> unabhängig, denn

$$0 = P(\underbrace{A_1 \cap A_2 \cap A_3}_{=\emptyset}) \neq P(A_1) \cdot P(A_2) \cdot P(A_3) > 0.$$

Beispiel 8.5. $\Omega = [0,1], \mathcal{A} = \mathcal{B}(\Omega), P = \lambda^1 = \lambda^1_{|[0,1]}$

$$A_n = \bigcup_{k=0}^{2^n-1} \left[\frac{2k}{2^n}, \frac{2k+1}{2^n}\right[$$



$\Rightarrow P(A_n) = \frac{1}{2}$

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = 2^{-k} = \prod_{l=1}^{k} P(A_{i_l}) \qquad (\forall i_1, \ldots, i_k \in \mathbb{N})$$

$\Rightarrow (A_n)_n$ <u>unabhängig</u> !

Beispiel 8.6. $\Omega = \prod_{i \in I} \Omega_i, \mathcal{A} = \bigotimes_{i \in I} \mathcal{A}_i, P = \bigotimes_{i \in I} P_i$

Für $A_i \in \mathcal{A}_i$ setze $\overline{A_i} = \pi_i^{-1}(A_i) = \ldots = \prod_{j \in I} B_j$ mit $B_j = \begin{cases} A_i & , \quad i = j \\ \Omega_j & , \quad i \neq j \end{cases}$

$(= \Omega_1 \times \ldots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \ldots$, falls $I = \mathbb{N})$
$\Rightarrow (\overline{A_i})_{i \in I}$ ist unabhängig (für belieb. $A_i \in \mathcal{A}_i$),

$$\text{denn} \quad P\left(\bigcap_{i \in J} \overline{A_i}\right) = P\left(\pi_J^{-1}\left(\prod_{i \in J} A_i\right)\right)$$
$$= \left(\bigotimes_{i \in J} P_i\right)\left(\prod_{i \in J} A_i\right) = \prod_{i \in J} P_i(A_i)$$
$$= \prod_{i \in J} P(\overline{A_i}).$$

Bemerkung: ("Verkleinern von Mengensystemen"):
Ist $(B_i)_{i \in I}$ eine unabhängige Familie von Mengensystemen $B_i \subset \mathcal{A}$ und ist $\mathcal{C}_i \subset \mathcal{B}_i$ $(\forall i \in I)$, so ist auch $(\mathcal{C}_i)_{i \in I}$ unabhängig.

Vergrößern von Mengensystemen zerstört i.a. die Unabhängigkeit.

Satz 8.7. *Sei $(\mathcal{B}_i)_{i\in I}$ unabhängige Familie von $\bigcap$-stabilen Mengensystemen $\mathcal{B}_i \subset \mathcal{A}$ $(i \in I)$. Dann gilt:*
(i) Die Familie $(\sigma(\mathcal{B}_i))_{i\in I}$ ist unabhängig.

(ii) Sei $I = \dot{\bigcup\limits_{k\in K}} J_k$ eine disjunkte Zerlegung von I.

Dann ist die Familie

$$\left(\sigma\left(\bigcup_{i\in J_k} \mathcal{B}_i\right)\right)_{k\in K} \quad \textit{unabhängig.}$$

Beweis: (i) o.B.d.A. $I \subset J$ endlich !
Betrachte für $j \in I$

$$\begin{aligned} \mathcal{D}_j &= \{A_j \in \sigma(\mathcal{B}_j) : \forall i \in I\setminus\{j\}, \forall A_i \in \mathcal{B}_i : \\ & \quad (*) \qquad P\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} P(A_i) \quad \} \end{aligned}$$

$\mathcal{D}_j$ ist ein Dynkin-System, denn
- $\Omega \in \mathcal{D}_j$

- Aus $A_j \in \mathcal{D}_j$ folgt

$$\begin{aligned} P\left(\bigcap_{i\neq j} A_i \cap A_j^c\right) &= P\left(\bigcap_{i\neq j} A_i\right) - P\left(\bigcap_{i} A_i\right) \\ &= \prod_{i\neq j} P(A_i) - \prod_{i} P(A_i) = \left(\prod_{i\neq j} P(A_i)\right)\cdot P(A_j^c), \text{ also } A_j^c \in D_j \end{aligned}$$

- Aus $A_{j_n} \in \mathcal{D}_j, n \in \mathbb{N}$ disjunkt, folgt $\ldots \bigcup\limits_n A_{j_n} \in \mathcal{D}_j$. Denn:

$$\begin{aligned} P\left(\bigcap_{i\neq j} A_i \cap \left(\bigcup_n A_{j_n}\right)\right) &= P\left(\dot{\bigcup_n}\left(A_{j_n} \cap \bigcap_{i\neq j} A_i\right)\right) = \sum_n P\left(A_{j_n} \cap \bigcap_{i\neq j} A_i\right) \\ &= \sum_n P(A_{j_n}) \prod_{i\neq j} P(A_i) = \prod_{i\neq j} P(A_i)\cdot P\left(\bigcup A_{j_n}\right) \end{aligned}$$

<u>Also</u>: $\mathcal{D}_j$ ist Dynkin-System mit $\mathcal{B}_j \subset \mathcal{D}_j \subset \sigma(\mathcal{B}_j)$ und $\mathcal{B}_j$ ist $\bigcap$-stabil.
$\Rightarrow \mathcal{D}_j = \sigma(\mathcal{B}_j)$ (d.h. $(*)$ gilt $\forall A_j \in \sigma(\mathcal{B}_j)$)

$(\tilde{\mathcal{B}}_i)_{i\in I} = (\sigma(\mathcal{B}_j), \mathcal{B}_i)_{\substack{i\neq j\\ i\in I}}$ ist unabhängig (für jedes $j \in I$)
$\Rightarrow$ (induktiv) $(\sigma(\mathcal{B}_i))_{i\in I}$ ist unabhängig
(ii) Betrachte $\mathcal{C}_k = \{\bigcap\limits_{i\in J} A_i : J \text{ endlich } \subset J_k, A_i \in \mathcal{B}_i\}$
$\Rightarrow \mathcal{C}_k$ ist $\bigcap$-stabil und $(\mathcal{C}_k)_{k\in K}$ unabhängig,

$$\begin{aligned}\text{denn } P(C_{k_1} \cap \ldots \cap C_{k_n}) &= P\left(\left(\bigcap_{i\in J_{k_1}} A_i\right) \cap \ldots \cap \left(\bigcap_{i\in J_{k_n}} A_i\right)\right) \\ &= \prod_{i\in J_{k_1}\cup\ldots\cup J_{k_n}} P(A_i) = P(C_{k_1}) \cdot \ldots \cdot P(C_{k_n}).\end{aligned}$$

$\Rightarrow$ (mit (i)): $(\sigma(\mathcal{C}_k))_{k\in K}$ unabhängig. □

Definition 8.8. *Sei $(\mathcal{B}_n)_{n\in\mathbb{N}}$ Folge von σ-Algebren auf Ω. Dann heißt $\mathcal{B}_\infty = \bigcap\limits_n \sigma\left(\bigcup\limits_{m\geq n} \mathcal{B}_m\right)$ σ-Algebra der* **terminalen Ereignisse** *(oder asymptotische σ-Algebra, tail-σ-Algebra) der Folge $(\mathcal{B}_n)_n$.*

Beispiel 8.9. $(\Omega, \mathcal{A})$ ∞-oft Münzwurf
$\mathcal{B}_n = \sigma(X_n)$ =“Information aus dem n-ten Wurf“
$\sigma\left(\bigcup\limits_{m\geq n} \mathcal{B}_m\right)$ =“Information aus den Würfen $n, n+1, \ldots$“
$\mathcal{B}_\infty$ =“Information aus der unendlich fernen Zukunft“
etwa {“*Kopf*“ kommt ∞-oft } $\in \mathcal{B}_\infty$
oder $\left\{\frac{1}{n}\sum\limits_{i=1}^{n} X_i \to \alpha \text{ für } n \to \infty\right\} \in \mathcal{B}_\infty$ für jedes α.

Satz 8.10 (0-1-Gesetz von Kolmogorov). *Ist $(\mathcal{B}_n)_{n\in\mathbb{N}}$ unabhängige Folge von σ-Algebren $\mathcal{B}_n \subset \mathcal{A}$, so gilt für jedes terminale Ereignis $A \in \mathcal{B}_\infty$*
entweder $P(A) = 0$ oder $P(A) = 1$.

Beweis: Nach Voraussetzung: $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n, \mathcal{B}_{n+m}, \ldots$ unabhängig.
$\Rightarrow \forall n \in \mathbb{N} : \mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n, \sigma\left(\bigcup\limits_{m>n} \mathcal{B}_m\right)$ unabhängig (nach Satz 8.7(ii))

$\Rightarrow \forall n \in \mathbb{N} : \mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_n, \mathcal{B}_\infty$ unabhängig ("Verkleinern")
$\Rightarrow \mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_\infty$ unabhängig
$\Rightarrow \sigma\left(\bigcup_n \mathcal{B}_n\right), \mathcal{B}_\infty$ unabhängig (nach Satz 8.7(ii))
$\Rightarrow \mathcal{B}_\infty, \mathcal{B}_\infty$ unabhängig ("Verkleinern")
$\Rightarrow \forall A \in \mathcal{B}_\infty : A$ unabhängig von $A \Rightarrow P(A) = P(A)^2 \Rightarrow P(A) \in \{0, 1\}$. □

Korollar 8.11. *$(A_n)_n$ unabhängige Folge von Ereignissen $A_n \in \mathcal{A} \Rightarrow$*

$$P(\limsup A_n) = 0 \qquad \textit{oder} \qquad = 1$$

Beweis: $(A_n)_n$ unabhängig $\Rightarrow (\sigma(A_n))_n$ unabhängig; $\limsup A_n$ terminales Ereignis. □

Bemerkung: $\limsup A_n \Leftrightarrow \exists \infty$ viele $n \in \mathbb{N}$ mit $\omega \in A_n$, d.h. "∞-viele der A_n treten ein"

Lemma 8.12 (von Borel-Cantelli).

(i) $\sum_n P(A_n) < \infty \Rightarrow P(\limsup A_n) = 0$.

(ii) Falls $(A_n)_n$ unabhängig:

$$\sum P(A_n) = \infty \Rightarrow P(\limsup A_n) = 1.$$

Beweis: von (ii): Sei $m < M < \infty$. Aus Unabhängigkeit folgt

$$P\left(\bigcap_{n=m}^{M} A_n^c\right) = \prod_{n=m}^{M} (1 - P(A_n)) \underset{\substack{\uparrow \\ \text{wegen } 1-x \le e^{-x} (\text{ für } x \ge 0)}}{\le} \prod_{n=m}^{M} e^{-P(A_n)} = e^{-\sum_{n=m}^{M} P(A_n)}$$

Für m fix und $M \to \infty$ ist nach Vor. $\sum_n^M P(A_n) \to \infty$ und damit

$$P\left(\bigcap_{n=m}^{\infty} A_n^c\right) = \lim_{M\to\infty} P\left(\bigcap_{n=m}^{M} A_n^c\right) = 0.$$

$$\Rightarrow \qquad P(\limsup A_n) = P\left(\bigcap_m \bigcup_{n=m}^{\infty} A_n\right) = \lim_{m\to\infty} P\left(\bigcup_{n=m}^{\infty} A_n\right) = 1. \square$$

Beispiel 8.13 („Affe an der Schreibmaschine").
$(\Omega, \mathcal{A}, P) = (\Omega_0, \mathcal{A}_0, P_0)^{\mathbb{N}}$
Ω_0 endlich, $\mathcal{A}_0 = \mathcal{P}(\Omega_0), P_0(\omega_0) > 0 (\forall \omega_0 \in \Omega_0)$ z.B. Gleichverteilumg auf Ω_0.
Jedes $\omega \in \Omega$ ist ein zufälliger (unendlich langer) Text,
Vorgabe eines n-Tupels $x = (x_1, \ldots, x_n) \in \Omega_0^n$
(Text der Länge n) z.B. Shakespeares Gesammelte Werke

$$\begin{aligned} T(\omega) &= \inf\{k \in \mathbb{N} \cup \{\infty\} : (\omega_k, \ldots, \omega_{k+n-1}) = (x_1, \ldots, x_n)\} \\ &= \text{ Zeitpunkt des ersten Auftretens des Textes } x \text{ im Text } \omega \end{aligned}$$

Wie groß ist $P(T < \infty)$?

$$= P(\text{ Text } x \text{ kommt irgendwann})$$

Lösung: Bilde Blöcke der Länge n

$$\underbrace{\qquad\qquad}_{\text{n}}\underbrace{\qquad\qquad}_{\text{n}}\underbrace{\qquad\qquad}_{\text{n}} \longrightarrow \mathbb{N}$$

$$A_i = \{ \text{ Text kommt im } i\text{-ten Block } \}$$

$\Rightarrow P(A_i) = \prod_{l=1}^{n} P_0(x_l) = \alpha > 0$ (unabhängig von i !)
$(A_i)_{i \in \mathbb{N}}$ unabhängig, $\sum P(A_i) = \infty$
$\Rightarrow P(\limsup A_i) = 1$
$\Rightarrow$ Text kommt ∞-oft.

Kapitel 9

Unabhängige ZV und ihre Verteilungen

a) Unabhängige ZV

Gegeben sei der Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$.

Definition 9.1. *Eine Familie $(X_i)_{i\in I}$ (I beliebig) von ZV X_i auf $(\Omega, \mathcal{A}, P)$ heißt* **unabhängig**, *falls $(\sigma(X_i))_{i\in I}$ unabhängig ist.*

Bemerkungen:

(1) Wegen $\sigma(X_i) = \{\{X_i \in A\} : A \in \mathcal{B}(\overline{\mathbb{R}})\}$ ist das äquivalent zu

$$P(\bigcap_{i\in J}\{X_i \in A_i\}) = \prod_{i\in J} P(\{X_i \in A_i\}) \qquad (*)$$

$\forall J$ endlich $\subset I$ und $\forall A_i \in \mathcal{B}(\overline{\mathbb{R}}), i \in J$

(2) Ist $(X_i)_{i\in I}$ unabhängig und $f_i : \overline{\mathbb{R}} \longrightarrow \overline{\mathbb{R}}$ meßbar ($\forall i \in I$), so ist auch $(f_i(X_i))_{i\in I}$ unabhängig.
(Denn $\sigma(f_i(X_i)) \subset \sigma(X_i)$ ($\forall i \in I$).)

b) Gemeinsame Verteilung von ZV

Gegeben sei eine ZV X_i auf $(\Omega, \mathcal{A}, P)$ $(i \in I)$ mit Verteilung $P_{X_i} = X_i(P)$ auf $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$.

Die Abbildung $X = \bigotimes_{i\in I} X_i : \begin{array}{l} \Omega \longrightarrow \overline{\mathbb{R}}^I \\ \omega \mapsto (X_i(\omega))_{i\in I} \end{array}$ heißt Produktabbildungen.

Sie ist $\mathcal{A}/\mathcal{B}(\overline{\mathbb{R}})^I$-meßbar.

Definition 9.2. *Das Bildmaß* $P_X = X(P)$ *auf* $(\overline{\mathbb{R}}^I, \mathcal{B}(\overline{\mathbb{R}})^I)$ *heißt* **gemeinsame Verteilung** *der Familie* $(X_i)_{i\in I}$.

Beispiel 9.3. $I = \{1, \ldots, n\}$

$$\Rightarrow X = (X_1, \ldots, X_n): \ \Omega \longrightarrow \overline{\mathbb{R}}^n$$
$$\omega \mapsto (X_1(\omega), \ldots, X_n(\omega))$$

P_X ist W-Maß auf $(\overline{\mathbb{R}}^n, \mathcal{B}(\overline{\mathbb{R}}^n))$, eindeutig bestimmt durch

$$P_X(A_1 \times \ldots \times A_n) = P(\bigcap_{i=1}^n \{X_i \in A_i\}) = P(X_1 \in A_1 \text{ und } \ldots \text{ und } X_n \in A_n) \quad \forall A_i \in \mathcal{B}(\overline{\mathbb{R}})$$

Satz 9.4. $(X_i)_{i\in I}$ *unabhängig* $\Longleftrightarrow P_X = \bigotimes_{i\in I} P_{X_i}$,
d.h. gemeinsame Verteilung = Produkt der einzelnen Verteilungen.

Beweis: Für beliebige J endlich $\subset I$ und $A_i \in \mathcal{B}(\overline{\mathbb{R}})$ gilt:

$$\begin{array}{ccc} P_X(\prod_{i\in J} A_i \times \overline{\mathbb{R}}^{I\setminus J}) & = & P(\bigcap_{i\in J}\{X_i \in A_i\}) \\ ? = (1) & & ? = (2) \\ (\bigotimes_{i\in I} P_{X_i})(\prod_{i\in J} A_i \times \overline{\mathbb{R}}^{I\setminus J}) & = & \prod_{i\in J} P(\{X_i \in A_i\}) \end{array}$$

Nun gilt ferner: $(X_i)_{i\in I}$ unabhängig $\Leftrightarrow$ (2) $\forall J$ endlich $\subset I$
$\forall A_i \in \mathcal{B}(\overline{\mathbb{R}})$

und $P_X = \bigotimes_{i\in I} P_{X_i} \Leftrightarrow$ (1) $\forall J$ endlich $\subset I$
$\forall A_i \in \mathcal{B}(\overline{\mathbb{R}})$ $\square$

Korollar 9.5. $X_1, \ldots, X_n$ *unabhängig,* P_{X_i} *absolut-stetig bzgl.* λ^1
mit Dichte f_i
$\Rightarrow P_X \ll \lambda^n$ *mit Dichte* $f(x_1, \ldots, x_n) = f_1(x_1) \cdot \ldots \cdot f_n(x_n)$.

Beweis:
$$\int_{A_1\times\ldots\times A_n} dP_X = P_X(A_1 \times \ldots \times A_n) = \prod P_{X_i}(A_i) = \prod \int_{A_i} f_i(x_i)\lambda^1(dx_i)$$
$$= \int_{A_1\times\ldots\times A_n} f_1(x_1) \cdot \ldots \cdot f_n(x_n)\lambda^n(dx).$$
$\square$

Beispiel 9.6. $X_1, \ldots, X_n$ unabhängig, normalverteilt auf $\mathbb{R}$ mit $E(X_i) = \alpha_i$, $\mathrm{var}(X_i) = \sigma^2$.

$\Rightarrow X = (X_1, \ldots, X_n)$ mit Verteilung P_X auf $\mathbb{R}^n$, absolut-stetig bzgl. λ^n

$\frac{dP_X(x)}{d\lambda^n(x)} = f(x) = (2\pi\sigma^2)^{-\frac{n}{2}} \cdot \exp\left(-\frac{|x-\alpha|^2}{2\sigma^2}\right)$

$|x-\alpha|^2 = \sum\limits_{i=1}^{n} |x_i - \alpha_i|^2 \qquad \alpha = (\alpha_1, \ldots, \alpha_n), x = (x_1, \ldots, x_n)$

("n-dimensionale Normalverteilung")

Beispiel 9.7. $X_1, \ldots, X_n$ unabhängig, gleichverteilt auf $[0,1]$
$\Rightarrow$ gemeinsame Verteilung = Gleichverteilung auf $[0,1]^n$.

Beispiel 9.8. $X_1 = X_2$ gleichverteilt auf $[0,1]$
$\Rightarrow$ gemeinsame Verteilung = Gleichverteilung auf der Diagonale von $[0,1]^2$.

c) Faltung von W-Maßen

Seien $\mu_1, \ldots, \mu_n$ W-Maße auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Dann ist das Produktmaß $\mu_1 \otimes \ldots \otimes \mu_n$ ein W-Maß auf $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.
Dessen Bildmaß unter der Abbildung

$$s_n : \begin{array}{l} \mathbb{R}^n \longrightarrow \mathbb{R} \\ (x_1, \ldots, x_n) \mapsto x_1 + \ldots x_n \\ \text{(stetig, also meßbar)} \end{array}$$

ist ein W-Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Definition 9.9. *Es wird mit*

$$\underbrace{\mu_1 * \ldots * \mu_n}_{\text{W-Maß auf } \mathbb{R}} = s_n(\underbrace{\mu_1 \otimes \ldots \otimes \mu_n)}_{\text{W-Maß auf } \mathbb{R}^n}$$

bezeichnet und heißt **Faltung** *(oder Faltungsprodukt) von* $\mu_1, \ldots, \mu_n$.

Proposition 9.10.

(i) Für jedes meßbare $f \geq 0 auf\, \mathbb{R}$ gilt

$$\begin{aligned} \int_{\mathbb{R}} f d(\mu_1 * \ldots * \mu_n) &= \int_{\mathbb{R}^n} f \circ s_n d(\mu_1 \otimes \ldots \otimes \mu_n) \\ &= \int_{\mathbb{R}^n} f(s_n(x))(\mu_1 \otimes \ldots \otimes \mu_n)(dx) \\ &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1 + \ldots + x_n)\mu_1(dx_1) \cdot \ldots \cdot \mu_n(dx_n) \end{aligned}$$

(ii) Im Fall $n = 2$ gilt für jedes $A \in \mathcal{B}(\mathbb{R})$

$$
\begin{aligned}
(\mu_1 * \mu_2)(A) &= \int_{\mathbb{R}} \int_{\mathbb{R}} 1\!\!1_A(x_1 + x_2)\mu_1(dx_1)\mu_2(dx_2) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} 1\!\!1_{A-x_1}(x_2)\mu_2(dx_2)\mu_1(dx_1) \\
&= \int_{\mathbb{R}} \mu_2(A - x_1)\mu_1(dx_1) = \int_{\mathbb{R}} \mu_1(A - x_2)\mu_2(dx_2)
\end{aligned}
$$

(iii) Ist einer der beiden Faktoren μ_i absolut-stetig bzgl. λ^1, etwa $\mu_1 = \varphi_1 \cdot \lambda^1$, so ist auch $\mu_1 * \mu_2$ absolut-stetig mit Dichte φ gegeben durch

$$\varphi(x) = \int \varphi_1(x - x_2)\mu_2(dx_2).$$

Denn:
$$
\begin{aligned}
(\mu_1 * \mu_2)(A) &= \int \mu_1(A - x_2)\mu_2(dx_2) \\
&= \int \int_{A-x_2} \varphi_1(x_1)\lambda^1(dx_1)\mu_2(dx_2) = \int \int_A \varphi_1(x_1 - x_2)\lambda^1(dx_1)\mu_2(dx_2) \\
&= \int_A \varphi(x_1)\lambda^1(dx_1)
\end{aligned}
$$

(iv) Ist $\mu_i = \varphi_i \cdot \lambda^1$ für $i = 1, 2$

$$\mu_1 * \mu_2 = (\varphi_1 * \varphi_2) \cdot \lambda^1$$

mit $\varphi_1 * \varphi_2(x) = \int \varphi_1(x - y)\varphi_2(y)dy$ = Faltung der Funktionen φ_1 und φ_2

Bemerkung: Obiges gilt nicht nur für W-Maße auf $\mathbb{R}$, sondern für endliche Maße auf $\mathbb{R}^d$.

Korollar 9.11. *Die Faltung von endlichen Maßen ist*

- *kommutativ:* $\mu_1 * \mu_2 = \mu_2 * \mu_1$
- *assoziativ:* $(\mu_1 * \mu_2) * \mu_3 = \mu_1 * (\mu_2 * \mu_3)$
- *distributiv:* $\mu_1 * (\mu_2 + \mu_3) = \mu_1 * \mu_2 + \mu_1 * \mu_3$

Analoges gilt für die Faltung von Funktionen.

Ferner gilt: $\delta_x * \delta_y = \delta_{x+y}$
und $\mu * \delta_0 = \mu$.

Bemerkung: Sei $\psi \geq 0$ in $\mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ mit $\int \psi d\lambda = 1$ und sei $\psi_\epsilon(x) = \epsilon \cdot \psi(\frac{x}{\epsilon})$. Dann gilt $\psi_\epsilon \cdot \lambda \longrightarrow \delta_0$ und

$$\mu * (\psi_\epsilon \cdot \lambda) \longrightarrow \mu \text{ schwach für } \epsilon \to 0$$

für jedes W-Maß μ.
Beachte: $\mu * (\psi_\epsilon \cdot \lambda) \ll \lambda$.

d) Summen unabhängiger ZV

Satz 9.12. *Seien $X_1, \ldots, X_n$ unabhängige, reelle ZV und $S_n = \sum\limits_{i=1}^n X_i$. Dann ist*

$$P_{S_n} = P_{X_1} * \ldots * P_{X_n}.$$

Beweis: Sei $s_n : \mathbb{R}^n \longrightarrow \mathbb{R}$, $(x_1, \ldots, x_n) \mapsto x_1 + \ldots + x_n$ und $X = (X_1, \ldots, X_n) : \Omega \longrightarrow \mathbb{R}^n$. Dann ist $S_n = s_n \circ X$

$$\begin{aligned}
\Rightarrow P_{S_n}(A) &= P_{s_n \circ X}(A) = P((s_n \circ X)^{-1}(A)) = P(X^{-1}(s_n^{-1}(A))) \\
&\overset{\text{unabhängig}}{=} P_X(s_n^{-1}(A)) = \underbrace{(P_{X_1} \otimes \ldots \otimes P_{X_n})(s_n^{-1}}_{\text{Bildmaß}}(A)) \\
&= (P_{X_1} * \ldots * P_{X_n})(A) \qquad (\forall A \in \mathcal{B}(\mathbb{R})). \square
\end{aligned}$$

Beispiel 9.13. Binomialverteilung $\beta_n^p = \sum\limits_{k=0}^n \binom{n}{k} p^k \cdot (1-p)^{n-k} \cdot \delta_k$.
Seien $X_1, \ldots, X_n$ unabhängig nach $\beta_1^p = p \cdot \delta_1 + (1-p) \cdot \delta_0$ verteilt.
(d.h. $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$ $\forall i = 1, \ldots, n$).
Dann ist S_n nach β_n^p verteilt.

$$\text{M.a.w.} \qquad \underbrace{\beta_1^p * \ldots * \beta_1^p}_{\text{n-mal}} = \beta_n^p$$

<u>Beweis durch Induktion:</u>

$$\begin{aligned}
\beta_{n-1}^p * \beta_1^p &= \left(\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} \cdot \delta_k\right) * (p \cdot \delta_1 + (1-p) \cdot \delta_0) \\
&= \sum_{k=0}^{n-1} \binom{n-1}{k} p^{k+1} \cdot (1-p)^{n-(k+1)} \cdot \delta_{k+1} + \sum_{k=0}^{n-1} \binom{n-1}{k} p^k \cdot (1-p)^{n-k} \cdot \delta_k \\
&= \sum_{k=0}^{n} \underbrace{[\binom{n-1}{k-1} + \binom{n-1}{k}]}_{\frac{k}{n}\binom{n}{k} + \frac{n-k}{n}\binom{n}{k}} \cdot p^k \cdot (1-p)^{n-k} \cdot \delta_k = \beta_n^p \square
\end{aligned}$$

Das hatten wir bereits immer wieder verwendet!
Aufgrund der Assoziativität der Faltung folgt:
Sind X_1 und X_2 unabhängig nach $\beta^p_{n_1}$ bzw. $\beta^p_{n_2}$ verteilt, so ist $X_1 + X_2$ nach $\beta^p_{n_1+n_2}$-verteilt.

Beispiel 9.14. Poisson-Verteilung $\pi(\lambda) = e^{-\lambda} \cdot \sum\limits_{k=0}^{\infty} \frac{\lambda^k}{k!}\delta_k$.
Sind X_1, X_2 unabhängig nach $\pi(\lambda_i)$-verteilt, so ist X_1+X_2 nach $\pi(\lambda_1+\lambda_2)$-verteilt.

<u>Beweis:</u> Sei $\mu_i = P_{X_i} = \pi(\lambda_i)$

$$\begin{aligned}
\Rightarrow P_{X_1+X_2}(\{n\}) &= \mu_1 * \mu_2(\{n\}) = \sum_{k=-\infty}^{+\infty} \mu_1(\{n-k\})\mu_2(\{k\}) \\
&= e^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{n} \frac{\lambda_1^{n-k} \cdot \lambda_2^k}{(n-k)!k!} \text{ mit } \lambda = \lambda_1 + \lambda_2 \\
&= e^{-\lambda} \cdot \frac{\lambda^n}{n!} \cdot \underbrace{\sum_{k=0}^{n} \binom{n}{k} \cdot \left(\frac{\lambda_1}{\lambda}\right)^{n-k} \cdot \left(\frac{\lambda_2}{\lambda}\right)^{k}}_{=1} \\
&= \pi(\lambda)(\{n\})\square
\end{aligned}$$

Beispiel 9.15. Normalverteilung $\nu_{\alpha,\sigma^2} = \varphi_{\alpha,\sigma^2} \cdot \lambda^1$ mit

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

Sind X_1, X_2 unabhängig nach $\nu_{\alpha_i,\sigma_i^2}$-verteilt, so ist $X_1 + X_2$ nach $\nu_{\alpha_1+\alpha_2,\sigma_1^2+\sigma_2^2}$-verteilt.

Beweis:

$$
\begin{aligned}
\varphi(x) &= (\varphi_{\alpha_1,\sigma_1^2} * \varphi_{\alpha_2,\sigma_2^2})(x) \\
&= (2\pi\sigma_1\sigma_2)^{-1} \cdot \int_{\mathbb{R}} \exp\left(-\frac{|x-y-\alpha_1|^2}{2\sigma_1^2} - \frac{|y-\alpha_2|^2}{2\sigma_2^2}\right) dy \\
&= \varphi_{\alpha_1+\alpha_2,\sigma_1^2+\sigma_2^2}(x) \cdot \left(2\pi \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right)^{-\frac{1}{2}} \cdot \\
&\quad \cdot \int_{\mathbb{R}} \exp\left(+\frac{|x-\alpha_1+\alpha_2|^2}{2\sigma_1^2+2\sigma_2^2} - \frac{|x-y-\alpha_1|^2}{2\sigma_1^2} - \frac{|y-\alpha_2|^2}{2\sigma_2^2}\right) dy \\
&= \varphi_{\alpha_1+\alpha_2,\sigma_1^2+\sigma_2^2}(x) \cdot \underbrace{\left(2\pi \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right)^{-\frac{1}{2}} \cdot \int_{\mathbb{R}} \exp\left(-\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\cdot\sigma_2^2}(y-\ldots)^2\right) dy}_{=1} \\
&= \varphi_{\alpha_1+\alpha_2,\sigma_1^2+\sigma_2^2}(x). \square
\end{aligned}
$$

Beispiel 9.16. Gamma-Verteilung $\Gamma(t,\alpha) = \varphi_{t,\alpha} \cdot \lambda^1$ mit

$$
\varphi_{t,\alpha}(x) = \begin{cases} \frac{\alpha^t}{\Gamma(t)} \cdot e^{-\alpha x} \cdot x^{t-1} & , \ x \geq 0 \\ o & , \ \text{sonst.} \end{cases}
$$

$$
E(X) = \int x d\Gamma(t,\alpha) = \frac{t}{\alpha}.
$$

- Sind X_1, X_2 unabhäng. nach $\Gamma(t_i,\alpha)$ verteilt, so ist $X_1 + X_2$ nach $\Gamma(t_1+t_2,\alpha)$ verteilt.
  ($\rightarrow$ Übungsstunde) Denn:

$$
\begin{aligned}
(\varphi_{t_1,\alpha} * \varphi_{t_2,\alpha})(x) &= \tfrac{\alpha^{t_1+t_2}}{\Gamma(t_1)\cdot\Gamma(t_2)} \cdot \int_0^x e^{-\alpha(x-y)} \cdot (x-y)^{t_1-1} \cdot e^{-\alpha y} \cdot y^{t_2-1} dy \\
&= \tfrac{\alpha t}{\Gamma(t)} \cdot e^{-\alpha x} \cdot x^{t_1} \cdot \underbrace{\frac{\Gamma(t)}{\Gamma(t_1)\cdot\Gamma(t_2)} \cdot \int_0^1 (1-y)^{t_1-1} \cdot y^{t_2-1} dy}_{=1 \text{ wegen } \int \varphi_{t_1,\alpha} * \varphi_{t_2,\alpha} d\lambda = 1}.
\end{aligned}
$$

- Für $t=1$: Exponentialverteilung: $\mathrm{Exp}(\alpha)$
  Sind $X_1,\ldots,X_n$ unabhängig nach $\mathrm{Exp}(\alpha)$ verteilt, so ist $X_1+\ldots+X_n$ nach $\Gamma(n,\alpha)$ verteilt. (Übung 7/2b)

- Für $t=\frac{1}{2}$, $\alpha = \frac{1}{2\sigma^2}$ Chiquadrat-Verteilung.
  Sind $X_1,\ldots,X_n$ unabhängig ν_{0,σ^2}-verteilt, so sind $X_1^2,\ldots,X_n^2$ unabhängig χ^2-verteilt $= \Gamma\left(\frac{1}{2},\frac{1}{2\sigma^2}\right)$-verteilt (Übung 7/4c)
  und $X_1^2+\ldots+X_n^2$ ist χ_n^2-verteilt $= \Gamma\left(\frac{n}{2},\frac{1}{2\sigma^2}\right)$-verteilt
  (= Chiquadrat-Verteilung mit n Freiheitsgraden).
  $|X| = (X_1^2+\ldots+X_n^2)^{\frac{1}{2}}$ ist <u>χ-verteilt</u> mit n Freiheitsgraden. (Übung 8/4a)

Beispiel 9.17. Cauchy-Verteilung $C(\alpha) = \varphi_\alpha \cdot \lambda^1$ mit

$$\varphi_\alpha(x) = \frac{1}{\pi} \frac{\alpha}{\alpha^2 + x^2}.$$

Sind X_1, X_2 unabhängig nach $C(\alpha_i)$-verteilt, so ist $X_1 + X_2$
nach $C(\alpha_1 + \alpha_2)$-verteilt.
Beweis später mit Fourier-Transformation!

Bemerkung: Ist X nach $C(\alpha)$-verteilt, so ist $2X$ nach $C(2\alpha)$-verteilt (nachrechnen!).
Folglich $\quad P_X * P_X = P_{2X}$.
Also: $P_{X_1+X_2} = P_{X_1} * P_{X_2} \not\Rightarrow X_1, X_2$ unabhängig, d.h. in obigem Satz gilt keine Umkehrung!

e) Produkte unabhängiger ZV

Satz 9.18.

(i) Seien $X_1, \dots, X_n$ unabhängig und ≥ 0. Dann ist

$$E(X_1 \cdot \ldots \cdot X_n) = E(X_1) \cdot \ldots \cdot E(X_n). \qquad (*)$$

(ii) Sind $X_1, \dots, X_n$ unabhängig und $\in \mathcal{L}^1$, so ist $X_1 \cdot \ldots \cdot X_n \in \mathcal{L}^1$ und $(*)$ gilt.

Beweis: Es genügt (i) zu zeigen.
Wegen Unabhängigkeit ist $P_X = \bigotimes_{i=1}^n P_{X_i}$ die gemeinsame Verteilung von
$X = (X_1, \dots, X_n)$.

$$\begin{aligned}\Rightarrow E(X_1 \cdot \ldots \cdot X_n) &= \int x_1 \cdot \ldots \cdot x_n P_X(dx) = \int \ldots \int x_1 \cdot \ldots \cdot x_n P_{X_1}(dx_1) \ldots P_{X_n}(dx_n) \\ &= \int x_1 P_{X_1}(dx_1) \cdot \ldots \cdot \int x_n P_{X_n}(dx_n) \\ &= E(X_1) \cdot \ldots \cdot E(X_n) \square\end{aligned}$$

Beachte:

- Allgemein gilt $X, Y \in \mathcal{L}^2 \Rightarrow X \cdot Y \in \mathcal{L}^1$.

- Falls X, Y unabhängig, so gilt:

$$X, Y \in \mathcal{L}^1 \Rightarrow X \cdot Y \in \mathcal{L}^1$$

$$(\text{bzw.} \qquad X, Y \in \mathcal{L}^2 \Rightarrow X \cdot Y \in \mathcal{L}^2\ )$$

In obigem Satz 9.18(i) gilt keine Umkehrung (siehe "Unkorreliertheit"). Aber

Korollar 9.19. *$X_1, \ldots, X_n$ unabhängig $\Longleftrightarrow \forall f_1, \ldots, f_n \geq 0$ meßbar auf $\overline{\mathbb{R}}$:*

$$E(f_1(X_1) \cdot \ldots \cdot f_n(X_n)) = E(f_1(X_1)) \cdot \ldots \cdot E(f_n(X_n))$$

Beweis: " $\Rightarrow$ " Aus der Unabhängigkeit von $(X_i)_i$ folgt die Unabhängigkeit von $(f_i(X_i))_i$ und damit die Behauptung wegen Satz 14.9.(i).

" $\Leftarrow$ " $\quad \forall A_1, \ldots, A_n \in \mathcal{B}(\overline{\mathbb{R}})$ gilt (mit $f_i = 1\!\!1_{A_i}$) :

$$\begin{aligned} P(\{X_1 \in A_1\} \cap \ldots \cap \{X_n \in A_n\}) &= E(f_1(X_1) \cdot \ldots \cdot f_n(X_n)) \\ &= E(f_1(X_1)) \cdot \ldots \cdot E(f_n(X_n)) = P(X_1 \in A_1) \cdot \ldots \cdot P(X_n \in A_n). \end{aligned}$$ □

f) Unabhängigkeit und Unkorreliertheit

Definition 9.20. *Für $X, Y \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ heißt*

$$\begin{aligned} cov(X, Y) &= E([X - E(X)] \cdot [Y - E(Y)]) \\ &= E(X \cdot Y) - E(X) \cdot E(Y) \end{aligned}$$

Kovarianz *von X und Y.*
X und Y heißen **unkorreliert**, *falls* $\operatorname{cov}(X, Y) = 0$.

<u>Elementare Eigenschaften</u>:
(i) X, Y unabhäng. $\Rightarrow$ unkorrel.

(ii) $\operatorname{var}(X) = \operatorname{cov}(X, X)$

(iii) X und X unkorreliert $\Leftrightarrow$ X f.s. konstant
$\Rightarrow$ X und Y unabhängig $(\forall Y)$

(iv) $\mathrm{var}(X+Y) = \mathrm{var}X + 2\cdot\mathrm{cov}(X,Y) + \mathrm{var}Y$
(denn: o.B.d.A. $E(X) = E(Y) = 0 \Rightarrow \mathrm{var}(X+Y) = E((X+Y)^2) = \ldots$)

(v) X, Y unkorreliert $\Leftrightarrow \mathrm{var}(X+Y) = \mathrm{var}X + \mathrm{var}Y$

(vi) Stets: $|\mathrm{cov}(X,Y)| \leq \sqrt{\mathrm{var}X}\cdot\sqrt{\mathrm{var}Y}$ (Cauchy-Schwarz-Ungleichung)
Die Zahl $\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}X}\cdot\sqrt{\mathrm{var}Y}} \in [-1,1]$ heißt **Korrelationskoeffizient**.

(vii) $\rho(X,Y) = \pm 1 \Leftrightarrow \exists \alpha, \beta : X = \alpha\cdot Y + \beta$ P-f.s. („völlige Abhängigkeit ")
(denn:

$$\begin{aligned}\Rightarrow \mathrm{var}[X + t\cdot Y] &= \mathrm{var}X + 2t\cdot\mathrm{cov}(X,Y) + t^2\cdot\mathrm{var}Y \\ &= \mathrm{var}X \pm 2t\cdot\sqrt{\mathrm{var}X}\cdot\sqrt{\mathrm{var}Y} + t^2\cdot\mathrm{var}Y \\ &= (\sqrt{\mathrm{var}X} \pm t\cdot\sqrt{\mathrm{var}Y})^2 = 0 \text{ für } t = \mp\frac{\sqrt{\mathrm{var}X}}{\sqrt{\mathrm{var}Y}}\end{aligned}$$

$\Rightarrow X + t\cdot Y = Const.$ P-f.s.)

Satz 9.21. *$(X_n)_n \subset \mathcal{L}^2$ (paarweise) unkorreliert $\Longrightarrow$*

$$var\left(\sum_{n=1}^{k} X_n\right) = \sum_{n=1}^{k} varX_n$$

(Gleichheit von Bienaymé).

Beweis: $\mathrm{var}(\left(\sum\limits_{n=1}^{k} X_n\right) = \sum\limits_{n} \mathrm{var}X_n + 2\sum\limits_{n=m}\mathrm{cov}(X_n, X_m)$ □

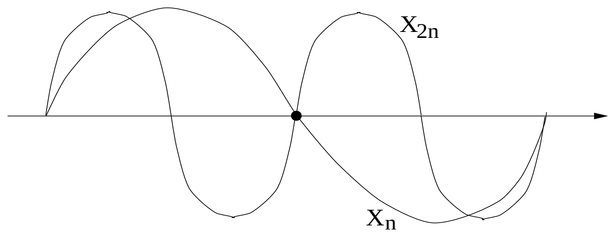
<u>Beachte</u>: Für Indikatorfunktion $\mathbb{1}_A, \mathbb{1}_B$ mit $A, B \in \mathcal{A}$ gilt:

$$\mathbb{1}_A, \mathbb{1}_B \text{ unabhängig } \Leftrightarrow \mathbb{1}_A, \mathbb{1}_B \text{ unkorreliert}$$

(denn: $\mathrm{cov}(\mathbb{1}_A, \mathbb{1}_B) = P(A\cap B) - P(A)\cdot P(B)$.)
Für allgemeine ZV gilt $\Rightarrow$ und $\not\Leftarrow$! (Keine maßtheoretische Induktion!)

Beispiel 9.22. $X_n(\omega) = \sin(2\pi n\cdot\omega)$ auf $(\Omega, \mathcal{A}, P) = ([0,1], \mathcal{B}([0,1]), \lambda^1)$
$\Rightarrow \forall n \neq m : X_n$ und X_m unkorreliert, denn $\mathrm{cov}(X_n, X_m) = E(X_n\cdot X_m) = 0$
Aber X_n und X_{2n} abhängig, denn für $\epsilon \to 0$:

$$\begin{aligned}P(|X_n| \leq \epsilon) &\approx \epsilon\cdot\frac{2}{\pi} \\ P(|X_{2n}| \leq \epsilon) &\approx \epsilon\cdot\frac{2}{\pi} \\ P(|X_n| \leq \epsilon \text{ und } |X_{2n}| \leq \epsilon) &\approx \epsilon\cdot\frac{1}{\pi}\end{aligned}$$



Beispiel 9.23. $\Omega = \{1,2,3\}$ mit Gleichverteilung P

$$X = \begin{cases} 1 \\ 0 \\ -1 \end{cases} \qquad Y = \begin{cases} 0 \\ 1 \\ 0 \end{cases}$$

Beispiel 9.24. $(\Omega, \mathcal{A}, P) = ([-\frac{1}{2}, \frac{1}{2}], \mathcal{B}(\Omega), \lambda^1)$
$X(\omega) = \omega,\ \ Y(\omega) = |\omega|$
$\Rightarrow E(X \cdot Y) = 0 = E(X) \cdot E(Y) \Rightarrow X, Y$ unkorreliert. Aber X, Y abhängig.

g) Das Konstruktionsprinzip für unabhängige ZV

Gegeben Verteilungen (=W-Maße) μ_i auf $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, $i \in I$, I beliebig.
Gesucht W-Raum $(\Omega, \mathcal{A}, P)$ und unabhängige ZV X_i mit Verteilung $P_{X_i} = \mu_i$, $i \in I$.

Lösung: $\Omega = \overline{\mathbb{R}}^I, \mathcal{A} = \mathcal{B}(\overline{\mathbb{R}})^I, P = \bigotimes_{i \in I} \mu_i$

und $X_i: \ \Omega \longrightarrow \overline{\mathbb{R}}$ Projektion.
$\omega \mapsto X_i(\omega) = \omega_i$

$\Rightarrow P_{X_i} = \mu_i$ (wegen der Definition des Produktmaßes) und
$\bigotimes_{i \in I} P_{X_i} = \bigotimes \mu_i = P = P_X$, da $X: \Omega \longrightarrow \overline{\mathbb{R}}^I$ Identität.

Ähnliches gilt: Beschreibt $(\Omega_i, \mathcal{A}_i, P_i)$ für jedes $i \in I$ ein Zufallsexperiment (mit $\Omega_i \subset \overline{\mathbb{R}}, \mathcal{A}_i = \mathcal{B}(\Omega_i)$),
so beschreibt $(\Omega, \mathcal{A}, P)$ mit $\Omega = \prod \Omega_i, \mathcal{A} = \bigotimes \mathcal{A}_i, P = \bigotimes P_i$ das Experiment, das aus der unabhängigen Durchführung der Einzelexperimente besteht.
Die Projektionen $X_i : \Omega \longrightarrow \Omega_i$ bilden eine Familie unabhängiger ZV.

Allgemein: Für jedes $i \in I$ wird ein Zufallsexperiment durch eine ZV $Z_i : \Omega_i \longrightarrow \overline{\mathbb{R}}$ auf einem W-Raum $(\Omega_i, \mathcal{A}_i, P_i)$ beschrieben.

Gesucht: Ein W-Raum $(\Omega, \mathcal{A}, P)$ und ZV $X : \Omega \longrightarrow \overline{\mathbb{R}}$, die die "unabhängige Durchführung der Einzelexperimente" beschreibt in folgendem Sinne:

- $\forall i \in I: \; Z_i \overset{\mathrm{d}}{=} X_i$, d.h. $P_{Z_i} = P_{X_i}$

- die Familie $(X_i)_i$ ist unabhängig

Lösung: $\Omega = \prod \Omega_i,\ \mathcal{A} = \bigotimes \mathcal{A}_i,\ P = \bigotimes P_i$
$X_i(\omega) = Z_i(\omega_i)$, d.h. $X_i = Z_i \circ \pi_i, X = \bigotimes Z_i \circ \pi_i$
$\Rightarrow P_{X_i} = P_{Z_i \circ \pi_i} = (P_{\pi_i})_{Z_i} = (P_i)_{Z_i}$, also $X_i \overset{\mathrm{d}}{=} Z_i$
$\Rightarrow \bigotimes P_{X_i} = \bigotimes (P_i)_{Z_i} = (\bigotimes P_i)_X = P_X$, also Unabhängigkeit

Obige Beispiele sind Spezialfälle: $Z_i =$id auf $\Omega_i = \overline{\mathbb{R}}$
bzw.
Alternative Lösung: $\Omega = \overline{\mathbb{R}}^I, \mathcal{A} = \mathcal{B}(\overline{\mathbb{R}})^I, P = \bigotimes (P_i)_{Z_i}, X_i(\omega) = \omega_i$ Projektion

h) Beispiel: "Maximum erraten"

Sie spielen mit Ihrem Tischnachbar folgendes Spiel:
- Er schreibt von Ihnen nicht einsehbar auf $n = 1000$ Zetteln n verschiedene, zufällige, reelle Zahlen X_i, $i = 1, \dots, n$; dann werden die Zettel umgedreht und gut durchmischt.
  (Die $X_1, \dots, X_n$ seien unabhängig, identisch verteilt mit stetiger Verteilung.)

- Sie sollen erraten, auf welchem Zettel die größte Zahl steht.
  Hierzu dürfen Sie der Reihe nach die Zettel aufdecken; jeweils nach dem Aufdecken des i-ten Zettels müssen Sie bekanntgeben, ob X_i Ihrer Meinug nach die größte der Zahlen $X_1, \dots, X_n$ ist - oder nicht.
  Natürlich können Sie sich nur für ein i entscheiden.
- Sie bekommen 10 DM, wenn Sie richtig raten und müssen 1 DM zahlen, falls Sie falsch liegen.
  Ist das für Sie gut?

Vorüberlegung
- $Y_1, \dots, Y_n$ seien die Zahlen, die Ihr Mitspieler der Reihe nach aufschreibt, es sind ZV auf $(\Omega_0, \mathcal{A}_0, P_0)$ mit $P_0(Y_i \neq Y_j) = 1 \ \forall i \neq j$.

- $$\Omega_1 = \{ \text{ Permutation von } \{1, \dots, n\}\} \text{ mit Gleichverteilung } P_1$$

- $X_1, \dots, X_n$ sind ZV auf $\Omega = \Omega_0 \times \Omega_1$ mit $P = P_0 \otimes P_1$ mit
  $$X_i(\omega_0, \omega_1) = Y_{\omega_1(i)}(\omega_0)$$
  Offenbar gilt $(X_i)_{i=1,\dots,n}$ identisch verteilt.
  (denn: $P(X_i \in A) = \int\int \mathbb{1}_A(Y_{\omega_1(i)}(\omega_0)) P_0(d\omega_0) P_1(d\omega_1) = \frac{1}{n} \sum\limits_{k=1}^{n} P_0(Y_k \in A) \qquad \forall i, \forall A)$

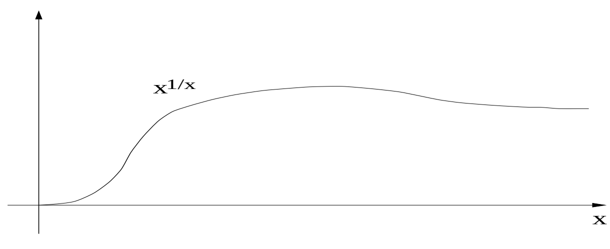
- Sei nun $A_i = \bigcap\limits_{k<i} \{X_i > X_k\}$ =“Rekord im i-ten Spiel“
  $\Rightarrow P(A_i) = \frac{1}{i}$ und $(A_i)_i$ unabhängig.

Strategie: Sie wählen $k \in \{1, \ldots, n\}$, warten die Spiele $X_1, \ldots, X_k$ ab und entscheiden sich dann für das erste $i \in \{k+1, \ldots, n\}$, für welches X_i den bisherigen Rekord $\max\limits_{j=1,\ldots,k} X_j$ bricht,

d.h. $i = \begin{cases} \inf\{l \in \{k+1, \ldots, n\} : X_l > \max\limits_{j=1,\ldots,k} X_j\} & , \text{ falls } \{k+1, \ldots, n\} \neq \emptyset \\ n & , \text{ sonst} \end{cases}$

Wie groß ist die Wahrscheinlichkeit, mit dieser Strategie zu gewinnen?

$$
\begin{aligned}
P(\text{Gewinnen}) &= P\Big(\dot{\bigcup_{l=k+1}^{n}} A_{k+1}^c \cap \ldots \cap A_{l-1}^c \cap A_l \cap A_l^c \cap \ldots \cap A_n^c\Big) \\
&\underset{\substack{\uparrow \\ \text{unabh.}}}{=} \left(\sum_{l=k+1}^{n} P(A_l)/P(A_l^c) \right) \cdot \prod_{j=k+1}^{n} P(A_j^c) \;\; (A_l = \text{Rekord im } l\text{-ten Spiel}) \\
&= \sum_{l=k+1}^{n} \frac{1/l}{1-1/l} \cdot \prod_{j=k+1}^{n} \left(1 - \frac{1}{j}\right) \\
&= \sum_{l=k+1}^{n} \frac{1}{l-1} \cdot \frac{k}{n} \\
&\approx (\ln n - \ln k) \cdot \frac{k}{n} = \ln\left(\left(\frac{n}{k}\right)^{(k/n)} \right) \underset{\substack{\uparrow \\ \text{“=“ für } \frac{n}{k} = e}}{\leq} \ln\left(e^{1/e}\right) = \frac{1}{e}
\end{aligned}
$$



Also: Optimales $k \approx \frac{n}{e}$ und damit $P(\text{Gewinn}) \approx \frac{1}{e} \approx 0,365$ und $E(\text{Gewinn}) \approx 3$.
$\Rightarrow$ Lukratives Spiel!

Kapitel 10

Die Gesetze der großen Zahlen

a) Motivation

Seien $(X_i)_{i\in\mathbb{N}}$ <u>identisch verteilt</u> (d.h. $\mu = P_{X_i}$ hängt nicht von $i \in \mathbb{N}$ ab) mit $E(X_i) = \alpha$.
Für $n \in \mathbb{N}$ betrachte $S_n = \sum\limits_{i=1}^{n} X_i$ und $\frac{1}{n}S_n$.

<u>Ziel</u>: Für $n \longrightarrow \infty$ "mittelt sich der Zufall weg", d.h. es gilt:

$$\begin{array}{ccc} \frac{1}{n}S_n(\omega) & \overset{(?)}{\longrightarrow} & \alpha = E(X_i) \\ \uparrow & & \uparrow \\ \text{empirischer Mittelwert} & & \text{Erwartungswert} \\ = \text{Mittelwert der Beobachtungswerte} & & = \text{Mittelwert der möglichen Aus-} \\ X_1(\omega), \dots, X_n(\omega) \text{ für festes } \omega \in \Omega & & \text{gänge } X_i(\omega), \omega \in \Omega, \\ (\text{gemittelt über } i = 1, \dots, n) & & \text{für festes } i \in \mathbb{N}. \\ & & (\text{gemittelt über } \omega \in \Omega) \end{array}$$

Bei einem Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p ist $E(X_i) = \alpha = p$ und

$$\frac{1}{n}S_n(\omega) = \text{relative (Erfolgs-)Häufigkeit}$$

D.h. $\qquad$ Relative Häufigkeit $\longrightarrow$ Wahrscheinlichkeit

b) Das Schwache Gesetz

<u>Voraussetzungen:</u>

- $X_i \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$, $i \in \mathbb{N}$,

- $\operatorname{var}(X_i) \le C < \infty, \alpha = E(X_i) \quad (\forall i \in \mathbb{N})$

- $(X_i)_{i\in\mathbb{N}}$ paarweise unkorreliert

Satz 10.1 („Schwaches Gesetz der Großen Zahlen“ (Khintchine)).

$$\frac{1}{n}S_n \longrightarrow \alpha \textit{ in } L^2 \textit{ und } P\textit{-stochastische für } n \to \infty.$$

Beweis: Stochastische Konvergenz folgt aus L^2-Konvergenz. Letztere folgt aus

$$E\left(\left[\frac{1}{n}S_n - \alpha\right]^2\right) = \mathrm{var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\mathrm{var}S_n \underset{\substack{\uparrow\\ \text{unkorr.}}}{=} \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}(X_i) \le \frac{C}{n} \to 0$$

□

Bemerkungen:

- Statt $\mathrm{var}X_i \le C$ genügt $\frac{1}{n^2}\sum\limits_{i=1}^{n}\mathrm{var}X_i \to 0$ $(n \to \infty)$.
- Falls $E(X_i)$ beliebig, betrachte $\tilde{X}_i = X_i - E(X_i)$, $\tilde{\alpha} = 0$, $\frac{1}{n}\tilde{S}_n \to 0$.

Anwendungen:

1. Ars conjectundi (J. Bernoulli (1654-1705)): Bei Bernoulli-Experimenten:

Relative Erfolgshäufigkeit $\frac{1}{n}S_n \longrightarrow$ Erfolgswahrscheinlichkeit p, P-stochastisch für $n \to \infty$

$\Rightarrow$ Interpretation der Wahrscheinlichkeit als relative Häufigkeit

2. Gleichmäßige Approximation durch Bernstein-Polynome

Sei $f \in \mathcal{C}([0,1])$ und definiere das n-te **Bernstein-Polynom** f_n durch

$$f_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\cdot\binom{n}{k}x^k\cdot(1-x)^{n-k} = E_x\left(f\left(\frac{S_n}{n}\right)\right)$$

falls $S_n = \sum\limits_{i=1}^{n} X_i$ mit $X_1, \ldots, X_n$ unabhängig, identisch Bernoulli-verteilt
$P_x(X_i = 1) = x = 1 - P_x(X_i = 0)$.

$$
\begin{aligned}
\Rightarrow |f_n(x) - f(x)| &= \left| E_x\left[f\left(\frac{S_n}{n}\right)\right] - f(x)\right| \\
&\leq E_x\left[\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\right] \\
&\leq \epsilon \cdot P_x\left[\left|\frac{S_n}{n} - x\right| \leq \delta\right] + 2 \cdot \|f\|_\infty \cdot P_x\left[\left|\frac{S_n}{n} - x\right| > \delta\right] \\
&\quad \text{da } \forall \epsilon > 0 : \exists \delta > 0 : |y - x| \leq \delta \Rightarrow |f(y) - f(x)| \leq \epsilon \\
&\leq \epsilon + 2 \cdot \|f\|_\infty \cdot C_\delta \\
\text{mit } C_\delta &= P_x\left[\left|\frac{S_n}{n} - x\right| > \delta\right] \underset{\substack{\uparrow \\ \text{Chebychev}}}{\leq} \frac{1}{\delta^2} \cdot \operatorname{var}\left(\frac{S_n}{n}\right) = \frac{x \cdot (1-x)}{\delta^2 \cdot n} \leq \frac{1}{4\delta^2 n} \to 0 \, (n \to \infty)
\end{aligned}
$$

$$
\Rightarrow \sup_x |f_n(x) - f(x)| \to 0 \quad (n \to \infty)
$$

Satz 10.2. $\forall f \in \mathcal{C}([0,1])$: *Die Bernstein-Polynome f_n konvergieren gleichmäßig gegen f.*

Korollar 10.3. Approximationssatz von Weierstraß

3. Monte Carlo-Integration

Sei $f \in \mathcal{L}^2([0,1], \mathcal{B}, \lambda_{|[0,1]})$.
$X_1, X_2, \ldots$ seien unabhängig, gleichverteilt in $[0,1]$ und
$I_n = \frac{1}{n}(f(X_1) + \ldots + f(X_n))$.

Satz 10.4. $\quad I_n \longrightarrow \int\limits_0^1 f(x)dx \quad$ *P-stochastisch*

Beweis:
Die ZV $f(X_i)_{i \in \mathbb{N}}$ sind unabhängig, identisch verteilt mit Erwartungswert

$$
\begin{aligned}
E(f(X_i)) &= \int f(X_i)dP = \int f(x)dP_{X_i}(x) = \int f(x)dx \qquad \text{und} \\
E(f(X_i)^2) &= \ldots = \int f^2(x)dx < \infty.
\end{aligned}
$$

Schwache Gesetz:

$$
I_n = \frac{1}{n}(f(X_1) + \ldots + f(X_n)) \longrightarrow E(f(X_i)) = \int f(x)dx \quad P\text{-stochastisch} \square
$$

Verschärfung des Schwachen Gesetzes (nun ohne $E(X_i^2) < \infty$).

Satz 10.5. $(X_i)_{i\in\mathbb{N}}$ *unabhängig, identisch verteilt mit*

$$x \cdot P(|X_1| > x) \to 0 \text{ für } x \to \infty.$$
$$S_n = \textstyle\sum_{i=1}^n X_i, \qquad \alpha_n = E(X_1 \cdot 1\!\!1_{\{|X_1|\le n\}})$$

Dann gilt: $$\frac{1}{n} S_n - \alpha_n \longrightarrow 0 \qquad P\text{-stochastisch}$$

Bemerkungen:

(1) $E(|X_1|) < \infty \Rightarrow x \cdot P(|X_1| > x) \le E(|X_1| \cdot 1\!\!1_{\{|X_1|>n\}}) \to 0$ und $\alpha_n \to E(X_1)$.

(2) Es gilt auch die Umkehrung: Wenn α_n existieren mit $\frac{1}{n}S_n - \alpha_n \to 0$ P-stochastisch, dann auch $x \cdot P(|X_1| > x) \to 0$, $x \to \infty$

Beweis:

1. Stutzen: $X_{n,k} = X_k \cdot 1\!\!1_{\{|X_k|\le n\}}$, $E(X_{n,k}) = \alpha_n$, $S_n' = \sum\limits_{k=1}^n X_{n,k}$
$\Rightarrow P(S_n \ne S_n') \le P(X_k \ne X_{n,k}$ für ein $k \le n) \le n \cdot P(|X_1| > n) \to 0$ nach Voraussetzung.

2.
$$\begin{aligned}
P\left(\left|\frac{1}{n}S_n' - \alpha_n\right| > \epsilon\right) &\le \frac{1}{\epsilon^2}\cdot \operatorname{var}\left(\frac{1}{n}S_n'\right) = \frac{1}{\epsilon^2 n}\cdot \operatorname{var} X_{n,1}\\
&\le \frac{1}{\epsilon^2 n}\cdot E(X_{n,1}^2) = \frac{1}{\epsilon^2 n}\cdot \int_0^\infty 2t\cdot P(|X_{n,1}| > t)dt\\
&= \frac{1}{\epsilon^2 n}\int_0^n 2tP(|X_1| > t)dt\\
&\le \underbrace{\frac{1}{\epsilon^2 n}\int_0^k 2tP(|X_1| > t)dt}_{\to 0 \text{ für } n\to\infty} + \frac{2}{\epsilon^2}\underbrace{\sup_{t>k} t\cdot P(|X_1| > t)}_{\to 0 \text{ für } k \to \infty} \to 0 \text{ für } n \to \infty \ (\forall \epsilon
\end{aligned}$$

3. Zusammen
$$P\left(\left|\frac{S_n}{n} - \alpha_n\right| > \epsilon\right) \le P\left(\left|\frac{S_n'}{n} - \alpha_n\right| > \epsilon\right) + P(S_n \ne S_n') \to 0 \ \ (\forall \epsilon > 0)\square$$

Bemerkung: Unkorreliertheit reicht nicht; geht verloren beim Stutzen.

Weitere Verschärfung

Satz 10.6. *$(X_n)_n$ unabhängig, $b_n > 0$ mit $b_n \to \infty$. Es gelte:*

- $\sum_{k=1}^{n} P(|X_k| > b_n) \to 0$ *und*
- $\frac{1}{b_n^2} \sum_{k=1}^{n} E(X_k^2 \cdot 1\!\!1_{\{|X_k| \leq b_n\}}) \to 0$ *für $n \to \infty$.*

Dann gilt für $S_n = \sum_{k=1}^{n} X_k$ und $a_n = \sum_{k=1}^{n} E(X_k \cdot 1\!\!1_{\{|X_k| \leq b_n\}})$

$$(S_n - a_n)/b_n \longrightarrow 0 \qquad \textit{stochastisch}$$

Beweis: Übung! Setze $X_{n,k} = X_k \cdot 1\!\!1_{\{|X_k| \leq b_n\}}$. Ansonsten wie eben. □

c) Das Starke Gesetz

Satz 10.7 (Kolmogorov). *$(X_i)_{i \in \mathbb{N}}$ seien paarweise unabhängig, identisch verteilt $\in \mathcal{L}^1$. Dann*

$$\frac{1}{n} S_n \longrightarrow E(X_1) \quad P\text{-f.s.}$$

Beweis: (nach Etemadi):

(1) o.B.d.A. $X_i \geq 0$ (ansonsten Übergang zu $(X_i^{\pm})_{i \in \mathbb{N}}$),

$Y_i = X_i \cdot 1\!\!1_{\{X_i \leq i\}}$ "Stutzen", $T_n = \sum_{i=1}^{n} Y_i$

$\Rightarrow P(X_i \neq Y_i$ für unendl. viele $i) = P(X_i = \infty) = 0$

$\Rightarrow$ genügt z.z. $\frac{1}{n} T_n \to E(X_1)$ P-f.s. mit $T_n = \sum_{i=1}^{n} Y_i$.

(2) Sei $\epsilon > 0, \alpha > 1$ und $k(n) = [\alpha^n]$ = ganzzahliger Anteil von α^n.

Chebyshev: $$\sum_{n=1}^{\infty} P(|T_{k(n)} - ET_{k(n)}| > \epsilon k(n)) \leq \frac{1}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{k(n)^2} \mathrm{var} T_{k(n)}$$

$$= \frac{1}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{k(n)^2} \sum_{i=1}^{k(n)} \mathrm{var} Y_i = \frac{1}{\epsilon^2} \sum_{i=1}^{\infty} \mathrm{var} Y_i \cdot \sum_{\substack{n=1 \\ k(n) \geq i}}^{\infty} \frac{1}{k(n)^2} \quad \text{da paarweise unabh.}$$

$$\leq \frac{1}{\epsilon^2} \cdot \frac{4\alpha^2}{1 - \alpha^{-2}} \cdot \sum_{i=1}^{\infty} \frac{1}{i^2} \cdot \mathrm{var} Y_i$$

denn $$\sum_{\substack{n=1 \\ k(n) \geq i}}^{\infty} \frac{1}{k(n)^2} \leq \sum_{\substack{n=1 \\ \alpha^n \geq i}}^{\infty} \frac{4}{\alpha^{2n}} = 4 \cdot \sum_{\substack{n=1 \\ n \geq \frac{\ln i}{\ln \alpha}}}^{\infty} (\alpha^{-2})^n \leq 4 \cdot \frac{\alpha^{+2} \cdot i^{-2}}{1 - \alpha^{-2}}$$

(3) Berechnung von $\sum\limits_{i=1}^{\infty} \frac{1}{i^2} \mathrm{var}(Y_i) \leq$

$$\begin{aligned}
\leq \sum_{i=1}^{\infty} \frac{1}{i^2} E(Y_i^2) &= \sum_{i=1}^{\infty} \frac{1}{i^2} \cdot \int_0^{\infty} 2t \cdot P(Y_i > t) dt \\
&= \sum_{i=1}^{\infty} \frac{1}{i^2} \int_0^{i} 2t \cdot P(X_1 > t) dt \\
&= \int_0^{\infty} \left(\sum_{\substack{i=1 \\ i \geq t}}^{\infty} \frac{1}{i^2} \right) \cdot 2t \cdot P(X_1 > t) dt \\
&\leq 4 \int_0^{\infty} P(X_1 > t) dt = 4E(X_1) < \infty \\
\text{denn: } \sum_{i \geq k} \frac{1}{i^2} &\leq \frac{1}{k^2} + \sum_{i \geq k} \frac{1}{i \cdot (i+1)} \\
&= \frac{1}{k^2} + \frac{1}{k} \sum_{i \geq k} \left(\frac{1}{i} - \frac{1}{i+1} \right) = \frac{1}{k^2} + \frac{1}{k} \leq \frac{2}{k} \ (\forall k \in \mathbb{N})
\end{aligned}$$

und daher $2t \cdot \sum\limits_{i \geq t} \frac{1}{i^2} \leq 4$. Also zusammen

$$\begin{aligned}
\sum_{n \in \mathbb{N}} P(|T_{k(n)} - E(T_{k(n)})| > \epsilon \cdot k(n)) &\leq \frac{1}{\epsilon^2} \cdot \frac{4\alpha^2}{1 - \alpha^{-2}} \sum_{i=1}^{\infty} \frac{1}{i^2} \mathrm{var} Y_i \\
&\leq \frac{1}{\epsilon^2} \cdot \frac{16\alpha^2}{1 - \alpha^{-2}} E(X_1) < \infty \qquad (\forall \epsilon > 0)
\end{aligned}$$

(4) Schließlich $E\left(\frac{1}{k} T_k\right) = \frac{1}{k} \sum\limits_{i=1}^{k} E(Y_i) \to E(X_1)$ für $k \to \infty$,
denn $E(Y_i) \to E(X_1)$ für $i \to \infty$ (monotone bzw. majorisierte Konvergenz)
Also $\forall \epsilon > 0$:

$$\begin{aligned}
&\sum_n P\left(\left| \frac{1}{k(n)} T_{k(n)} - E(X_1) \right| > \epsilon \right) \\
&\quad \leq \sum_n P\left(\left| \frac{1}{k(n)} T_{k(n)} - \frac{1}{k(n)} E(T_{k(n)}) \right| > \frac{\epsilon}{2} \right) \\
&\quad\quad + \# \left\{ n \in \mathbb{N} : \left| \frac{1}{k(n)} E(T_{k(n)}) - E(X_1) \right| > \frac{\epsilon}{2} \right\} < \infty
\end{aligned}$$

d.h. "schnelle stochastische Konvergenz" von $(T_{k(n)})_{n \in \mathbb{N}}$ gegen $E(X_1)$.
$\Rightarrow$ (mit folgendem Lemma): $\frac{1}{k(n)} T_{k(n)} \to E(X_1)$ P-f.s.

also "fast sichere Konvergenz" entlang der Teilfolge $(k(n))_{n\in\mathbb{N}}$

(5) Für die restlichen $k \in \mathbb{N} : \exists n : \; k(n) \le k < k(n+1)$

$$\Rightarrow \quad \frac{1}{\alpha} = \frac{1}{k(n+1)} T_{k(n)} \le \frac{1}{k} T_k \le \frac{1}{k(n)} T_{k(n+1)} = \alpha \cdot \frac{1}{k(n+1)} T_{k(n+1)}$$

$$\Rightarrow \quad \frac{1}{\alpha} E(X_1) \le \liminf_{k\to\infty} \frac{1}{k} T_k \le \limsup_{k\to\infty} \frac{1}{k} T_k \le \alpha \cdot E(X_1) \; P\text{-f.s.} \;\; (\forall \alpha > 1)$$

$$\Rightarrow \quad \frac{1}{k} T_k \to E(X_1) \; P\text{-f.s.} \square$$

Lemma 10.8. $(X_n)_{n\in\mathbb{N}}$ *bel. ZV mit*

$$\sum_{n=1}^{\infty} P(|X_n| \ge \epsilon) < \infty \qquad (\forall \epsilon > 0)$$

("schnelle stochastische Konvergenz")

$$\Longrightarrow \qquad X_n \to 0 \quad P\text{-}f.s.$$

Beweis: Wegen der Borel-Cantelli-Lemma folgt aus der Annahme

$$P(|X_n| \ge \epsilon \text{ für unendlich viele } n \in \mathbb{N}) = 0,$$

daß $\limsup_{n\to\infty} |X_n| \le \epsilon$ P-f.s., $\forall \epsilon > 0$. $\Rightarrow$ Behauptung. $\square$

Korollar 10.9. $(X_i)_{i\in\mathbb{N}}$ *(paarweise) unabhängig, identisch verteilt,* ≥ 0

$$\Rightarrow \; \frac{1}{n} \sum_{i=1}^{n} X_i \longrightarrow E(X_1) \in [0, \infty] \qquad P\text{-}f.s.$$

Beweis: Nur noch zu zeigen, falls $E(X_1) = +\infty$. In diesem Fall

$$\frac{1}{n} \sum_{i=1}^{n} X_i \ge \frac{1}{n} \sum_{i=1}^{n} (X_i \wedge N) \stackrel{n\to\infty}{\longrightarrow} E(X_1 \wedge N) \; P\text{-f.s. (für jedes } N \in \mathbb{N})$$

Aber $E(X_1 \wedge N) \stackrel{N\to\infty}{\longrightarrow} E(X_1) = +\infty$.

$$\Rightarrow \frac{1}{n} \sum_{i=1}^{n} X_i \stackrel{n\to\infty}{\longrightarrow} +\infty \; P\text{-f.s..} \qquad \square$$

Beispiel 10.10. $X_1, X_2, \dots$ unabhängig, Cauchy-verteilt $C(\alpha)$

$$\begin{aligned} &\Rightarrow \quad S_n \text{ nach } C(n\cdot\alpha)\text{-verteilt} \\ &\Rightarrow \quad \frac{1}{n} S_n \text{ nach } C(\alpha)\text{-verteilt} \end{aligned}$$

d.h. $\frac{1}{n}S_n$ denkt nicht daran, gegen irgendeine Zahl zu konvergieren!
Beachte: $E(|X_i|) = \int |x| dP_{X_i}(x) = \frac{1}{\pi} \int\limits_{-\infty}^{+\infty} \frac{\alpha\cdot|x|}{\alpha^2+x^2} dx = +\infty$
(und dann natürlich auch $E(X_i^2) = +\infty$)

Alternative Version des Starken Gesetzes (für nicht identisch verteilte ZV):

Satz 10.11. $(X_i)_{i\in\mathbb{N}}$ *unabhängig* $\in \mathcal{L}^1$ *mit* $E(X_i) = 0$, $b_i \nearrow \infty$
$\sum\limits_{i=1}^{\infty} \frac{1}{b_i^2} \mathrm{var} X_i < \infty$. *Dann gilt*

$$\frac{1}{b_n} \cdot S_n \longrightarrow 0 \qquad P\text{-}f.s.$$

Beweis: s. Bauer, Durrett
Typisch: $b_n = n \Rightarrow$ Kolmogorov-Kriterium: $\sum\limits_{i=1}^{\infty} \frac{1}{i^2} \mathrm{var} X_i < \infty$. □

d) Konvergenz der empirischen Verteilungen

Starke Gesetz sagt: Empirischer Mittelwert kann zur Schätzung des Erwartungswertes $E(X)$ verwendet werden.
Womit kann man die Verteilung P_X schätzen?

Definition 10.12. *Für ZV* $(X_i)_{i\in\mathbb{N}}$ *heißt die Abbildung*

$$\begin{aligned} \rho_n : \quad & \Omega \times \mathcal{B}(\overline{\mathbb{R}}) \longrightarrow [0,1] \\ & (\omega, A) \mapsto \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A(X_i(\omega)) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(\omega)}(A) \end{aligned}$$

empirische Verteilung *der ersten* n *Beobachtungen.*

Für jedes $\omega \in \Omega$ ist $\rho_n(\omega, .) = \frac{1}{n} \sum\limits_{i=1}^{n} \delta_{X_i(\omega)}$ ein W-Maß auf $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$.

Satz 10.13 (Glivenko-Cantelli). *Seien* $(X_i)_{i\in\mathbb{N}}$ *unabhängig, identisch verteilt und reell.*
Dann gilt für P*-fast alle* $\omega \in \Omega$:

$$\rho_n(\omega, .) \longrightarrow P_{X_1} \text{ schwach}$$

d.h. die empirische Verteilung konvergiert schwach gegen die Verteilung.

Beweis: Wir zeigen Konvergenz der "empirischen Verteilungsfunktion"

$$F_n(\omega, x) = \rho_n(\omega,]-\infty, x]) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\mathbb{1}_{]-\infty,x]}(X_i(\omega))}_{Y_i(\omega)\in\mathcal{L}^1}$$

gegen die "theoretische" Verteilungsfunktion

$$F(x) = P_{X_1}(]-\infty, x]) = P(\{X_1 \le x\})$$

an jeder Stetigkeitsstelle x von F.
Nun gilt $\forall x$ wege des Starken Gesetzes:

$$(*) \qquad F_n(\omega, x) \longrightarrow E(\mathbb{1}_{]-\infty,x]}(X_1(\omega))) = F(x) \qquad P\text{-f.s.},$$

d.h $\forall x : \exists N_x \in \mathcal{A}$, $P(N_x) = 0 : \forall \omega \in \Omega \setminus N_x :$ $\quad (*)$
Wähle $N = \bigcup\limits_{x \text{ rational}} N_x \Rightarrow P(N) = 0$ und $\forall \omega \in \Omega \setminus N$:
$\forall x, \forall s, r \in \mathbb{Q}$ mit $s \le x \le r$:

$$\begin{aligned} F(s) &= \lim_n F_n(\omega, s) \le \liminf_n F_n(\omega, x) \le \limsup_n F_n(\omega, x) \\ &\le \lim_n F_n(\omega, r) = F(r) \end{aligned}$$

$\Rightarrow$ falls F stetig in x: $F(x) = \lim\limits_n F_n(\omega, x)$. □

Einschub: Einige Paradoxa der W-Theorie

a) Petersburger Paradoxon

Die ZV X beschreibe die Auszahlung beim sogenannten Petersburger Spiel:
Sie bekommen 2^n DM, falls Sie bei einer Serie von Münzwürfen beim n-ten Wurf zum 1. Mal "Zahl" erhalten, also $X = 2^T$ mit T=Wartezeit auf "Zahl".
Was ist ein vernünftiger Eintrittsprreis c, um dieses Spiel spielen zu dürfen?
Kanonischer Kandidat:

$$E(X) = \sum_{n=1}^{\infty} 2^n \cdot P(X_i = 2^n) = +\infty$$

Aber: Niemand zahlt ∞ viel, um dieses Spiel zu spielen.
(Beachte: mit $p = \frac{1}{2}$ gewinnt man genau 2 DM,
$p = \frac{3}{4}$ gewinnt man max. 4 DM, ...)

Erklärung 1: Betrachte Folge $(X_i)_{i \in \mathbb{N}}$ unabhängig, identisch verteilt mit obiger Verteilung
Dann gilt zwar $\frac{1}{n} \sum_{i=1}^{n} X_i \to E(X) = +\infty$ P-f.s.
(was für $c = \infty$ spricht), aber auch

$$\underbrace{E(X) - \frac{1}{n} \sum_{i=1}^{n} X_i}_{\text{"gemittelter Verlust"}} = \infty \qquad \text{sicher}$$

Man kann zeigen:

$$\begin{array}{lll} \nexists\, \alpha_n : & \frac{1}{n} \cdot S_n - \alpha_n \to 0 & P\text{-f.s.} \\ \text{aber} & \frac{\log 2}{n \cdot \log n} S_n - 1 \to 0 & P\text{-stochastisch} \end{array}$$

Etwas gewagt!

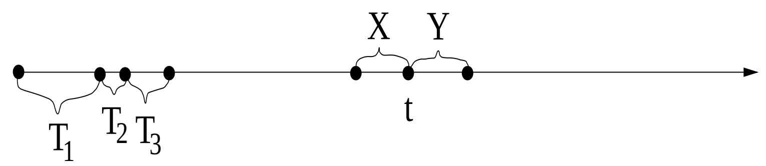
Erklärung 2: Das mathematische Modell ist unrealistisch, denn Ihr Mitspieler besitz maximal $M = 2^m$ DM
$\Rightarrow$ realistische ZV $\tilde{X} = X \wedge M$
$\Rightarrow E(\tilde{X}) = \sum_{n=1}^{m} 2^n \cdot P(X_i = 2^n) + 2^m \cdot P(X_i > 2^m) = m + 1$
Also: fairer Preis $c = m + 1$.
z.B. $M = 2$ Billionen DM $= 2^{41}$ DM ($\approx$ Staatsverschuldung der BRD)
$c = 42$ DM
(Eine weitere Beschränkung folgt aus der Anzahl der Spiele, die Sie in Ihrem Leben schaffen.)

Bemerkung: Auch unter dieser Modifikation ist das Spiel sehr riskant, denn $E(\tilde{X}^2) = \sum_{n=1}^{m} 2^{2n} \cdot 2^{-n} + 2^{2m} \cdot 2^{-m} = 3 \cdot 2^m - 2 \Rightarrow \text{var}\tilde{X} \approx 3 \cdot M$
Chebyshev: $P\left(\left|\frac{1}{n}\tilde{S}_n - E(\tilde{X})\right| \geq \epsilon\right) \leq \frac{1}{\epsilon^2 \cdot n} \cdot 3 \cdot M$ ($\epsilon = 10$)
d.h. ungefähr $n = M = 2^{41}$ Spiele, um einigermaßen sicher durchschnittlich $c = 42$ DM zu gewinnen.

b) Das Wartezeitenproblem

An einer Bushaltestelle fahren im Mittel pro Stunde α Busse ab.

Die Wartezeiten $T_1, T_2, \ldots$ zwischen den aufeinanderfolgenden Bussen seien unabhängig Exp(α)-verteilt ($\Rightarrow E(T_i) = \frac{1}{\alpha}$)



Wie lange muß man im Mittel auf einen Bus warten, falls man zu einer bestimmten Zeit t kommt?

Antwort (1): $\frac{1}{2}\frac{1}{\alpha}$, denn $X + Y = T_k$ für ein $k \in \mathbb{N}$ und $X \approx Y$

Antwort (2): $\frac{1}{\alpha}$, denn X spielt keine Rolle (Gedächtnislosigkeit der Exponentialverteilung)

(2) ist richtig! (1) falsch.

Satz 10.14. *a) X und Y unabhängig*

b) Y Exp(α)*-verteilt, also $E(Y) = \frac{1}{\alpha}$*

c) $E(X) = \frac{1}{\alpha}(1 - e^{-\alpha t})$

Also: Typischerweise erwischt man bei gegebenen t ein $k = k(t) \in \mathbb{N}$ mit $T_{k(t)}$ groß, nämlich

$$E(X) + E(Y) = E(T_{k(t)}) = \frac{1}{\alpha}(2 - e^{-\alpha t}) \sim \frac{2}{\alpha}$$

Beweis: Bestimme gemeinsame Verteilung von X und Y.
Sei hierzu $S_n = T_1 + \ldots + T_n$, ist $\Gamma(n, \alpha)$-verteilt

$$\Rightarrow \qquad P(S_n \leq x) = \int_0^x f_n(s)ds \text{ mit } f_n(s) = \frac{\alpha^n}{(n-1)!} s^{n-1} \cdot e^{-\alpha s}.$$

$$\text{Beachte:} \qquad \sum_{n=1}^{\infty} f_n(s) = \alpha \qquad (\forall s)$$

Betrachte nun für $x \in [0,t]$ und $y \in [0,\infty[$

$$\begin{aligned}
P(X \geq x, Y \geq y) &= P(T_1 \geq y+t) + \sum_{n=1}^{\infty} P(S_n \leq t-x, T_{n+1} \geq y+t-S_n) \\
&= e^{-\alpha(t+y)} + \sum_{n=1}^{\infty} P_{(S_n,T_{n+1})}(\{(s,r) : s \leq t-x, r \geq y+t-s\}) \\
&= e^{-\alpha(t+y)} + \int\int_{\left\{(s,r):\ \substack{s \leq t-x,\\ r \geq y+t-s}\right\}} f_n(s) \cdot \alpha \cdot e^{-\alpha r} ds dr \\
&= e^{-\alpha(t+y)} + \int_{\{s \leq t-x\}} \sum_{n=1}^{\infty} f_n(s) \cdot e^{-\alpha(y+t-s)} ds \\
&= e^{-\alpha(t+y)} \cdot (1 + \alpha \cdot \int_0^{t-x} e^{\alpha s} ds) = e^{-\alpha y} \cdot e^{-\alpha x}
\end{aligned}$$

$\overset{x=0}{\Rightarrow} P(Y \geq y) = e^{-\alpha y}$ $(\forall y \geq 0)$, d.h. Y ist Exp(α)-verteilt

$\overset{y=0}{\Rightarrow} P(X \geq x) = \begin{cases} e^{-\alpha x} & , \ 0 \leq x \leq t \\ e^{-\alpha t} & , \ x > t \end{cases}$ “gestauchte“ Exp(α)-Verteilung

$$\Rightarrow E(X) = \int_0^t x \cdot \alpha \cdot e^{-\alpha x} dx + t \cdot e^{-\alpha t} = \frac{1}{\alpha}(1 - e^{-\alpha t})$$

Ferner folgt: X, Y unabhängig. $\square$

c) Simpsons Paradoxon

Vereinfacht:

	$M = ♂$		$W = ♀$	
A	90	40%	10	80%
B	10	4%	90	8%
Ges	100	36.4%	100	15.2%

genauer:

A	n_{AM}	p_{AM}	n_{AW}	p_{AW}
B	n_{BM}	p_{BM}	n_{BW}	p_{BW}
Ges	n_M	p_M	n_W	p_W

$$\begin{aligned}
n_M &= n_{AM} + n_{BM} \\
p_M &= \frac{p_{AM} \cdot n_{AM} + p_{BM} \cdot n_{BM}}{n_M} = \frac{90 \cdot 0.4 + 10 \cdot 0.04}{100} = 36.4 \\
p_W &= \ldots = \frac{10 \cdot 0.8 + 90 \cdot 0.08}{100} = 15.2
\end{aligned}$$

M.a.W. $P(\text{Erfolg}|A \cap M) < P(\text{Erfolg}|A \cap W)$

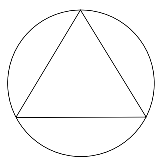
$P(\text{Erfolg}|B \cap M) < P(\text{Erfolg}|B \cap W)$

aber $P(\text{Erfolg}|M) > P(\text{Erfolg}|W)$

Erklärung:
Aus $\frac{m_a}{M_a} < \frac{w_a}{W_a}$ und $\frac{m_b}{M_b} < \frac{w_b}{W_b}$ folgt nicht $\frac{m_a+m_b}{M_a+M_b} < \frac{w_a+w_b}{W_a+W_b}$.

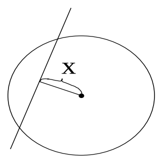
d) Bertrands Paradoxon

Bertrand 1889:
Gegeben ein Kreis und ein eingeschriebenes gleichseitiges Dreieck



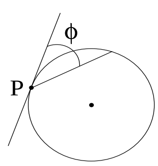
Gesucht P(zufällig gewählte Sehne des Kreises > Dreiecksseite)

Lösung 1. Sehne $S = S(x, \phi)$ mit x= Abstand zu 0= Mittelpunkt
mit $x \in [0, r]$, $\phi \in [0, 2\pi]$



$$P(\ldots) = P(x \leq \frac{r}{2}) = \frac{1}{2}$$

Lösung 2. Sehne $S = S(\phi, P)$ mit $\phi = \sphericalangle_p$ (Sehne, Tangente) $\in [0, \frac{\pi}{2}]$, P= Sphäre



$$\Rightarrow P(\ldots) = P(\phi \in [\frac{\pi}{3}, \frac{\pi}{2}]) = \frac{1}{3}$$

Kapitel 11

Große Abweichungen

Stets $(\Omega, \mathcal{A}, P)$ W-Raum, $(X_i)_{i\in\mathbb{N}}$ unabhängig, identisch verteilte, integrierbDare ZV mit Verteilung $\mu = P_{X_i}$ und $\alpha = E(X_i)$.

$$S_n = \sum_{i=1}^{n} X_i$$

a) Vorüberlegung

Das Gesetz der GZ besagt: $\frac{1}{n}S_n \longrightarrow \alpha$ P-f.s. für $n \to \infty$.

Aber: keine Aussage über Konvergenzgeschwindigkeit (vgl. Petersburger Paradoxon)
Falls $X_i \in \mathcal{L}^2$ (und nicht nur $\in \mathcal{L}^1$), so besagt

$$\text{Chebyshev:} \qquad P\left(\left|\frac{1}{n}S_n - \alpha\right| \geq \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2} \cdot \frac{1}{n} \qquad (\forall n \in \mathbb{N})$$

mit $\sigma = \text{var} X_i = E(X_i^2) - \alpha^2 = \int x^2 \mu(dx) - \alpha^2$.

Also: Konvergenz wie $\frac{1}{n}$.

Tatsächlich ist die Konvergenz meist wesentlich schneller.
Wir werden (unter stärkeren Voraussetzungen) zeigen.
Die Wahrscheinlichkeit großer Abweichungen (engl. large deviation) ist exponentiell klein, d.h.

$$P\left(\left|\frac{1}{n}S_n - \alpha\right| \geq \epsilon\right) \sim e^{-I_\mu(\epsilon)\cdot n} \text{ für } n \to \infty$$

mit geeigneter Ratenfunktion $I_\mu : \mathbb{R} \longrightarrow [0, \infty]$ bestimmt durch $\mu = P_{X_i}$.

b) Die Raten-Funktion

Wir wählen als Raten-Funktion die "Cramér-Transformierte von μ"

$$I_\mu(x) = \sup_{t\in\mathbb{R}}[tx - \log\check{\mu}(t)] \quad (\in \overline{\mathbb{R}})$$

("Legendre-Transformierte von $\log\check{\mu}(t)$"), wobei

$$\check{\mu}(t) = \int e^{tx}\mu(dx) = E(e^{tX_i})$$

die (verallgemeinerte) Laplace-Transformierte von μ ist.
Offenbar ist $\check{\mu}(t) \in]0,\infty]$ $(\forall t \in \mathbb{R})$ und $\check{\mu}(0) = 1 \Rightarrow I_\mu(x) \in [0,\infty]$ $(\forall x \in \mathbb{R})$.
Wegen Konvexität von $x \mapsto e^{tx}$ folgt aus Jensen-Ungleichung

$$\check{\mu}(t) = E(e^{tX_i}) \geq e^{tE(X_i)} = e^{t\alpha}$$

und damit

$$t\cdot\alpha - \log\check{\mu}(t) \leq 0 \quad (\forall t \in \mathbb{R}).$$

Also: $\quad I_\mu(\alpha) = 0.$

Beispiel 11.1. $\mu = \delta_\alpha \quad \Rightarrow \quad I_\mu(x) = \begin{cases} 0 & , \ x = \alpha \\ +\infty & , \ x \neq \alpha \end{cases}$
denn $\check{\mu}(t) = e^{t\alpha}$

Beispiel 11.2. $\mu = \mathrm{Exp}(\lambda)$, $\lambda > 0$

$$\begin{aligned} \Rightarrow \check{\mu}(t) &= \int_0^\infty e^{tx}\cdot\lambda e^{-\lambda x}dx = \begin{cases} +\infty & , \ t \geq \lambda \\ \frac{\lambda}{\lambda - t} & , \ t < \lambda \end{cases} \\ \Rightarrow I_\mu(x) &= \sup_{t<\lambda}\left[tx + \log\frac{\lambda - t}{\lambda}\right] \underset{\substack{\uparrow \\ t = \lambda - \frac{1}{x}}}{=} \begin{cases} \lambda x - 1 - \log(\lambda x) & , \ x > 0 \\ +\infty & , \ x \leq 0 \end{cases} \end{aligned}$$

Lemma 11.3.
(i) $\log\check{\mu}$ *ist konvex, ebenso* I_μ.

(ii) I_μ *ist antiton auf* $]-\infty,\alpha]$ *und isoton auf* $[\alpha,\infty[$

(iii) Falls $\check{\mu}(t) < \infty$ für alle $t \in [-t_0, t_0]$ (für ein $t_0 > 0$), so ist

$$I_\mu(x) \longrightarrow \infty \text{ für } |x| \to \infty.$$

Falls $\check{\mu}(t) < \infty$ für alle $t \in \mathbb{R}$, so ist

$$\frac{I_\mu(x)}{|x|} \longrightarrow \infty \text{ für } |x| \to \infty.$$

(iv) Falls $\check{\mu}(t) < \infty$ $(\forall t \in \mathbb{R})$, so ist $\check{\mu} \in \mathcal{C}^\infty(\mathbb{R})$.

Beweis:

(i) Konvexität von $\log \check{\mu}$ folgt aus der Hölder-Ungleichung:

$$\begin{aligned}
\log \check{\mu}(\lambda t_1 + (1-\lambda)t_2) &= \log \int e^{\lambda t_1 \cdot x} \cdot e^{(1-\lambda)t_2 \cdot x} \mu(dx) \\
&\leq \log \left(\left[\int e^{t_1 x} \mu(dx) \right]^\lambda \cdot \left[\int e^{t_2 x} \mu(dx) \right]^{1-\lambda} \right) \\
&= \lambda \cdot \log \check{\mu}(t_1) + (1-\lambda) \cdot \log \check{\mu}(t_2) \quad (\forall \lambda \in [0,1]) \\
\lambda I(x_1) + (1-\lambda) I(x_2) &= \sup_{t_1, t_2} (\lambda [x_1 t_1 - \log \check{\mu}(t_1)] + (1-\lambda)[x_2 t_2 - \log \check{\mu}(t_2)]) \\
&\geq \sup_t ([\lambda x_1 + (1-\lambda) x_2] t - \log \check{\mu}(t)) = I(\lambda x_1 + (1-\lambda) x_2)
\end{aligned}$$

(ii) folgt aus der Konvexität von I_μ und aus $I_\mu(\alpha) \geq 0 = I_\mu(a)$.

(iii) Aus $\check{\mu}(t_0) < \infty$ folgt

$$\begin{aligned}
\liminf_{x \to \infty} \frac{I(x)}{x} &= \liminf_{x \to \infty} \; [sup_t (tx - \log \check{\mu}(t))/x] \\
&\geq \liminf_{x \to \infty} \frac{t_0 x - \log \check{\mu}(t_0)}{x} = t_0 \\
\Rightarrow \lim_{x \to \infty} I(x) = +\infty \quad &, \quad \text{analog } \lim_{x \to -\infty} I(x) = +\infty.
\end{aligned}$$

Mit selbem Argument folgt nun $\frac{I(x)}{|x|} \to \infty$ für $|x| \to \infty$, falls $\check{\mu} < \infty$ auf $\mathbb{R}$.

(iv) Übung!

c) Der Satz von Cramér

Proposition 11.4. *$\forall n \in \mathbb{N}$:*

$$\begin{aligned}
\forall x \geq \alpha: \quad & P\left(\frac{1}{n}S_n \geq x\right) \leq \exp(-n \cdot I_\mu(x)) \\
\forall x \leq \alpha: \quad & P\left(\frac{1}{n}S_n \leq x\right) \leq \exp(-n \cdot I_\mu(x))
\end{aligned}$$

Beweis: Es genügt eine der beiden letzten Ungleichungen zu beweisen. O.B.d.A. ist hierbei $x = 0$ und $\alpha < 0$.
(Ansonsten: $X_n \to -X_n$ bzw. $X_n \to X_n - x$.)
Dann gilt $\forall t \geq 0$:

$$\begin{aligned}
P\left(\frac{1}{n}S_n \geq 0\right) &= P(X_1 + \ldots + X_n \geq 0) = P(e^{t(X_1+\ldots+X_n)} \geq 1) \\
&\leq E(e^{t(X_1+\ldots+X_n)}) = \int \ldots \int e^{tx_1} \cdot \ldots \cdot e^{tx_n} \mu(dx_1) \ldots \mu(dx_n) \\
&= \check{\mu}(t)^n \\
\Rightarrow \log P\left(\frac{1}{n}S_n \geq 0\right) &\leq n \cdot \log \check{\mu}(t) \qquad (\forall t \geq 0) \\
\Rightarrow -\frac{1}{n}\log P\left(\frac{1}{n}S_n \geq 0\right) &\geq \sup_{t \geq 0}(-\log \check{\mu}(t)) = I_\mu(0) \\
& \quad \text{denn } \forall t < 0: -\log \check{\mu}(t) \leq \underbrace{t \cdot \alpha}_{>0} - \log \check{\mu}(t) \leq I_\mu(\alpha) = 0
\end{aligned}$$

$\Rightarrow$ Behauptung. □

Satz 11.5 (von Cramér). *Sei $\check{\mu}(t) < \infty$ für alle $t \in \mathbb{R}$. Dann gilt für alle Borelmengen $B \subset \mathbb{R}$:*

$$\begin{aligned}
-\inf_{x \in \mathring{B}} I(x) &\leq \liminf_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n}S_n \in B\right) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n}S_n \in B\right) \leq -\inf_{x \in \bar{B}} I(x)
\end{aligned}$$

(mit $\mathring{B}$ = Inneres von B, $\bar{B}$ = Abschluß von B).

Beweis: Obere Abschätzung bereits bewiesen:
Nämlich $\alpha \notin B \Rightarrow \exists \alpha_1 < \alpha < \alpha_2,\ \alpha_1, \alpha_2 \in \partial B,\ B \subset]-\infty, \alpha_1] \cup [\alpha_2, \infty[$

$$\begin{aligned}
\Rightarrow P\left(\frac{1}{n}S_n \in B\right) &\leq \exp(-n \cdot I(\alpha_1)) + \exp(-n \cdot I(\alpha_2)) \\
&\leq 2 \cdot \exp(-n \cdot \inf_{x \in \bar{B}} I(x))
\end{aligned}$$

Für untere Abschätzung zeigen wir: $\forall x_0 \in \mathbb{R}$, $\forall \delta > 0$:

$$\liminf_{n\to\infty} \frac{1}{n} \log P\left(\left|\frac{1}{n}S_n - x_0\right| < \delta\right) \geq -I(x_0) \qquad (*)$$

Denn daraus folgt $\forall G$ offen $\subset \mathbb{R}$:

$$\liminf_{n\to\infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \in G\right) \geq -\inf_{x\in G} I(x), \qquad (**)$$

und hieraus natürlich die Behauptung.

Zum Beweis von $(*)$ für fixes $x_0 \in \mathbb{R}$:

1. Fall: $\exists t \geq 0 : I(x_0) = tx_0 - \log \breve{\mu}(t)$.
Betrachte das W-Maß

$$\mu_t(dx) = \frac{1}{\breve{\mu}(t)} \cdot e^{tx} \cdot \mu(dx) \qquad \text{auf } (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

Für $(\tilde{X}_i)_{i\in\mathbb{N}}$ unabhängig, identisch verteilt nach μ_t dürfen wir erwarten, daß $\frac{1}{n}\sum\limits_{i=1}^{n} \tilde{X}_i \to x_0$.

Denn: $\tilde{E}(|\tilde{X}_i|) = \dfrac{1}{\breve{\mu}(t)} \cdot \displaystyle\int_{-\infty}^{+\infty} |x| \cdot e^{tx}\mu(dx) < \infty$

und $\tilde{E}(\tilde{X}_i) = \dfrac{1}{\breve{\mu}(t)} \displaystyle\int_{-\infty}^{+\infty} x \cdot e^{tx}\mu(dx) = \frac{d}{ds}(\log \breve{\mu}(s))_{|s=t}$

Nun gilt nach Annahme: $s \mapsto sx_0 - \log \breve{\mu}(s)$ nimmt Maximum an der Stelle $s = t$ an.
Also $\frac{d}{ds}(sx_0 - \log \breve{\mu}(s))_{|s=t} = 0$ und damit

$$\tilde{E}(\tilde{X}_i) = x_0$$

Nach dem schwachen Gesetz gilt

$$\tilde{P}\left(\left|\frac{1}{n}\tilde{S}_n - x_0\right| < \delta\right) \to 1 \qquad (n \to \infty)$$

(Also: das für μ-verteilte ZV Ereignis einer großen Abweichung wird für μ_t-verteilte ZV zu einem höchstwahrscheinlichem Ereignis)
Nun ist für $f : \mathbb{R} \to \overline{\mathbb{R}}_+$ meßbar

$$\begin{aligned}
\int f(\tilde{S}_n)d\tilde{P} &= \int \ldots \int f(x_1 + \ldots + x_n)\mu_t(dx_1)\ldots\mu_t(dx_n) \\
&= \int \ldots \int \frac{e^{t(x_1+\ldots+x_n)}}{\breve{\mu}(t)^n} \cdot f(x_1 + \ldots + x_n)\mu(dx_1)\ldots\mu(dx_n) \\
&= \breve{\mu}(t)^{-n} \cdot \int e^{t\cdot S_n} \cdot f(S_n)dP
\end{aligned}$$

und daher (wegen $t \geq 0$) für alle $n \in \mathbb{N}$

$$\begin{aligned}
P\left(\left|\frac{1}{n}S_n - x_0\right| < \delta\right) &= \int\limits_{\{|\frac{1}{n}S_n - x_0|<\delta\}} dP \\
&\geq \int\limits_{\{|\frac{1}{n}S_n - x_0|<\delta\}} \exp(t(\underbrace{S_n - n\cdot(x_0+\delta)}_{<0}))dP \\
&= \tilde{P}\left(\left|\frac{1}{n}\tilde{S}_n - x_0\right| < \delta\right)\cdot \check{\mu}(t)^n \cdot \exp(-tn\cdot(x_0+\delta)) \\
&= \tilde{P}\left(\left|\frac{1}{n}\tilde{S}_n - x_0\right| < \delta\right)\cdot \exp(-n\cdot I(x_0) - n\cdot t\delta)
\end{aligned}$$

$$\Rightarrow \qquad \liminf_{n\to\infty}\frac{1}{n}\log P\left(\left|\frac{1}{n}S_n - x_0\right| < \delta\right) \geq -I(x_0) - t\delta$$

(Beachte: links isoton in δ, rechts antiton)

$$\Rightarrow \qquad \liminf_{n\to\infty}\frac{1}{n}\log P\left(\left|\frac{1}{n}S_n - x_0\right| < \delta\right) \geq -I(x_0),$$

$\Rightarrow$ Behauptung.

<u>2. Fall</u>: $\exists t \leq 0$: analog

<u>3. Fall</u>: $\forall t \in \mathbb{R} : I(x_0) > tx_0 - \log\check{\mu}(t)$
Wegen $I(x_0) = \sup\limits_{t\in\mathbb{R}}(\underbrace{tx_0 - \log\check{\mu}(t)}_{\text{konkav,endlich,stetig}})$ folgt: $\exists t_n \to +\infty$ oder $\exists t_n \to -\infty$,
o.B.d.A.:

$$\exists t_n \to \infty \text{ mit } t_n x_0 - \log\check{\mu}(t_n) \to I(x_0) \;\Rightarrow \int\limits_{\{X_1<x_0\}} \exp(t_n(X_1 - x_0))dP \to 0$$

$$\text{und } \int \exp(t_n(X_1 - x_0))dP = \exp(-(t_n x_0 - \log\check{\mu}(t_n))) \to \exp(-I(x_0))$$

Also zusammen

$$\begin{aligned}
&\int\limits_{\{X_1\geq x_0\}} \exp(t_n(X_1 - x_0))dP \to \exp(-I(x_0)) \\
\Rightarrow\;& P(X_1 > x_0) = 0 \text{ und } P(X_1 = x_0) = \exp(-I(x_0)) \\
\Rightarrow\;& P\left(\frac{1}{n}S_n = x_0\right) \geq \exp(-n\cdot I(x_0)) \quad (\forall n) \\
\Rightarrow\;& \liminf_{n\to\infty}\frac{1}{n}\log P\left(\left|\frac{1}{n}S_n - x_0\right| < \delta\right) \geq -I(x_0)\square
\end{aligned}$$

d) Entropie

Ziel: Alternative Charakterisierung der Raten-Funktion

Gegeben: W-Maße μ, ν auf $(\Omega, \mathcal{A})$.

Definition 11.6.

$$\mathrm{H}(\nu|\mu) = \begin{cases} \int \log \frac{d\nu}{d\mu} d\nu & , \textit{ falls } \nu \ll \mu \\ +\infty & , \textit{ sonst} \end{cases}$$

heißt **relative Entropie** *von ν bezgl. μ.*

Lemma 11.7.
(i) $\mathrm{H}(\nu|\mu) \in [0, \infty]$

(ii) $\mathrm{H}(\nu|\mu) < \infty \Leftrightarrow \exists f : \nu = f \cdot \mu$ *und* $\int f \log f d\mu < \infty$ *(mit* $0 \cdot \log 0 = 0$*)*

(iii) $\mathrm{H}(\nu|\mu) = 0 \Leftrightarrow \nu = \mu$

Beweis: $\mathrm{H}(\nu|\mu) < \infty \Rightarrow \exists f : \nu = f \cdot \mu$ und

$$\mathrm{H}(\nu|\mu) = \int \log f d\nu = \int f \cdot \log f d\mu \geq \int (f-1) d\mu = 0$$

wobei die Ungleichung aus $\log x \geq 1 - \frac{1}{x}$ folgt. ($\Rightarrow$ (i) & (ii)). Ferner gilt

$$\begin{aligned} & \mathrm{H}(\nu|\mu) = 0 \\ \Rightarrow\ & f \cdot \log f = f - 1 \quad \mu\text{-f.ü.} \\ \Rightarrow\ & f = 1 \quad \mu\text{-f.ü.} \\ \Rightarrow\ & \nu = \mu \end{aligned}$$

□

Beispiel 11.8. *em* $\mu = N(\alpha, \sigma^2)$, $\nu = N(\beta, \tau^2)$

$$\begin{aligned} \Rightarrow\ & f(x) = \frac{d\nu}{d\mu}(x) = \frac{\sigma}{\tau} \cdot \exp\left(-\frac{(x-\beta)^2}{2\tau^2} + \frac{(x-\alpha)^2}{2\sigma^2} \right) \\ \Rightarrow\ & H(\nu|\mu) = \log \frac{\sigma}{\tau} + \frac{1}{2\sigma^2}(\tau^2 \cdot \sigma^2 + (\alpha - \beta)^2) \end{aligned}$$

Also: $H(\nu|\mu)$ = "Abstand zwischen ν und μ", jedoch nicht symmetrisch.

Bemerkung:

(i) Ist $\nu = \sum\limits_{k\in\mathbb{Z}} p_k\delta_k$ und $\mu = \sum\limits_{k\in\mathbb{Z}} q_k\delta_k$, so ist

$$\mathrm{H}(\nu|\mu) = \sum \left(\log\frac{p_k}{q_k}\right) p_k, \text{ falls } \frac{p_k}{q_k} < \infty \;\; \forall k, \text{ und } = +\infty \text{ sonst}$$

(ii) Ist ν ein W-Maß auf $\mathbb{Z}$, so definiert man die **Entropie** von ν durch

$$\begin{aligned}\mathrm{H}(\nu) &= -\sum_{k\in\mathbb{Z}} p_k\log p_k \text{ mit } p_k = \nu(\{k\}) \text{ und mit } 0\cdot\log 0 = 0.\\ &= -\mathrm{H}(\nu|\sum\delta_k) \geq 0, \text{ denn } p_k\log p_k \geq 0 \; (\forall p_k \in [0,1])\end{aligned}$$

(iii) Unter allen W-Maßen ν auf $\{1,\ldots,n\}$ hat die Gleichverteilung $\nu_0 = \frac{1}{n}\sum\limits_{k=1}^{n}\delta_k$ <u>maximale Entropie</u>

<u>Denn</u>: $\mathrm{H}(\nu) = -\mathrm{H}(\nu|\nu_0) + \mathrm{H}(\nu_0) \leq \mathrm{H}(\nu_0) \quad \forall\nu$ und $= \mathrm{H}(\nu_0) \Leftrightarrow \nu = \nu_0$

(iv) Unter allen W-Maßen ν auf $N = \{1,2,\ldots\}$ mit vorgegeben Erwartungswert (EW) $\sum\limits_{k=1}^{\infty} k\cdot p_k = \alpha$ hat die geometrische Verteilung $\nu_0 = \sum\limits_{k\in\mathbb{N}}(1-p)^{k-1}\cdot p\cdot\delta_k$ mit $p = \frac{1}{\alpha}$ max. Entropie.

<u>Denn</u>: $$\begin{aligned}\mathrm{H}(\nu) - \mathrm{H}(\nu_0) &= -\sum p_k\log p_k + \sum(1-p)^{k-1}\cdot p[(k-1)\log(1-p) + \log p]\\ &= -\sum p_k\log p_k + (1-p)\log(1-p)\cdot\frac{1}{p} + \log p\\ &= -\sum p_k\log p_k + \log(1-p)\cdot\underbrace{\sum p_k\cdot(k-1)}_{=\frac{1}{p}-1} + \log p\cdot\underbrace{\sum p_k}_{=1}\\ &= -\sum p_k\log p_k + \sum p_k\log((1-p)^{k-1}\cdot p) = -\mathrm{H}(\nu|\nu_0) \leq 0\\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \nu = \nu_0 \Leftrightarrow = 0\end{aligned}$$

(v) Ist $\nu = f\cdot\lambda^1$ ein absolut-stetiges Maß auf $\mathbb{R}$ mit beschränkter Dichte, so definiert man die "Entropie" von ν durch

$$\mathrm{H}(\nu) = -\int_{-\infty}^{+\infty} f(x)\log f(x)dx = -\mathrm{H}(\nu|\lambda).$$

(vi) Unter allen absolut-stetigen W-Maßen ν auf $\mathbb{R}_+$ mit EW $= \alpha$ hat die Verteilung $\mathrm{Exp}(\alpha)$ maximaler Entropie.

(vii) Unter allen absolut-stetigen W-Maßen auf $\mathbb{R}_+$ mit EW$=\alpha$ und Var$=\sigma^2$ hat die Normalverteilung ν_{α,σ^2} maximale Entropie.

Lemma 11.9. *Sei μ W-Maß mit $\check{\mu}(t)<\infty$ $(\forall t\in\mathbb{R})$. Dann*

$\forall t, \forall$ W-Maße ν

$$\begin{aligned} H(\nu|\mu)-t\cdot\int x\cdot\nu(dx)+\log\check{\mu}(t) &= H(\nu|\mu_t)\geq 0 \\ \textit{und} \quad \text{“}&=0\text{“} \Leftrightarrow \nu=\mu_t \end{aligned}$$

Beweis: o.B.d.A. $\mathrm{H}(\nu|\mu)<\infty\Rightarrow\nu\ll\mu\Rightarrow\nu\ll\mu_t$

$$\begin{aligned} \Rightarrow \mathrm{H}(\nu|\mu) &= \int\log\frac{d\nu}{d\mu}d\nu \\ &= \int\log\frac{d\nu}{d\mu_t}d\nu+\int\overbrace{\log\frac{d\mu_t}{d\mu}}^{=\log\frac{1}{\check{\mu}(t)}e^{tx}=tx-\log\check{\mu}(t)}d\nu \\ &= \underbrace{\mathrm{H}(\nu|\mu_t)}_{\geq 0 \text{ und } =0\Leftrightarrow\nu=\mu_t}+t\cdot\int xd\nu-\log\check{\mu}(t).\square \end{aligned}$$

Satz 11.10. *μ wie oben*

(*i*) $H(\mu_t|\mu)=\min\{H(\nu|\mu):\nu$ *W-Maß mit* $\int xd\nu=\int xd\mu_t=:\alpha_t\}$

d.h. μ_t hat minimale relative Entropie unter allen W-Maßen mit gleichem EW

(*ii*) $H(\mu_t|\mu)=\sup_s[s\cdot\alpha_t-\log\check{\mu}(s)]=I_\mu(\alpha_t)$

Beweis:

(i) folgt sofort aus Lemma 11.9

(ii) Nach Lemma 11.9 (mit s statt t und μ_t statt ν):

$$\mathrm{H}(\mu_t|\mu)=s\cdot\alpha_t-\log\check{\mu}(s)+\mathrm{H}(\mu_t|\mu_s).\square$$

e) Satz von Sanov

Satz von Cramér = Prinzip der Großen Abweichung vom Starken Gesetz der GZ

Satz von Sanov = Prinzip der Großen Abweichung vom Satz von Glivenko-Cantelli

Gegeben: $(\Omega, \mathcal{A}, P), (X_i)_{i\in\mathbb{N}}$ unabhängig, identisch verteilt, reell mit Verteilung $\mu = P_{X_i}$.
Sei

$$\rho_n(\omega, .) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)} \in \mathcal{M}_1(\mathbb{R}) = \{\text{W-Maße auf } (\mathbb{R}, \mathcal{B}(\mathbb{R}))\}$$

die empirische Verteilung $(\forall\omega \in \Omega)$.
Glivenko-Cantelli besagt: Für P-fast alle $\omega \in \Omega$: $\rho_n(\omega, .) \to \mu$ schwach.

Satz 11.11 (Satz von Sanov). *$\forall B \subset \mathcal{M}_1(\mathbb{R})$:*

$$\begin{aligned} -\inf_{\nu\in\mathring{B}} H(\nu|\mu) &\le \liminf_{n\to\infty} \frac{1}{n}\log P(\rho_n \in B) \\ &\le \limsup \frac{1}{n}\log P(\rho_n \in B) \le -\inf_{\nu\in\bar{B}} H(\nu|\mu) \end{aligned}$$

$\mathring{B}, \bar{B}$ = Inneres bzw. Abschluß von B in $\mathcal{M}_1$ bzgl. schwacher Topologie!

<u>Ohne Beweis</u>! (s.Deuschel/Stroock)

Teil IV

Grenzwertsätze

Kapitel 12

Fourier-Transformierte und charakteristische Funktionen

Gegeben: W-Raum $(\Omega, \mathcal{A}, P)$, reelle ZV X, W-Maß μ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $\mu = P_X$.

Definition 12.1. Fourier-Transformierte *von μ bzw.* **charakteristische Funktion** *von X heißt die Abbildung $\mathbb{R} \longrightarrow \mathbb{C}$ definiert durch*

$$\begin{aligned} \widehat{\mu}(t) &= \int e^{itx}\mu(dx) = \int \cos(tx)\mu(dx) + i \cdot \int \sin(tx)\mu(dx) \\ &= E(e^{itX}) = E(\cos tX) + i \cdot E(\sin tX) \\ &= \varphi_X(t). \end{aligned}$$

Satz 12.2 (von Bochner). *Eine Funktion $\varphi : \mathbb{R} \longrightarrow \mathbb{C}$ ist die Fourier-Transformierte $\widehat{\mu}$ eines W-Maßes μ auf $\mathbb{R}$ genau dann, wenn gilt:*

(i) $|\varphi(t)| \leq \varphi(0) = 1$

(ii) φ ist gleichmäßig stetig

(iii) φ ist "positiv semidefinit", d.h. $\forall n, \forall t_1, \dots, t_n \in \mathbb{R}, \forall \lambda_1, \dots, \lambda_n \in \mathbb{C}$:

$$\sum_{l,k=1}^{n} \lambda_l \bar{\lambda}_k \cdot \varphi(t_l - t_k) \geq 0.$$

Beweis: von "$\Rightarrow$" (Umkehrung wird nicht benötigt.)

(i) $|\varphi(t)| = |E(e^{itX})| \leq E(|e^{itx}|) = 1, \quad \varphi(0) = 1$
Geometrisch: e^{itX} auf Einheitssphäre in $\mathbb{C} \Rightarrow$ EW in Einheitskreis

(ii) $|\varphi(t+h) - \varphi(t)| = |Ee^{i(t+h)X} - e^{itX}| \leq E|e^{i(t+h)X} - e^{itX}| = E|\underbrace{e^{ihX} - 1}_{\to 0}| \to 0$ für $h \to 0$

(majorisierte Konvergenz)

(iii)
$$\begin{aligned}
\sum_{l,k} \lambda_l \bar{\lambda}_k \cdot \varphi(t_l - t_k) &= \sum_{l,k} \lambda_l \bar{\lambda}_k \cdot \int e^{i(t_l - t_k)\cdot X} \mu(dx) \\
&= \int \left[\sum_l \lambda_l \cdot e^{it_l X} \right] \cdot \overline{\left[\sum_k \lambda_k \cdot e^{it_k X} \right]} \mu(dx) \\
&= \int F(x) \cdot \bar{F}(x) \mu(dx) \geq 0. \square
\end{aligned}$$

Weitere Eigenschaften:

(iv) $\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$

(v) $\varphi_{aX+b}(t) = e^{itb} \cdot \varphi(at)$

(vi) X, Y *unabhängig* $\Rightarrow \varphi_{X+Y} = \varphi_X \cdot \varphi_Y$, *allgemein:*

$$\begin{aligned}
\widehat{\mu * \nu} &= \widehat{\mu} \cdot \widehat{\nu} \\
\textit{denn: } \widehat{\mu * \nu}(t) &= \int e^{itx} (\mu * \nu)(dx) = \int \int e^{it(x_1 + x_2)} \mu(dx_1) \nu(dx_2) \\
&= \int e^{itx_1} \mu(dx_1) \cdot \int e^{itx_2} \nu(dx_2) = \widehat{\mu}(t) \cdot \widehat{\nu}(t) \\
\textit{und damit } \varphi_{X+Y} &= \widehat{P_{X+Y}} = \widehat{P_X * P_Y} = \widehat{P_X} \cdot \widehat{P_Y} = \varphi_X \cdot \varphi_Y.
\end{aligned}$$

Beispiele 12.3.

1. $\mu = \delta_\alpha \Rightarrow \widehat{\mu}(t) = \cos \alpha t + i \cdot \sin \alpha t = e^{i\alpha t}$
   Insbesondere: $\mu = \delta_0 \Rightarrow \widehat{\mu}(t) \equiv 1$

2. $\mu =$ Gleichverteilung auf $[a, b] \Rightarrow \widehat{\mu}(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}$
   Speziell: $a = -b \Rightarrow \widehat{\mu}(t) = \frac{\sin(tb)}{tb}$

3. $\mu = \text{Exp}(1) \Rightarrow \widehat{\mu}(t) = \frac{1}{1-it}$

4. $\mu = \nu_{0,1} = g \cdot \lambda^1$ Normalverteilung mit $g(x) = (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{x^2}{2}\right)$

$$\Rightarrow \qquad \widehat{\mu}(t) = \exp\left(-\frac{t^2}{2}\right)$$

$$\begin{aligned}
\underline{\text{Denn}}\text{: } \varphi(t) &= \frac{1}{\sqrt{2\pi}} \int e^{itx} \cdot e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int \cos(tx) \cdot e^{-\frac{x^2}{2}} dx \\
\Rightarrow \varphi'(t) &= \frac{1}{\sqrt{2\pi}} \int -x \cdot \sin(tx) \cdot e^{-\frac{x^2}{2}} dx \\
&= -\frac{1}{\sqrt{2\pi}} \int t \cdot \cos(tx) \cdot e^{-\frac{x^2}{2}} dx + \frac{1}{\sqrt{2\pi}} \left[\sin(tx) \cdot e^{-\frac{x^2}{2}}\right]_{-\infty}^{+\infty} \\
&= -t \cdot \varphi(t) \\
\Rightarrow \log \varphi(t) &= -\frac{t^2}{2} + \underbrace{\log \varphi(0)}_{=0} \Rightarrow \varphi(t) = \exp\left(-\frac{t^2}{2}\right)
\end{aligned}$$

5. Cauchy-Verteilung $\mu(dx) = \frac{1}{\pi} \frac{1}{1+x^2} dx = C_1(dx)$

$$\Rightarrow \qquad \widehat{\mu}(t) = \exp(-|t|)$$

(Übung!)

Folgerung 12.4. (unter Verwendung des folgenden Eindeutigkeitssatzes)

ad 4.

$$\begin{aligned}
\nu_{\alpha,\sigma^2} * \nu_{\beta,\tau^2} &= \nu_{\alpha+\beta,\sigma^2+\tau^2}, \\
\textit{denn: } \widehat{\nu_{\alpha,\sigma^2} * \nu_{\beta,\tau^2}}(t) &= e^{it(\alpha+\beta)} \cdot \widehat{\nu_{0,1}}(t\sigma) \cdot \widehat{\nu_{0,1}}(t\tau) \\
&= e^{it(\alpha+\beta)} \cdot \exp\left(-\frac{t^2}{2}(\sigma^2+\tau^2)\right) = \widehat{\nu_{\alpha+\beta,\sigma^2+\tau^2}}(t)
\end{aligned}$$

ad 5.

$$\begin{aligned}
C_\alpha * C_\beta &= C_{\alpha+\beta}, \text{ denn} \\
\widehat{C_\alpha * C_\beta}(t) &= \widehat{C_1}(\alpha t) \cdot \widehat{C_1}(\beta t) = \exp(-(\alpha+\beta)|t|) = \widehat{C_{\alpha+\beta}}(t)
\end{aligned}$$

Satz 12.5 (Eindeutigkeitssatz und Inversionsformel).
(i) μ ist eindeutig durch φ_μ bestimmt.

(ii) Für $a < b$ gilt:

$$\mu(]a,b[) + \frac{1}{2}\mu(\{a,b\}) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi_\mu(t) dt.$$

Beweis: Offenbar (ii)$\Rightarrow$(i). Zu (ii): Sei

$$\begin{aligned}
I_T &= \int_{-T}^{T} \frac{e^{-ita}-e^{-itb}}{it}\varphi(t)dt = \int_{-T}^{T}\int \underbrace{\frac{e^{-ita}-e^{-itb}}{it}}_{\substack{=\int_a^b e^{-ity}dy \\ \Rightarrow \left|\int_a^b e^{-ity}dy\right| \le b-a}} e^{itx}\mu(dx)dt \\
&= \int\int_{-T}^{T} \frac{e^{-ita}-e^{-itb}}{it} e^{itx} dt\, \mu(dx) \\
&= 2\int\left[\int_0^T \frac{\sin(t(x-a))}{t}dt - \int_0^T \frac{\sin(t(x-b))}{t}dt\right]\mu(dx) \\
&\to 2\int\left[\begin{array}{l} \frac{\pi}{2}\cdot \mathbb{1}_{]a,\infty[}(x) - \frac{\pi}{2}\mathbb{1}_{]b,\infty[}(x) \\ -\frac{\pi}{2}\mathbb{1}_{]-\infty,a[}(x) + \frac{\pi}{2}\mathbb{1}_{]-\infty,b[}(x)\end{array}\right]\mu(dx) = 2\pi\cdot\mu(]a,b[) + \pi\cdot\mu(\{a,b\}) \\
&\quad \text{denn: } \int_0^T \frac{\sin(ty)}{t}dt \to \int_0^\infty \frac{\sin(ty)}{t}dt = \begin{cases} +\frac{\pi}{2} & , \ y>0 \\ 0 & , \ y=0 \\ -\frac{\pi}{2} & , \ y<0 \end{cases} \quad \square
\end{aligned}$$

Korollar 12.6. *Ist $|\varphi| = |\widehat{\mu}| \in L^1(\mathbb{R}, \lambda^1)$, so ist $\mu = f\cdot\lambda^1$ mit*

$$f(x) = \frac{1}{2\pi}\cdot\widehat{\varphi}(-x) \qquad \lambda^1\text{-f.a. } x\in\mathbb{R}$$

Beweis: Für λ^1-f.a. $x\in\mathbb{R}$ gilt

$$\begin{aligned}
f(x) &= \lim_{\epsilon\to 0}\frac{1}{2\epsilon}\mu(]x-\epsilon, x+\epsilon[) \\
&= \lim_{\epsilon\to 0}\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T} e^{-itx}\cdot\underbrace{\frac{1}{2it\epsilon}\left[e^{it\epsilon}-e^{-it\epsilon}\right]}_{\frac{1}{t\epsilon}\cdot\sin(t\epsilon)}\varphi(t)dt \\
&= \lim_{\epsilon\to 0}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\cdot\underbrace{\frac{1}{t\epsilon}\cdot\sin(t\epsilon)}_{\to 1}\varphi(t)dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\cdot\varphi(t)dt = \frac{1}{2\pi}\widehat{\varphi}(-x). \square
\end{aligned}$$

Die Dichte $f = \frac{d\mu}{d\lambda}$ ist dabei beschränkt und stetig!

Bemerkung: $\mu(\{x\}) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e^{-itx}\varphi(t)dt$

Proposition 12.7. *Ist das n-te Moment $\int |x|^n \mu(dx) < \infty$, so ist*

$$\varphi = \widehat{\mu} \in \mathcal{C}^n(\mathbb{R}) \quad \textit{und} \quad \varphi^{(n)}(0) = i^n \cdot \int x^n d\mu = i^n \cdot \mathrm{E}(X^n)$$

Beweis: Übung! $\square$

Satz 12.8 (Stetigkeitssatz von Lévy). *(i) Aus $\mu_n \to \mu$ (schwach) folgt $\widehat{\mu_n} \to \widehat{\mu}$ (punktweise).*

(ii) Falls $\widehat{\mu_n} \to \varphi$ (punktweise) mit $\varphi : \mathbb{R} \longrightarrow \mathbb{C}$ stetig in 0, so folgt:

$$\exists \mu \textit{ mit } \varphi = \widehat{\mu} \textit{ und } \mu_n \to \mu \textit{ (schwach).}$$

Bemerkungen:

(a) In (i) gilt sogar: $\widehat{\mu_n} \to \widehat{\mu}$ lokal gleichmäßig

(b) In (ii) ist Stetigkeit von φ in 0 wesentlich.
Beispiel: $\mu_n = \nu(0, n)$ Normalverteilung

$$\Rightarrow \widehat{\mu_n}(t) = \exp(-n\frac{t^2}{2}) \to \begin{cases} 0 & , \quad t \neq 0 \\ 1 & , \quad t = 0 \end{cases} = \varphi(t)$$

Aber $\nexists\ \mu : \mu_n \to \mu$!
(Denn $\mu_n(]-\infty, x]) \to \frac{1}{2} \ \forall x \in \mathbb{R}$)

Beweis:

(i) $\forall t \in \mathbb{R} : x \mapsto e^{itx}$ ist beschränkt und stetig. Also wegen $\mu_n \to \mu$:

$$\widehat{\mu_n}(t) = \int e^{itx} \mu_n(dx) \to \int e^{itx} \mu(dx) = \widehat{\mu}(t),$$

d.h. $\widehat{\mu_n} \to \widehat{\mu}$ pkt.weise.

(ii) Annahme: $\widehat{\mu_n} \to \varphi$ punktweise.
Sei $\widetilde{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ die Einpunkt-Kompaktifizierung von $\mathbb{R}$.
Setze $\mu_n \in \mathcal{M}_1(\mathbb{R})$ fort zu $\tilde{\mu}_n \in \mathcal{M}_1(\widetilde{\mathbb{R}})$ durch $\tilde{\mu}_n(\{\infty\}) = 0 \ \ (\forall n)$.

Zum Beweis von (ii) verwenden wir folgende elementare Folgerung aus dem Satz von Prokorov (den wir zu Beginn des nächsten Semesters herleiten).

Lemma 12.9. $\mathcal{M}_1(\widetilde{\mathbb{R}})$ *ist kompakt.*

Fortsetzung des Bew. von (ii):

Also ist $\{\tilde{\mu}_n\}_{n\in\mathbb{N}}$ relativ-kompakt, $\exists$ Teilfolge $\tilde{\mu}_{n_k}$ und Maß $\tilde{\mu} \in \mathcal{M}_1(\widetilde{\mathbb{R}})$:

$$\tilde{\mu}_{n_k} \to \tilde{\mu}.$$

Für $\delta > 0$ gilt:

$$\begin{aligned}
\frac{1}{2\delta}\int_{-\delta}^{\delta} \widehat{\mu_{n_k}}(t)dt &= \frac{1}{2\delta}\int_{-\delta}^{\delta}\int e^{itx}\mu_{n_k}(dx)dt \\
&= \frac{1}{2\delta}\int\int_{-\delta}^{\delta} e^{itx}dt\,\mu_{n_k}(dx) = \int \frac{\sin(\delta x)}{\delta x}\mu_{n_k}(dx) \\
&= \int_{\widetilde{\mathbb{R}}} f_\delta(x)\tilde{\mu}_{n_k}(dx) \text{ mit } f_\delta(x) = \left\{\begin{array}{lll} \frac{\sin(\delta x)}{\delta x} & , & x \in \mathbb{R}\setminus\{0\} \\ 0 & , & x \in \widetilde{\mathbb{R}}\setminus\mathbb{R} \\ 1 & , & x = 0 \end{array}\right\} \in \mathcal{C}_b(\widetilde{\mathbb{R}})
\end{aligned}$$

Daher

$$\begin{aligned}
\frac{1}{2\delta}\int_{-\delta}^{\delta}\varphi(t)dt &= \lim_{k\to\infty}\frac{1}{2\delta}\int_{-\delta}^{\delta}\widehat{\mu_{n_k}}(t)dt \quad \text{Majorisierte Konvergenz wegen } |\widehat{\mu_{n_k}}| \le 1 \\
&= \lim_{k\to\infty}\int_{\widetilde{\mathbb{R}}} f_\delta(x)\tilde{\mu}_{n_k}(dx) \\
&= \int_{\widetilde{\mathbb{R}}} f_\delta(x)\tilde{\mu}(dx) = \int_{\mathbb{R}} f_\delta(x)\mu(dx) \\
&\quad \text{mit } \mu = \mathbb{1}_{\mathbb{R}}\cdot\tilde{\mu} = \text{Einschränkung von } \tilde{\mu} \text{ auf } \mathbb{R} \in \mathcal{M}(\mathbb{R})
\end{aligned}$$

<u>Beh.</u>: $\mu(\mathbb{R}) = 1$

$$\begin{aligned}
\text{Denn: } \mu(\mathbb{R}) &= \lim_{\delta\to 1}\int_{\mathbb{R}} \underbrace{f_\delta(x)}_{\to 1}\mu(dx) \\
&= \lim_{\delta\to 0}\frac{1}{2\delta}\int_{-\delta}^{\delta}\varphi(t)dt = \varphi(0) = 1 \text{ (wegen Stetigkeit in 0!)}
\end{aligned}$$

Also $\mu \in \mathcal{M}_1(\mathbb{R})$.
Ferner gilt wegen $\tilde{\mu}_{n_k} \to \tilde{\mu}$ in $\mathcal{M}_1(\widetilde{\mathbb{R}})$: $\forall G$ offen $\subset \mathbb{R}$:

$$\liminf_{k\to\infty}\mu_{n_k}(G) = \liminf\tilde{\mu}_{n_k}(G) \ge \tilde{\mu}(G) = \mu(G)$$

d.h. $\quad \mu_{n_k} \to \mu$ in $\mathcal{M}_1(\mathbb{R})$.
Also nach (i): $\widehat{\mu_{n_k}} \to \widehat{\mu}$. Andererseits nach Voraussetzung: $\widehat{\mu_{n_k}} \to \varphi$.

Also $\widehat{\mu} = \varphi$ und damit μ eindeutig!
Schließlich $\forall f \in \mathcal{C}_b(\mathbb{R})$: $\forall$ Teilfolge (n_k) : $\exists$ Teilfolge (n_{k_l}):

$$\begin{gathered} \mu_{n_{k_l}} \to \mu \\ \Rightarrow \int f d\mu_{n_{k_l}} \to \int f d\mu \\ \Rightarrow \int f d\mu_n \to \int f d\mu \\ \Rightarrow \mu_n \to \mu. \square \end{gathered}$$

Kapitel 13

Der Zentrale Grenzwertsatz

Ausgangsfrage: Gegeben $(X_i)_{i\in\mathbb{N}}$ unabhängig, identisch verteilt $\in \mathcal{L}^2$.

Was kann man über $S_n = \sum_{i=1}^{n} X_i$ sagen ?

- $\frac{1}{n}S_n \to \alpha = E(X_i)$ P-f.s. (Starke Gesetz)
- $P\left(\left|\frac{1}{n}S_n - \alpha\right| \geq \epsilon\right) \sim e^{-I(\epsilon)\cdot n}$,

  falls $E(\exp tX_i) < \infty$ $(\forall t)$ (Große Abweichung, Satz von Cramér)
- $\text{var}\left(\frac{1}{\sqrt{n}}S_n\right) = \frac{1}{n}\text{var}S_n = \text{var}X_1 \quad (\forall n \in \mathbb{N})$

  Konvergiert $\frac{1}{\sqrt{n}}[S_n - E(S_n)]$?

Setze $\frac{S_n - E(S_n)}{\sqrt{\text{var}S_n}} = S_n^* \Rightarrow E(S_n^*) = 0, \text{var}S_n^* = 1$

Satz 13.1. (ZGS) *Unter obiger Voraussetzung:*

$$P_{S_n^*} \to \nu_{0,1} \qquad \textit{schwach}$$

oder, äquivalent dazu: $\forall x \in \mathbb{R}$:

$$P(S_n^* \leq x) \to \Phi(x) = (2\pi)^{-\frac{1}{2}} \cdot \int_{-\infty}^{x} e^{-\frac{y^2}{2}}\,dy.$$

Bemerkungen:

(i) Für Binomialverteilte X_i kann man $P(S_n^* \leq x)$ explizit berechnen. Der ZGS folgt dann aus der Stirlingschen Formel (= Satz von DeMoivre-Laplace).

(ii) Obiger ZGS ist Spezialfall des folgenden Satzes von Lindeberg-Feller.

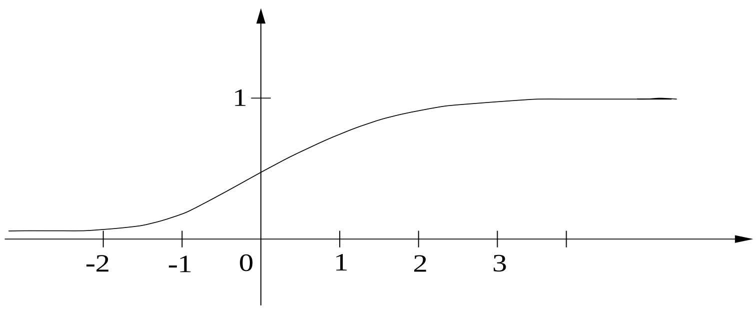
Beweis: o.B.d.A. $E(X_k) = 0$, $\mathrm{var}(X_k) = 1$
(ansonsten Übergang zu $X_k' = \frac{X_k - E(X_k)}{\sqrt{\mathrm{var} X_k}}$) $\Rightarrow S_n^* = \frac{1}{\sqrt{n}} S_n$
Wegen $X_k \in \mathcal{L}^2$ ist $\varphi_{X_k} \in \mathcal{C}^2$ mit $\varphi_{X_k}(0) = 1$,

$$\begin{aligned}
\varphi'_{X_k}(0) &= i \int x dP_{X_k} = i \cdot E(X_k) = 0 \text{ und} \\
\varphi''_{X_k}(0) &= - \int x^2 dP_{X_k} = -E(X_k^2) = -1.
\end{aligned}$$

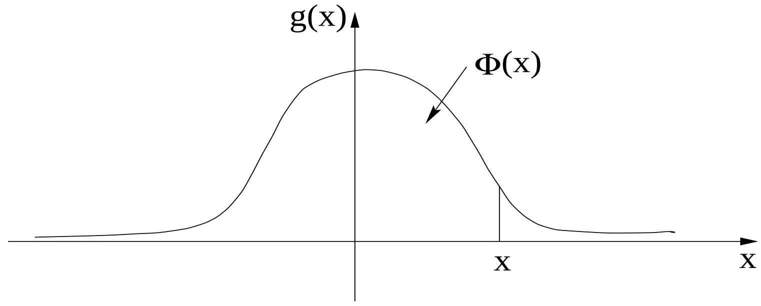
Also

$$\begin{aligned}
& E(e^{itX_k}) = \varphi_{X_k}(t) = 1 - \frac{t^2}{2} + o(t^2) = \left(1 - \frac{t^2}{2}\right) \cdot \left(o(t^2)\right) \text{ für } t \to 0. \\
\Rightarrow \quad & E(e^{itS_n^*}) = \varphi_{S_n}\left(\frac{t}{\sqrt{n}}\right) = \left[\varphi_{X_k}\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
& = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n = \underbrace{\left(1 - \frac{t^2}{2n}\right)^n}_{\to e^{-\frac{t^2}{2}}} \cdot \underbrace{\exp\left(n \cdot o\left(\frac{t^2}{n}\right)\right)}_{\to 1} \\
& \to \exp\left(-\frac{t^2}{2}\right) = \widehat{\nu_{0,1}}(t) \text{ für } n \to \infty \text{ (und } t \text{ fix)} \\
\Rightarrow \quad & P_{S_n^*} \to \nu_{0,1} \qquad \square
\end{aligned}$$

Die Funktion $\Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = \nu_{0,1}(]-\infty, x])$ ist tabelliert.



$$\Phi(0) = 0,5,\ \Phi(1) = 0,8413\ldots,\ \Phi(2) = 0,9773\ldots,\ \Phi(3) = 0,9987\ldots$$
$$\Phi(-x) = 1 - \Phi(x)$$



Beispiel 13.2. Münzwurf
$(X_k)_{k\in\mathbb{N}}$ unabhängig, identisch verteilt mit

$$P(X_k = 0) = P(X_k = 1) = \frac{1}{2},$$

$$S_n = \sum_{k=1}^{n} X_k \text{ Anzahl von “Zahl“}$$
$$E(X_k) = \frac{1}{2},\ \mathrm{var}(X_k) = \frac{1}{4}$$
$$S_n^* = \frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{4}}}$$

<u>ZGS</u>: $$P\left(\left|S_n - \frac{n}{2}\right| \le x \cdot \frac{\sqrt{n}}{2}\right) \approx \nu_{0,1}([-x,x])$$

Konkret $x = 2$ (“2σ-Regel“):

$$P\left(\left|S_n - \frac{n}{2}\right| \le \sqrt{n}\right) \approx 0,95$$

“3σ-Regel“:

$$\begin{aligned} P\left(\frac{1}{\sqrt{n}} S_n > 3\sigma\right) &= P\left(\frac{1}{\sqrt{n}} S_n < -3\sigma\right) \sim 0,001 \\ P\left(\frac{1}{\sqrt{n}} S_n > 2\sigma\right) &= P\left(\frac{1}{\sqrt{n}} S_n < -2\sigma\right) \sim 0,02 \end{aligned}$$

falls $EX_1 = 0$ Konkret $n = 10.000$:

$$P(4.900 \le S_n \le 5.100) \approx 0,95$$

Beispiel 13.3. Roulette
36 nummerierte Felder (18 rot,18 schwarz) + 2 grüne Felder (0 und 00)

$$\begin{aligned}
\Rightarrow \quad & P(X_i = 1) = \frac{18}{38},\ P(X_i = -1) = \frac{20}{38} \\
& E(X_i) = -\frac{1}{19},\ \mathrm{var} X_i = 1 - \left(\frac{1}{19}\right)^2 \approx 0,9972 \approx 1 \\
\Rightarrow \quad & P(S_n \geq 0) = P\left(S_n^* \geq \sqrt{n} \cdot \frac{1/19}{0,9972}\right) \approx 1 - \Phi\left(\frac{\sqrt{n}}{19}\right)
\end{aligned}$$

z.B. $n = 19^2 = 361$:

$$P(S_{361} \geq 0) \approx 1 - \Phi(1) \approx 0,1587$$

$n = 100$:

$$P(S_{100} \geq 0) \approx 1 - \Phi\left(\frac{10}{19}\right) \approx 0,3$$

Beispiel 13.4.
Vergleich mit exakten Werten: $n = 16$ (“klein“)
Exakte Binomialverteilung

$$P(S_{16} = 8) = \binom{16}{8} \cdot 2^{-16} = 0,1964$$

Normalapproximation

$$\begin{aligned}
P(S_{16} = 8) &= P(|S_{16} - 8| \leq \overbrace{0,5}^{\text{Histogramm-Korrektur}}) = P(|S_{16}^*| \leq 0,25) \\
&\approx \Phi(0,25) - \Phi(-0,25) = 0,1974.
\end{aligned}$$

Beispiel 13.5. Geburten in der Schweiz zwischen 1871-1900
m= 1.359.670 ♂, w= 1.285.086 ♀
Verträgt sich das mit der Hypothese $P(♂) = P(♀) = \frac{1}{2}$?
M.a.W.: Wie wahrscheinlich ist obiges Ereignis unter diesem W-Maß?
Insgesamt $n = 2.644.756 = m + w$

$$\begin{aligned}
E(S_n) &= \frac{n}{2} = 1.322.378 = \frac{m+w}{2} \\
\sqrt{\mathrm{var} S_n} &= \frac{1}{2}\sqrt{n} \approx 813
\end{aligned}$$

$$P(S_n \geq m) = P(S_n^* \geq 46) \approx 1 - \Phi(46) \approx \frac{1}{46} e^{-\frac{46^2}{2}} \approx e^{-1000} \approx 10^{-500}$$

<u>Antwort</u>: Die Hypothese

$$P(♂) = P(♀)$$

ist definitiv auszuschließen.

Allgemeine Version des ZGWS (Lindeberg-Feller)

$X_n, n = 1, 2, \ldots$ seien nur noch unabhängig, nicht notwendig identisch, verteilte zentrierte ($E[X_n] = 0$) ZV.

Endliche Varianz:
$$\sigma_n^2 := E[X_n^2] < \infty$$
$$\tau_n^2 := E[S_n^2] = \sum_{i=1}^n \sigma_i^2, \quad n \geq 0,\ \tau_n > 0 \text{ ab einem } n_0$$
$$(= n\tau_1^2 \text{ für i.i.d.})$$

Hinreichende Bedingung für ZGWS ist die Lindeberg-Bedingung:

(L) $\forall \epsilon > 0$ gilt
$$L_n(\epsilon) = \sum_{j=1}^n \int\limits_{\{|X_j| \geq \epsilon\tau_n\}} X_j^2 dP = o(\tau_n^2)$$
$$\left(\text{d.h. } \frac{L_n(\epsilon)}{\tau_n^2} = \frac{n \cdot \int_{\{|X_1| \geq \epsilon\sqrt{n}\sigma\}} X_1^2 dP}{n \cdot \sigma^2} = \int\limits_{\{|X_1| \geq \epsilon\sqrt{n}\sigma\}} X_1^2 dP \to 0\right)$$

(L) impliziert neben dem ZGWS (s.u.) auch die Fellersche Bedingung

Lemma 13.6.
$$\textit{(L) impliziert (F)} \lim_{n\to\infty} \max_{1\leq j\leq n} \frac{\sigma_j^2}{\tau_n^2} = 0$$

Beweis:
$$\max_{1\leq j\leq n} \frac{\sigma_j^2}{\tau_n^2} \leq \max_{1\leq j\leq n} \frac{1}{\tau_n^2}\left[\epsilon^2\tau_n^2 + \int\limits_{\{|X_j| > \epsilon\tau_n\}} |X_j|^2 dP\right]$$
wegen (L) $\stackrel{n\to\infty}{\longrightarrow} \epsilon^2$. Da ϵ beliebig, folgt (F). $\square$

Bemerkungen:

(F) bedeutet, daß keine Einzelvarianz asymptotisch von der gleichen Größenordnung wie die Gesamtvarianz ist.

(F) impliziert dann, daß $\left\{\frac{X_j}{\tau_n}\right\}_{j,n\in\mathbb{N}}$ "asymptotisch vernachlässigbar" ist, vgl. Bauer, Loeve,... (klassische Lehrbuchliteratur).

Satz 13.7 (Lindeberg-Feller). *Äquivalent sind:*
(i) Die Lindeberg-Bedingung gilt.

(ii) Die Fellersche Bedingung und der Zentrale Grenzwertsatz

$$\frac{S_n}{\tau_n} \to \nu_{0,1} \qquad \textit{in Verteilung}$$

gelten.

Beweis:
1. Sei (L) erfüllt. Das vorhergehende Lemma 13.6 zeigt dann die Gültigkeit von (F). Zeige noch ZGWS.
   Abschätzen der Differenz zwischen den charakteristischen Funktionen:

$$\begin{aligned}
\Delta_n(t) \quad &:= \quad \left|\varphi_{\frac{S_n}{\tau_n}}(t) - \widehat{\nu_{0,1}}(t)\right| = \left|\prod_{j=1}^n \varphi_{\frac{X_j}{\tau_n}}(t) - \prod_{j=1}^n e^{-\frac{\sigma_j^2}{\tau_n^2}\frac{t^2}{2}}\right| \\
&\qquad\qquad \uparrow \\
&\qquad \text{Unabhängig von } X_j \text{ und } \sum_{j=1}^n \frac{\sigma_j^2}{\tau_n^2} = 1. \\
&\leq \quad \sum_{j=1}^n \left|\varphi_{\frac{X_j}{\tau_n}}(t) - e^{-\frac{\sigma_j^2}{\tau_n^2}\frac{t^2}{2}}\right|
\end{aligned}$$

$$\begin{aligned}
\text{(denn: } \left|\prod_{j=1}^n a_j - \prod_{j=1}^n b_j\right| \quad &= \quad |(a_1-b_1)a_2 \cdot \ldots \cdot a_n + b_1(a_2-b_2)a_3 \cdot \ldots \cdot a_n + \ldots + \\
&\qquad + b_1 \cdot \ldots \cdot b_{n-1}(a_n - b_n)| \\
&\qquad\qquad \text{falls: } a, b \leq 1 \leq \sum_{j=1}^n |a_j - b_j|
\end{aligned}$$

Taylorentwicklung von charakteristischen Funktionen. (vgl. Übung, Serie 12, A2).
Ist μ ein Wahrscheinlichkeitsmaß mit $\int x d\mu = 0$, $\int x^2 \mu(dx) =: \sigma_\mu^2$ gilt

$$\widehat{\mu}(t) = 1 - \sigma_\mu^2 \frac{t^2}{2} + \Theta_\mu(t)\frac{t^2}{2}$$

$$\begin{aligned}
\text{mit } |\Theta_\mu(t)| \quad &\leq \quad \sup_{0\leq\theta\leq 1} |\widehat{\mu}''(\theta t) - \widehat{\mu}''(0)| \leq \sup_{0\leq\theta\leq 1} \int |e^{i\theta tx} - 1| x^2 \mu(dx) \\
&\leq \quad \sup_{0\leq\theta\leq 1} \int_{\{|x|<\delta\}} |e^{i\theta tx} - 1| x^2\mu(dx) + \sup_{0\leq\theta\leq 1} \int_{\{|x|\geq\delta\}} |e^{i\theta tx} - 1| x^2 \mu(dx) \\
&\leq \quad \epsilon\sigma^2 + 2\int_{\{|x|\geq\delta\}} x^2\mu(dx)
\end{aligned}$$

bei vorgegebenem $\epsilon > 0$ durch Wahl von $\delta = \delta(\epsilon, t) = \frac{\epsilon}{t}$,
(denn $\left|\int_0^{t\theta x} e^{i\theta t} dt\right| = |e^{i\theta tx} - 1| \leq |\theta tx| < |t||x| < \epsilon$ für $|x| < \delta(\epsilon, t) := \frac{|\epsilon|}{|t|}$)
$e^{-\frac{\sigma_j^2}{\tau_n^2} \cdot \frac{t^2}{2}}$ ist charakteristische Funktion von $\nu_{0,\frac{\sigma_j^2}{\tau_n^2}}$. Wende Taylorentwicklung auf $\widehat{\nu}_{0,\frac{\sigma_j^2}{\tau_n^2}}$ und $\varphi_{\frac{X_j}{\tau_n}}$ an. Dann gilt wegen $\operatorname{var}\left(\frac{X_j}{\tau_n}\right) = \operatorname{var}(\nu_{0,\frac{\sigma_j^2}{\tau_n^2}})$ für vorgegebenes ϵ und $\delta = \delta(\epsilon, t)$.

$$
\begin{aligned}
\Delta_n(t) &\leq \sum_{j=1}^n \left| \Theta_{\frac{X_j}{\tau_n}}(t) - \Theta_{\nu_{0,\frac{\sigma_j^2}{\tau_n^2}}}(t) \right| \frac{t^2}{2} \\
&\leq t^2[\epsilon + \underbrace{\sum_{j=1}^n \int\limits_{\{\frac{X_j}{\tau_n} > \delta\}} \frac{X_j^2}{\tau_n^2} dP}_{\substack{\to 0 \text{ für } n \to \infty \\ \text{wegen } \frac{L_n(\delta)}{\tau_n^2} \to 0}} + \underbrace{\sum_{j=1}^n \int\limits_{\{|x| > \delta\}} x^2 \nu_{0,\frac{\sigma_j^2}{\tau_n^2}}(dx)}_{R_n(\delta)}]
\end{aligned}
$$

Zeige: $R_n(\delta) \to 0$ ("Lindeberg für Normalverteilung")

$$
\begin{aligned}
R_n(\delta) &= \sum_{j=1}^n \frac{\sigma_j^2}{\tau_n^2} \int\limits_{\{|x| \geq \frac{\delta \tau_n}{\sigma_j^2}\}} x^2 \nu_{0,1}(dx) \leq \sum_{j=1}^n \frac{\sigma_j^2}{\tau_n^2} \int\limits_{\left\{|x| \geq \frac{\delta \tau_n}{\max\limits_{1 \leq i \leq n} \sigma_i}\right\}} x^2 \nu_{0,1}(dx) \\
&= \int\limits_{\underbrace{\{|x| > \underbrace{\frac{\delta \tau_n}{\max_{1 \leq i \leq n} \sigma_j}}_{\to \infty \text{ (da (F) gilt)}} \}}_{\{\ldots\} \to \emptyset}} x^2 \nu_{0,1}(dx) \stackrel{n \to \infty}{\longrightarrow} 0
\end{aligned}
$$

Insgesamt:

$$
\begin{aligned}
&\lim_{n \to \infty} \Delta_n(t) \leq \epsilon t^2 \text{ für jedes } \epsilon \text{ somit} \\
&\lim_{n \to \infty} \Delta_n(t) = 0
\end{aligned}
$$

und der ZGWS folgt aus dem Stetigkeitssatz von Levy.

2. ZGWS + (F) $\Rightarrow$ (L).

(F) liefert

$$\left|1-\varphi_{\frac{X_j}{\tau_n}}(t)\right| \underset{\substack{\uparrow\\ \text{Taylor}}}{\leq} \frac{\sigma_j^2 t^2}{2\tau_n^2} \underset{\substack{\uparrow\\ \text{(F)}}}{\leq} \frac{t^2}{2}\max_{1\leq h\leq n}\frac{\sigma_h^2}{\tau_n^2} < \epsilon$$

denn $\varphi(t) = 1-\sigma^2\cdot\frac{t^2}{2}+\Theta(t)\cdot\frac{t^2}{2}$ mit $|\Theta(t) \leq \sup\limits_{0\leq s\leq t}|\varphi''(s)-\varphi''(0)| \leq \int|e^{isx}-1|x^2\mu(dx) \leq 2\int x^2\mu(dx) = 2\sigma^2$ für $n > n(\epsilon,t), j=1,\ldots,n$
Reihenentwicklung von $\log\varphi_{\frac{X_j}{\tau_n}}(t)$, $n > n(\epsilon,t)$

$$\begin{aligned}\left|\sum_{j=1}^n \log\varphi_{\frac{X_j}{\tau_n}}(t)-\left(\varphi_{\frac{X_j}{\tau_n}}(t)-1\right)\right| \quad &\leq \quad \sum_{j=1}^n\left|\varphi_{\frac{X_j}{\tau_n}}(t)-1\right|^2\\ &\leq \quad \epsilon\sum_{j=1}^n\left|\varphi_{\frac{X_j}{\tau_n}}(t)-1\right| \leq \epsilon\sum_{j=1}^n\frac{\sigma_j^2t^2}{2\tau_n^2} = \epsilon\frac{t^2}{2}\ \text{d.h.}\end{aligned}$$

denn $\sum\limits_{j=1}^n \log\varphi_{\frac{X_j}{\tau_n}}(t) \overset{\text{ZGWS}}{\longrightarrow} -\frac{t^2}{2}$

$$-\frac{t^2}{2} = \lim_{n\to\infty}\sum_{j=1}^n\left(\varphi_{\frac{X_j}{\tau_n}}(t)-1\right)$$

Betrachtet man den Realanteil, so folgt

$$(**)\quad \frac{t^2}{2}-\sum_{j=1}^n\int\limits_{\{|X_j|<\delta\tau_n\}}\left(1-\cos t\frac{X_j}{\tau_n}\right)dP = \sum_{j=1}^n\int\limits_{\{|X_j|\geq\delta\tau_n\}}\left(1-\cos t\frac{X_j}{\tau_n}\right)dP + R_n(t)$$

mit $R_n(t)\to 0$ für $n\to\infty$
Da $1-\cos z \leq \frac{z^2}{2}$, so ist

$$\sum_{j=1}^n\int\limits_{\{|X_j|\leq\delta\tau_n\}}\left(1-\cos\frac{tX_j}{\tau_n}\right)dP \leq \frac{t^2}{2\tau_n^2}\sum_{j=1}^n\int\limits_{\{|X_j|\leq\delta\tau_n\}}X_j^2dP = \frac{t^2}{2}\left(1-\frac{L_n(\delta)}{\tau_n^2}\right).$$

Andererseits

$$\sum_{j=1}^n\int\limits_{\{|X_j|>\delta\tau_n\}}\left(1-\cos\frac{tX_j}{\tau_n}\right)dP \leq 2\sum_{j=1}^n P[|X_j|\geq\delta\tau_n] \underset{\substack{\uparrow\\ \text{Chebyshev}}}{\leq} \frac{2}{\delta^2\tau_n^2}\sum_{j=1}^n\sigma_j^2 = \frac{2}{\delta^2}$$

Diese zwei Ungleichungen in $(**)$ eingesetzt ergibt

$$\frac{2}{\delta^2}+R_n(t) \geq \frac{t^2}{2}-\sum_{i=1}^{n} \int_{\{|X_i|<\delta\tau_n\}} \left(1-\cos\left(t\frac{X_i}{\tau_n}\right)\right) dP \geq \frac{t^2}{2}-\frac{t^2}{2}\left(1-\frac{L_n(\delta)}{\tau_n^2}\right) = \frac{t^2}{2}\frac{L_n(\delta)}{\tau_n^2}$$

damit $\lim\limits_{n\to\infty} \frac{L_n(\delta)}{\tau_n^2} \leq \frac{4}{t^2\delta^2}$, $t \to \infty$ liefert dann die Gültigkeit von (L). □

ZGWS enthält keine Aussage über Konvergenzgeschwindigkeit. Diese Aussage liefert der folgende

Satz 13.8 (Berry, Esséen). *Für jedes $\delta > 0$ existiert ein $0 < C_\delta < \infty$ mit*

$$\sup_{x\in\mathbb{R}} \left| P\left[\frac{S_n}{\tau_n} < x\right] - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy \right| \leq C_\delta \frac{1}{\tau_n^{2+\delta}} \sum_{j=1}^{n} E[|X_j|^{2+\delta}]$$

(z.B. $C_1 = 6$.) (nur sinnvoll, falls $(2+\delta)$-te Momente existieren)

Beweisskizze: für $\delta = 1, X_n$ i.i.d. (vgl. Ganssler/ Stute: W-Theorie)

1. Esséen Ungleichung
   Seien F und G Verteilungsfunktion mit $|G'(x)| \leq m$, dann

$$\sup_{x\in\mathbb{R}} |F(x) - G(x)| \leq \frac{1}{\pi} \int_{-T}^{T} \left| \frac{\varphi_F(t) - \varphi_G(t)}{t} \right| dt + \frac{24m}{\pi T} \quad \text{für alle } T > 0$$

2. Verfeinerung der Abschätzung mit Taylorentwicklung bis zur dritten Ableitung

$$\Delta_n(t) \leq \frac{8}{9T} \cdot \left(\frac{|t|^3}{6} + \frac{5t^4}{72}\right) e^{-\frac{t^2}{8}} \text{ mit } T = \frac{8\tau_n^3}{9\sum_{j=1}^{n} E[|X_j|^3]}$$

3. Setze dies in Esséen-Ungleichung ein.
   Ergibt Berry-Esséen. □

Beispiel 13.9. $(X_k)_k$ iid $\Rightarrow \tau_n = \sqrt{n}\sigma$,

$C_\delta \frac{1}{\tau_n^{2+\delta}} \sum\limits_{j=1}^{n} E[|X_j|^{2+\delta}] = C_\delta \frac{1}{n^{\frac{\delta}{2}}} E|X_1|^{2+\delta}$

Kapitel 14

Poissonkonvergenz*

Satz 14.1. *Seien $X_{nj}, j = 1, \ldots, n$, unabhängig Bernoulli-verteilt mit Erfolgsparameter $p_{n,j}$,*

$$\max_{1 \leq m \leq n} p_{n,m} \to 0, \quad \sum_{m=1}^{n} p_{n,m} \to \lambda, \quad 0 \leq p_{n,j} \leq 1, \lambda > 0$$

Dann gilt mit $S_n = X_{n1} + \ldots + X_{nn}$

$$S_n \to X \qquad \textit{schwach} \ ,$$

wobei X die Poisson-Verteilung mit Parameter λ hat.

Beweis:

$$E[\exp(itS_n)] = \prod_{j=1}^{n} E[\exp(itX_{nj})] = \prod_{j=1}^{n} (it + p_{n,j}(e^{it} - 1)).$$

Nun gilt

$$\begin{aligned}
&\left| \exp\left(\sum_{j=1}^{n} p_{n,j}(e^{it} - 1) \right) - \prod_{j=1}^{n} (1 + p_{n,j}(e^{it} - 1)) \right| \\
\leq \quad & \sum_{j=1}^{n} \left| \exp(p_{n,j}(e^{it} - 1)) - (1 + p_{n,j})(e^{it} - 1) \right| \ \text{(vgl. Beweis von Lindeberg-Feller)} \\
\leq \quad & \sum_{j=1}^{n} p_{n,j}^2 \cdot |e^{it} - 1|^2 \leq 2^2 \left(\sum_{j=1}^{n} p_{n,j} \right) \left(\max_{1 \leq j \leq n} p_{n,j} \right) \to 0
\end{aligned}$$

Außerdem konvergiert wegen $\sum_{j=1}^{n} p_{n,j} \to \lambda$ der Ausdruck

$$\begin{aligned}\exp\left(\sum_{j=1}^{n} p_{n,j}(e^{it}-1)\right) \to \exp(\lambda(e^{it}-1)) &= \sum_{k=0}^{\infty} e^{-\lambda}\frac{(\lambda e^{it})^k}{k!} \\ &= \sum_{k=0}^{\infty}\left(e^{-\lambda}\frac{\lambda^k}{k!}\right)\cdot e^{itk} = \left[\int e^{itX}\right]\end{aligned}$$

$\square$

Oder auch, falls $p_{n,j} = p_n$

$$\begin{aligned}P[S_n = k] &= \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{1}{k!}\left(1-\frac{np_n}{n}\right)^{n-k}\prod_{j=1}^{k}(n-k+j)p_n \\ &= \frac{1}{k!}\underbrace{\prod_{j=1}^{k}\underbrace{(n-k+j)p_n}_{\to\lambda}}_{\to\lambda^k}\cdot\underbrace{\left(1-\frac{np_n}{n}\right)^n}_{\to e^{-\lambda}}\cdot\underbrace{\left(1-\frac{np_n}{n}\right)^{-k}}_{\to 1} = \frac{1}{k!}\lambda^k e^{-\lambda}\square\end{aligned}$$

Der Satz heißt "Schwaches Gesetz der kleinen Zahlen" oder auch "Gesetz der seltenen Ereignisse", da eine Summe von ZV, die mit kleiner Wahrscheinlichkeit positiv sind, betrachtet wird.
Dieses Gesetz wurde von Siméon D. Poisson (1781-1840) in seinem Buch (erschienen 1837)
"Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précedées des règles génératés du calcul des probabilités" formuliert.

Beispiele 14.2.

a) Seien $\xi_{n1}, \ldots, \xi_{nn}$ unabhängige und uniform auf $[-n, n]$ verteilte Zufallsvariablen. Definiere

$$\begin{aligned}X_{n,i} &= \mathbb{1}_{(a,b)}(\xi_{ni}) \text{ mit } -n < a < b < n, \\ \text{d.h. } p_{n,i} &= \frac{(b-a)}{2n} \text{ und} \\ \sum_{i=1}^{n} p_{n,i} &= \frac{b-a}{2}.\end{aligned}$$

Für $S_n := X_{n,1} + \ldots + X_{n,n}$ gilt dann

$$S_n \Rightarrow Z \text{ mit } Z\ P\left(\frac{b-a}{2}\right)\text{-verteilt.}$$

Räumliche Verteilung zufälliger Punkte.

b) Bakterienkolonie auf einer Petrischale. (Schale mit Nährflüssigkeit. Lasse Testflüssigkeit mit Bakterien darauf tropfen. Wo Bakterien hinkommen, entstehen Bakterienkolonien)
N_k = Anzahl von Quadraten mit k schwarzen Punkten

k	0	1	2	3	4	5	6
N_k	5	19	26	26	21	13	8
$N_p(k, 2.9623)$	6.1	18.0	26.7	26.4	19.6	11.7	9.5

$$N = \sum N_k = 118, \quad \frac{1}{N}\sum kN_k = 2.93$$

Weitere Beispiele für räumliche Verteilung.

c) Bombentreffer auf London (eingeteilt in $N = 576$ Quadrate)

k	0	1	2	3	4	5 und mehr
	229	211	93	35	7	1
$N_p(k, 0.9223)$	226.74	211.29	98.54	30.62	7.14	1.57

$$\frac{1}{N}\sum kN_k = \lambda = 0.9323$$

d) Ein anderes Beispiel für ein binomialverteiltes Ereignis:
Radioaktiver Zerfall. $N = 2608$ Zeitintervalle von 7.5 Sekunden/Länge.
Für jedes Zeitintervall wurde die Anzahl der Zerfälle gezählt (d.h. Klicken des Geigenzählers). N_k = Anzahl der Zeitintervalle, in denen es k-mal klickte.

k	0	1	2	3	4	5	6	7	8	9	≤ 10
N_k	57	203	383	525	532	408	273	139	45	27	16
$N_p(k, 3.870)$	54.3	210.5	407.3	525.49	508.41	393.51	253.81	140.32	67.8	29.18	17.07

Güte der Approximation eines μ mit Hilfe der χ^2-Statistik.

r-Klassen: $\mathbb{R} = A_1 \cup \ldots \cup A_r$.
N Beobachtungen, davon N_k in Klasse $k = 1, \ldots, r$

$$T_{Beob.} = \sum_{j=1}^{r} \frac{(N_j - N\mu(A_j))^2}{N\mu(A_j)}$$

$$(\Rightarrow \quad X^2_{r-1} \overset{\mathrm{D}}{=} \sum_{i=1}^{r-1} Y_i, \quad Y_i \text{ i.i.d. } \nu_{0,1})$$

In unserem Fall $A_k = \{k\}$, $k = 1, \ldots, r-1$, $A_r = [r, \infty[$.

Beispiele 14.3.

(1) Petrischale

$$P[X^2_{r-1} \geq T_{Beob.}] = 0.97$$

(d.h. Produziert man Poisson (2.9623)-verteilte Zufallsvariablen und berechnet deren T_{th}, so ist deren T_{th} in 97% der Fälle größer als unsere beobachteten $T_{Beob.}$)

(2) Bomben auf London
88% der theoretischen Experimente hätten eine größere T_{th},

d.h. beide Beispiele können sehr gut durch Poisson-Verteilungen beschrieben werden

Vergleich mit ZGWS

(Aus der Telefonsteinzeit)
2000 Kunden wollen möglicherweise von A nach B telefonieren. Minimiere Anzahl N der Leitungen, so daß unter normalen Bedingungen nur einer von 100 Anrufen nicht verbunden werden kann.
“Normale Bedingungen“, d.h. pro Stunde telefoniert jeder 2 Minuten.
Für einen festen Zeitpunkt ist die Anzahl der Anrufe eine $B(2000, \frac{1}{30})$-verteilte Zufallsvariable. Finde also kleinstes N mit

$$P[S_{2000} \geq N] < 0.01, \quad S_{2000} \overset{\mathrm{D}}{=} B(2000, \frac{1}{30})$$

Poisson-Approximation $\qquad \lambda = \frac{2000}{30} \approx 66.67$

$$\Rightarrow \quad P[X_{\frac{2000}{30}} \quad \geq \quad 87] < 0.0097$$
$$P[X_{\frac{2000}{30}} \geq 86] \approx 0.013$$

Also reichen 87 Leitungen.
Normal-Approximation $\quad 1 - \Phi(x) = 0.01 \Leftrightarrow x \approx 2.327$
Dann ist gefordert, daß

$$\frac{N - np}{\sqrt{npq}} \geq 2.327$$

mit $n = 2000$, $p = \frac{1}{30} \Rightarrow N \geq 85.350527$, d.h. 86 würden reichen.

Poisson-Verteilung und Warteschlangen (vgl. Übung, Serie 10, A3)

Satz 14.4. *Ist $N(s,t)$ die Anzahl der Kunden, die an einem Schalter in der Zeit zwischen s und t ankommen, und setzen wir voraus, daß*
(i) Die Anzahl der Kunden in disjunkten Zeitintervallen ist unabhängig und

(ii) ihre Verteilung hängt nur von $t-s$ ab,
(iii) $P[N(0,h)=1]=\lambda h+o(h)$,

(iv) $P[N(0,h)\geq 2]=o(h)$,

Nun ist $N(t):=N(0,t)$ Poisson-verteilt mit Parameter λt.

Denn: t fest

$$\begin{aligned} X_{n,m} &:= 1, \text{ falls } N((m-1)t/n, mt/n)=1 \\ X_{n,m} &:= 0 \text{ sonst} \end{aligned}$$

Dann ist $p_{n,m}=P[X_{n,m}=1]=\lambda\frac{t}{n}+o\left(\frac{t}{n}\right)$ und somit $\sum\limits_{m=1}^{n} p_{n,m}\to\lambda t$, woraus:

$$\sum_{m=1}^{n} X_{n,m}\Rightarrow Z$$

mit Poisson-Verteilung (λt) folgt:

$$\begin{aligned} & P\left[\sum_{m=1}^{n} X_{n,m}\neq N(t)\right]\leq\sum_{m=1}^{n} P[X_{n,m}\neq N((m-1)t/n, mt/n)] \\ & \sum_{m=1}^{n} n\cdot o\left(\frac{t}{n}\right)\to 0, \text{ d.h.} \\ & \sum_{m=1}^{n} X_{n,m}-N(t)\to 0 \quad P\text{-stochastisch} \end{aligned}$$

Anwendung des Satzes von Slutsky (Übung, Serie 9, A2)

$$N(t)=\underbrace{N(t)-\sum_{m=1}^{n} X_{n,m}}_{\substack{\to 0 \\ \text{in Wahrscheinlichkeit}}}+\underbrace{\sum_{m=1}^{n} X_{n,m}}_{Y\sim\text{Poisson}(\lambda t)\text{-verteilt}}\Rightarrow Y$$

$(N_t)_{t\geq 0}$ heißt dann Poisson-Prozeß.
Dieser ergibt sich auch aus

$$N_t=\max\{k: X_1+\ldots+X_k\leq t\} \text{ mit } X_i \text{ i.i.d. } \exp(\lambda).$$

Das sieht man aus dem Beispiel a) über die Verteilung von Punkten auf $[0, 2n]$.

$$\xi_{n,i}, \ \ i = 1, \dots, n \text{ unabhängig und gleichverteilt auf } [0, 2n].$$

Ordne

$$\xi_{[1],n} < \dots < \xi_{[n],n}$$

Dann für $s < t$

$$\begin{aligned} P[\xi_{[k],n} < s < t < \xi_{[k+1],n}] &= \frac{n!}{k!(n-k)!} \left(\frac{s}{2n}\right)^k \left(\frac{2n-t}{2n}\right)^{n-k} \\ &= \int_0^s \int_t^{2n} \binom{n}{k} k(n-k) \left(\frac{n}{2n}\right)^{k-1} \left(\frac{2n-v}{2n}\right)^{n-k-1} dndv \end{aligned}$$

Aus dieser gemeinsamen Verteilung erhält man die Verteilung von

$$P[\xi_{[k+1],n} - \xi_{[k],n} \leq x] = \int_0^x \frac{1}{2} \left(1 - \frac{y}{2n}\right)^{n-1} dy \overset{n\to\infty}{\to} \int_0^x \frac{1}{2} e^{-\frac{1}{2}y} dy,$$

d.h. exponential $(\frac{1}{2})$-verteilt.

Kapitel 15

Der Satz vom iterierten Logarithmus

Vorgehen nach:

- V.S. Borkar: Prob. Theory. An advanced course. Springer

- K.L. Chung: A first course in Prob. Theory. Academic Press '74

a) Ausgangsproblem

Gegeben $(\Omega, \mathcal{A}, P)$, $(X_i)_{i \in \mathbb{N}}$ unabhängig, identisch verteilt $\in \mathcal{L}^2$ mit $E(X_i) = 0$ und $0 < \sigma^2 = \mathrm{var}(X_i) < \infty$.
Wie verhält sich $S_n(\omega)$ für große n?

Antworten:

1.a) : $\forall \epsilon > 0, \forall c > 0, \forall$ hinreichend großen n :

$$P\left(\frac{1}{n}|S_n| < c\right) \geq 1 - \epsilon.$$

Also Aussage für individuelles $n \in \mathbb{N}$.

b) (Starkes Gesetz): $P\left(\limsup_n \frac{1}{n}|S_n| = 0\right) = 1$ Aussage für individuelles $\omega \in \Omega$

2.a) (ZGS): $\forall c > 0, \forall$ hinreichend großen n :

$$P\left(\frac{1}{\sqrt{n}}|S_n| < c \cdot \sigma\right) \approx \nu_{0,1}([-c, c]) \text{ Aussage für individuelles } n \in \mathbb{N}$$

(z.B. "3σ-Regel": $P\left(\frac{1}{\sqrt{n}} S_n > 3\sigma\right) \approx 0.001$

$= P\left(\frac{1}{\sqrt{n}} S_n < -3\sigma\right)$)

Gesucht Folge $(a_n)_n$ isoton, reell mit

$$P\left(\limsup_{n\to\infty} \frac{1}{a_n}|S_n| = c\right) = 1 \text{ Aussage für individuelles } \omega \in \Omega$$

für ein $0 < c < \infty$.

Hinweis:
Die Funktion $\limsup \frac{1}{a_n}|S_n|$ ist stets meßbar bzgl. der terminalen σ-Algebra.
Also gilt: $\exists c \in [0, \infty]$ mit $\limsup_{n\to\infty} \frac{1}{a_n}|S_n|(\omega) = c$ für P-fast alle $\omega \in \Omega$.

Gesucht sind also $(a_n)_n$ mit $0 < c < \infty$.

Nun gilt:

- $a_n = n \Rightarrow c = 0$ (s.o.)

- $a_n = \sqrt{n} \cdot \log n \Rightarrow c = 0$, denn nach dem Starken Gesetz der GZ von Kolmogorov gilt $\frac{1}{a_n} S_n \to 0$ P-f.s., falls $\sum \frac{1}{a_n^2} < \infty$.

- $a_n = \sqrt{n} \Rightarrow c = \infty$, denn $\forall c < \infty$ gilt mit $A_n = \{\omega : \frac{1}{\sqrt{n}}|S_n| < c\}$.

$$\begin{aligned}
P\left(\left\{\limsup_{n\to\infty} \frac{1}{\sqrt{n}}|S_n| < c\right\}\right) &= P\left(\bigcup_n \bigcap_{k\geq n} A_k\right) \\
&= \lim_{n\to\infty} P\left(\bigcap_{k\geq n} A_k\right) \leq \limsup_{n\to\infty} P(A_n) \\
&= \nu_{0,\sigma^2}(]-c,c[) < 1 \\
&\Rightarrow P\left(\left\{\limsup \frac{1}{\sqrt{n}}|S_n| < c\right\}\right) = 0
\end{aligned}$$

- $a_n = \sqrt{2n \cdot \log(\log n)}$ (“iterierter Logarithmus“) $\Rightarrow c = \sigma$

  Khintchine ’24 (“normale Zahlen“)
  Kolmogorov ’29 ($\mathcal{L}^3$, unabhängig)
  Hartman, Wintner ’41 (i.i.d., $\mathcal{L}^2$)
  Strassen ’64

b) Voraussetzungen

$(\Omega, \mathcal{A}, P)$, $(X_i)_{i\in\mathbb{N}}$ unabhängig $\in \mathcal{L}^3$ (<u>nicht</u> identisch verteilt) mit $E(X_i) = 0$. Sei $S_n = \sum_{i=1}^{n} X_i$. Setze

$$\begin{aligned}
s_n &= \left[\sum_{i=1}^{n} E(|X_i|^2)\right]^{\frac{1}{2}} = \left[\sum \text{var}X_i\right]^{\frac{1}{2}} = [\text{var}S_n]^{\frac{1}{2}} \\
\text{und} \quad \xi_n &= \left[\sum_{i=1}^{n} E(|X_i|^3)\right]^{\frac{1}{3}}.
\end{aligned}$$

<u>Annahmen</u>:

(i) $\quad s_n \to \infty \quad (n\to\infty)$
und $\exists \epsilon > 0$:
(ii) $\quad (\log s_n)^{1+\epsilon} \cdot \left(\frac{\xi_n}{s_n}\right)^3 \to 0 \quad (n \to \infty)$.

Bemerkung: Sind die $(X_i)_i$ identisch verteilt mit $\sigma = (\text{var}X_i)^{\frac{1}{2}} > 0$ und $\tau = (E|X_i|^3)^{\frac{1}{3}}$, so ist

$$s_n = \sqrt{n} \cdot \sigma \to \infty$$

und

$$(\log s_n)^{1+\epsilon} \cdot \left(\frac{\xi_n}{s_n}\right)^3 = \left(\frac{1}{2}\log n + \log\sigma\right)^{1+\epsilon} \cdot \left(\frac{\tau \cdot n^{\frac{1}{3}}}{\sigma \cdot n^{\frac{1}{2}}}\right)^3 = \frac{c}{n^{\frac{1}{2}}} \to 0 \text{ für } n \to \infty.$$

Also: Voraussetzung (i) und (ii) erfüllt.

Definition 15.1. $\varphi_\lambda(x) = \sqrt{2\lambda \cdot x^2 \cdot \log\log x}$ *für* $\lambda > 0$ *und* $x > e$.

Satz 15.2 (Gesetz vom iterierten Logarithmus).
Unter obigen Vor.:

- $\limsup_{n\to\infty} \frac{S_n}{\sqrt{2s_n^2 \log\log s_n}} = 1 \qquad P\text{-f.s.}$
- $\liminf_{n\to\infty} \frac{S_n}{\sqrt{2s_n^2 \log\log s_n}} = -1 \qquad P\text{-f.s.}$

Lemma 15.3.
a) Aus Bedingung (ii) folgt die "Lyapunov-Bedingung":

$$\exists \delta > 0: \qquad \frac{1}{s_n^{2+\delta}} \cdot \sum_{i=1}^n E(|X_i|^{2+\delta}) \to 0 \quad (n \to \infty)$$

(und zwar mit $\delta = 1$).

b) Aus der Lyapunov-Bedingung folgt die Lindeberg-Bedingung.

Beweis:
a) Offensichtlich.

b) $\forall \epsilon > 0$ und $|x| \geq \epsilon \cdot s_n$ gilt: $|x|^{2+\delta} \geq |x|^2 \cdot (\epsilon \cdot s_n)^\delta$

Also:

$$\begin{aligned}\frac{1}{s_n^2}\sum_{j=1}^n \int_{\{|X_j| \geq \epsilon\cdot s_n\}} X_j^2 dP &\leq \frac{1}{\epsilon^\delta \cdot s_n^{2+\delta}} \sum_{j=1}^n \int_{\{|X_j| \geq \epsilon\cdot s_n\}} |X_j|^{2+\delta} \\ &\leq \frac{1}{\epsilon^\delta} \cdot \frac{1}{s_n^{2+\delta}} \sum_{j=1}^n E(|X_j|^{2+\delta}).\end{aligned}$$

□

Zum Beweis des Satzes werden 3 Propositionen und 3 Lemmata benötigt.

Lemma 15.4.

$$\begin{aligned}\int_x^\infty e^{-\frac{y^2}{2}} dy &\sim \frac{1}{x} e^{-\frac{x^2}{2}} \qquad (\forall x > 1)\\ \left(\frac{1}{x} - \frac{1}{x^3}\right) e^{-\frac{x^2}{2}} &\leq \int_x^\infty e^{-\frac{y^2}{2}} dy \leq \frac{1}{x} e^{-\frac{x^2}{2}} \quad (\forall x > 0)\end{aligned}$$

Beweis:

- $\int_x^\infty e^{-\frac{y^2}{2}} dy = \int_x^\infty e^{-\frac{x^2}{2}} \cdot e^{-\frac{y^2}{2}} \cdot e^{-xy} dy$
  $\leq e^{-\frac{x^2}{2}} \int_x^\infty e^{-xy} dy = \frac{1}{x} e^{-\frac{x^2}{2}}$
- $\int_x^\infty e^{-\frac{y^2}{2}} dy \geq \int_x^\infty \left(1 - \frac{3}{y^4}\right) e^{-\frac{y^2}{2}} = \left(\frac{1}{x} - \frac{1}{x^3}\right) e^{-\frac{x^2}{2}}$ □

Proposition 15.5. $\forall \delta \in]0, \epsilon[: \exists C_1, C_2, n_0 : \forall n \geq n_0$:

$$\begin{aligned} P(S_n > \varphi_{1+\delta}(s_n)) &\leq \frac{C_1}{(\log s_n)^{1+\delta}} \\ P(S_n > \varphi_{1-\delta}(s_n)) &\geq \frac{C_2}{(\log s_n)^{1-\frac{\delta}{2}}}. \end{aligned}$$

Beweis: Nach Berry-Esséen gilt: $\exists C : \forall x, \forall n$:

$$\left| P(S_n > x \cdot s_n) - \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{y^2}{2}} dy \right| \leq C \cdot \left(\frac{\xi_n}{s_n}\right)^3 \tag{15.1}$$

Ferner gilt für $x \to \infty$ (siehe Übung)

$$\int_x^\infty e^{-\frac{y^2}{2}} dy \approx \frac{1}{x} \cdot e^{-\frac{x^2}{2}}$$

Für $x = \sqrt{2(1 \pm \delta) \log \log s_n}$ gilt daher

$$\frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{y^2}{2}} dy \approx \frac{1}{\sqrt{4\pi(1 \pm \delta) \log \log s_n}} \cdot \frac{1}{(\log s_n)^{1 \pm \delta}}. \tag{15.2}$$

Nach Voraussetzung (ii) gilt für $\epsilon > \delta : \exists C'$:

$$\left(\frac{\xi_n}{s_n}\right)^3 \leq \frac{C'}{(\log s_n)^{1+\epsilon}} \tag{15.3}$$

Also mit (15.1)-(15.3):

$$\begin{aligned} P(S_n > \varphi_{1\pm\delta}(s_n)) &\approx \frac{1}{\sqrt{2\pi}} \int\limits_{\sqrt{2(1\pm\delta)\log\log n}} e^{-\frac{y^2}{2}} dy \\ &\approx \frac{1}{\sqrt{4\pi(1\pm\delta)\log\log s_n}} \cdot \frac{1}{(\log s_n)^{1\pm\delta}} \begin{cases} \leq & \frac{C_1}{(\log s_n)^{1+\delta}} \\ \geq & \frac{C_2}{(\log s_n)^{1-\frac{\delta}{2}}} \end{cases} \end{aligned}$$

□

Lemma 15.6. *Seien $E_i, F_i \in \mathcal{A}$ $(i = 1, \ldots, n)$ Ereignisse mit:*

- $\forall j : F_j$ *ist unabhängig von* $E_1^c \cap E_2^c \cap \ldots \cap E_{j-1}^c \cap E_j$ *und*
- $\exists a > 0 : \forall j : P(F_j) \geq a.$

Dann: $$P\left(\bigcup_{j=1}^n E_j \cap F_j\right) \geq a \cdot P\left(\bigcup_{j=1}^n E_j\right).$$

Beweis:

$$\begin{aligned} P\left(\bigcup_{j=1}^n (E_j \cap F_j)\right) &= P\left(\dot{\bigcup_{j=1}^n} (E_1 \cap F_1)^c \cap \ldots \cap (E_{j-1} \cap F_{j-1})^c \cap (E_j \cap F_j)\right) \\ &\geq P\left(\dot{\bigcup_{j=1}^n} E_1^c \cap \ldots \cap E_{j-1}^c \cap E_j \cap F_j\right) \\ &= \sum_{j=1}^n P(E_1^c \cap \ldots \cap E_{j-1}^c \cap E_j) \cdot P(F_j) \\ &\geq \left[\sum_{j=1}^n P(E_1^c \cap \ldots \cap E_{j-1}^c \cap E_j)\right] \cdot a \\ &= a \cdot P\left(\bigcup_{j=1}^n E_j\right). \end{aligned}$$ □

Betrachte nun die Mengen

$$\begin{aligned} E_n^+ &= \{\omega : S_n(\omega) > \varphi_{1+\delta}(s_n)\} \text{ und} \\ E_n^- &= \{\omega : S_n(\omega) > \varphi_{1-\delta}(s_n)\}. \end{aligned}$$

Proposition 15.7.

$$P\left(\bigcap_k \bigcup_{n\geq k} E_n^+\right) = P(\limsup E_n^+) = P(E_n^+ \textit{ unendlich oft }) = 0$$

Beweis: Sei $c > 1$ beliebig vorgegeben.
$\Rightarrow \exists$ isotone Folge $(n(k))_k \subseteq \mathbb{N}$ mit

$$s_{n(k)} \leq c^k < s_{n(k)+1}.$$

Behauptung 1: $s_{n(k)} \approx c^k$ für $k \to \infty$.
Denn nach Lemma 15.3 gilt die Lindeberg-Bedingung und somit : $\max\limits_{1 \leq j \leq n} \frac{\sigma_j}{s_n} \to 0$ $(\forall n \to \infty)$.

Daher $\dfrac{s_{n(k)+1}}{s_{n(k)}} \to 1$ $(k \to \infty)$ und folglich $\dfrac{c^k}{s_{n(k)}} \to 1$.

Sei nun $k \in \mathbb{N}$ (fix) und $j \in \mathbb{N}$ mit $n(k) \leq j < n(k+1)$.
Definiere

$$F_j = \{\omega : |S_{n(k+1)}(\omega) - S_j(\omega)| < s_{n(k+1)}\}.$$

Nach Chebyshev:

$$P(F_j) \geq 1 - \frac{\operatorname{var}(S_{n(k+1)} - S_j)}{s^2_{n(k+1)}} \geq 1 - \frac{s^2_{n(k+1)} - s^2_{n(k)}}{s^2_{n(k+1)}} = \left(\frac{s_{n(k)}}{s_{n(k+1)}}\right)^2 \to \frac{1}{c^2}$$

Also für hinreichend große k:

$$P(F_j) \geq \frac{1}{2c^2} \qquad (\forall j = n(k), \ldots, n(k+1) - 1)$$

Mit Lemma 15.6 folgt hieraus: Behauptung 2:

$$P\left(\bigcup_{j=n(k)}^{n(k+1)-1} E_j^+ \cap F_j\right) \geq \frac{1}{2c^2} \cdot P\left(\bigcup_{j=n(k)}^{n(k+1)-1} E_j^+\right)$$

denn $F_j \in \sigma(X_i : i \geq j+1)$ ist unabhängig von $E_1, \ldots, E_j \in \sigma(X_i : i \leq j)$.
Nun ist

$$E_j^+ \cap F_j \subset \{\omega : S_{n(k+1)}(\omega) > S_j(\omega) - s_{n(k+1)} > \varphi_{1+\delta}(s_j) - s_{n(k+1)}\}. \tag{15.1}$$

Behauptung 3: $\forall \beta \in]0,1[$: $\forall$ hinreichend große k : $\forall j = n(k), \ldots, n(k+1) - 1$:

$$\varphi_{1+\delta}(s_j) - s_{n(k+1)} > \varphi_{\beta \cdot \left(\frac{1+\delta}{c^2}\right)}(s_{n(k+1)}). \tag{15.2}$$

Denn:

$$\begin{aligned}
\varphi_{1+\delta}(s_j) &\geq \varphi_{1+\delta}(s_{n(k)}) \approx \varphi_{1+\delta}\left(\frac{s_{n(k+1)}}{c}\right) \approx \varphi_{\frac{1+\delta}{c^2}}(s_{n(k+1)}) \\
\varphi_{\beta \cdot \left(\frac{1+\delta}{c^2}\right)}(s_{n(k+1)}) &= \sqrt{\beta} \cdot \varphi_{\frac{1+\delta}{c^2}}(s_{n(k+1)}) \\
s_{n(k+1)} &= o(\varphi_{\frac{1+\delta}{c^2}}(s_{n(k+1)}))
\end{aligned}$$

Wähle nun $\beta < 1$ und $c > 1$ beide hinreichend nahe bei 1, so daß

$$\beta \cdot \left(\frac{1+\delta}{c^2}\right) > 1 + \frac{\delta}{2} \tag{15.3}$$

und setze

$$G_k = \{\omega : S_{n(k+1)}(\omega) > \varphi_{1+\frac{\delta}{2}}(s_{n(k+1)})\}.$$

[Dann enspricht G_k dem $E^+_{n(k+1)}$ mit δ ersetzt durch $\frac{\delta}{2}$.]
Aus (15.1.),(15.2.),(15.3.) folgt:

$$E_j^+ \cap F_j \subset G_k$$

für k hinreichend groß und $j = n(k), \ldots, n(k+1) - 1$.
Insbesondere also

$$\bigcup_{j=n(k)}^{n(k+1)-1} E_j^+ \cap F_j \subset G_k. \tag{15.4}$$

Also

$$\begin{aligned}
\sum_{k=1}^{\infty} P\left(\bigcup_{j=n(k)}^{n(k+1)-1} E_j^+\right) &\overset{\text{Beh.(2)}}{\leq} 2c^2 \cdot \sum_{k=1}^{\infty} P\left(\bigcup_{j=n(k)}^{n(k+1)-1} E_j^+ \cap F_j\right) \\
&\overset{(15.4)}{\leq} 2c^2 \cdot \sum_{k=1}^{\infty} P(G_k) \\
&\overset{\text{Prop.15.5}}{\leq} 2c^2 \cdot \sum_{k=1}^{\infty} \frac{C_1}{(\log s_{n(k+1)})^{1+\frac{\delta}{2}}} \\
&\leq 2c^2 \cdot \sum_{k=2}^{\infty} \frac{C_1}{(k \cdot \log c)^{1+\frac{\delta}{2}}} < \infty.
\end{aligned}$$

Nach Borel-Cantelli-Lemma (trivialer Teil) folgt daraus:

$$P\left(\left(\bigcup_{j=n(k)}^{n(k+1)-1} E_j^+\right) \text{ unendlich oft}\right) = 0.$$

Das ist äquivalent zur Behauptung. $\square$

Sei nun wieder $c > 1$ beliebig und $(n(k))_{k \in \mathbb{N}}$ wie eben.
Setze $t_k^2 = s^2_{n(k+1)} - s^2_{n(k)}$ und

$$D_k = \{\omega : S_{n(k+1)}(\omega) - S_{n(k)}(\omega) > \varphi_{1-\frac{\delta}{2}}(t_k)\}.$$

Lemma 15.8. $P(D_k \text{ unendlich oft }) = 1$

Beweis: Die ZV $(S_{n(k+1)} - S_{n(k)})_{k\in\mathbb{N}}$ sind unabhängig, folglich auch die Mengen $(D_k)_{k\in\mathbb{N}}$. Ferner gilt

$$t_k^2 \approx \left(1-\frac{1}{c^2}\right) s_{n(k+1)}^2 \approx \left(1-\frac{1}{c^2}\right) c^{2(k+1)}.$$

Daher nach Voraussetzung (ii):

$$\begin{aligned}\frac{\xi_{n(k+1)}^3 - \xi_{n(k)}^3}{t_k^3} &\leq \left(\frac{\xi_{n(k+1)}}{t_k}\right)^3 \approx \left(1-\frac{1}{c^2}\right)^{\frac{3}{2}} \cdot \left(\frac{\xi_{n(k+1)}}{s_{n(k+1)}}\right)^3 \\ &= o\left(\frac{1}{(\log s_{n(k+1)})^{1+\epsilon}}\right) = o\left(\frac{1}{(\log t_k)^{1+\epsilon}}\right)\end{aligned}$$

für große k.
Anwendung von Proposition 15.5 auf $S_{n(k+1)} - S_{n(k)}$ (statt S_n) und $\frac{\delta}{2}$ (statt δ) liefert

$$P(D_k) \geq \frac{C}{(\log t_k)^{1-\frac{\delta}{4}}} \sim \frac{C'}{k^{1-\frac{\delta}{4}}}$$

$\Rightarrow \sum_k P(D_k) = \infty$
$\Rightarrow$ (nicht-trivialer Teil von Borel-Cantelli:) $P(D_k$ unendlich oft $) = 1$ □

Proposition 15.9. $P(E_n^- \text{ unendlich oft }) = 1$

Beweis: Nach Proposition 15.7 (angewandt auf $-X_n$ mit $\delta = 1$) gilt P-f.s.

$$S_{n(k)} \geq -\varphi_2(s_{n(k)}) \quad \text{für alle hinreichend großen } k. \tag{15.1}$$

Nach Lemma 15.8 gilt P-f.s.

$$S_{n(k+1)} - S_{n(k)} \geq \varphi_{1-\frac{\delta}{2}}(t_k) \quad \text{für unendlich viele } k. \tag{15.2}$$

Also P-f.s.

$$S_{n(k+1)} \geq \varphi_{1-\frac{\delta}{2}}(t_k) - \varphi_2(s_{n(k)}) \quad \text{für unendlich viele } k.$$

Nun ist $\log\log t_k^2 \approx \log\log s_{n(k)}^2$ und $s_{n(k)} \approx c^k$.
Also ist $\forall \alpha < 1, \forall$ hinreichend große k:

$$\varphi_{1-\frac{\delta}{2}}(t_k) - \varphi_2(s_{n(k)}) \geq \alpha \cdot \left(\sqrt{1-\frac{\delta}{2}} \cdot \sqrt{1-\frac{1}{c^2}} - \sqrt{\frac{2}{c^2}}\right) \cdot \varphi_1(s_{n(k+1)}).$$

Letzteres ist $> \varphi_{1-\delta}(s_{n(k+1)})$, falls man c hinreichend groß und α hinreichend nahe an 1 wählt.
Dann folgt P-f.s.

$$S_{n(k+1)} \geq \varphi_{1-\delta}(s_{n(k+1)}) \quad \text{für unendlich viele } k. \square$$

Beweis des Satzes 15.2:
Nach Proposition 15.7 und 15.9 gilt $\forall \delta < \epsilon$:

$$1 - \delta \leq \limsup_{n\to\infty} \frac{S_n}{\varphi_1(s_n)} \leq 1 + \delta \qquad P\text{-f.s.}$$

Wähle nun $\delta = \frac{\epsilon}{n}$ für $n = 2, 3, \ldots$. Dann folgt

$$\limsup_{n\to\infty} \frac{S_n}{\varphi_1(s_n)} = 1 \qquad P\text{-f.s..}$$

Die zweite Behauptung des Satzes folgt daraus durch Übergang von $(X_i)_i$ zu $(-X_i)_i$. $\square$

Satz 15.10. (von Hartman, Wintner) *Seien $(X_i)_{i\in\mathbb{N}}$ unabhängig, identisch verteilt $\in \mathcal{L}^2$ mit $E(X_i) = 0$ und*
$\sigma^2 = var(X_i) < \infty$. Dann gilt P-f.s.

$$\begin{aligned} \limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} &= \sigma \qquad \textit{und} \\ \liminf_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} &= -\sigma. \end{aligned}$$

Beweis: Falls zusätzlich $X_i \in \mathcal{L}^3$: s.o, ansonsten: siehe Bauer $\square$

c) Ausblicke

Satz 15.11. (Strassen) *Unter den eben gemachten Voraussetzungen gilt:*
Die Menge der Häufungspunkte der Folge

$$\left(\frac{S_n(\omega)}{\sqrt{2n\log\log n}}\right)_{n\in\mathbb{N}}$$

ist genau das Intervall $[-\sigma, \sigma]$.

Seien $(X_i)_{i\in\mathbb{N}}$ unabhängig, identisch verteilt mit $E(X_i) = 0, \operatorname{var}(X_i) = 1$.
Sei $S_n = \sum_{i=1}^{n} X_i$ und $S : t \mapsto S(t)$ die lineare Interpolation von $n \mapsto S_n$.

Betrachte $Z_n : \; [0,1] \times \Omega \longrightarrow \mathbb{R}$,
$(t,\omega) \mapsto Z_n(t)(\omega) = \frac{S(n\cdot t)}{\sqrt{2n \log\log n}}$,

d.h. für jedes $\omega \in \Omega$ ist $Z_n(.)(\omega)$ ein Element aus $\mathcal{C}([0,1],\mathbb{R})$.

Satz 15.12.
(Strassens Gesetz vom iteriertem Logarithmus, Strassens Invarianzprinzip):
Für P-f.a. $\omega \in \Omega$ ist die Menge der Häufungspunkte der Folge $(Z_n(.)(\omega))_{n\in\mathbb{N}}$ genau

$$\mathcal{K} = \{f \in \mathcal{C}([0,1],\mathbb{R}) : f(x) = \int_0^x g(y)dy \textit{ mit } \int_0^1 g^2(y)dy \leq 1\}.$$

Hinweis: Für $f \in \mathcal{K}$ gilt $f(1)^2 = \left(\int_0^1 g(y)dy\right)^2 \leq \int_0^1 g^2(y)dy \leq 1$.