Wahrscheinlichkeitstheorie

Skript[1] zur Vorlesung von

Prof. Dr. Michael Kohler

Wintersemester 2011/12

[1]Dieses Skript basiert auf dem Exzerpt zur Vorlesung "Wahrscheinlichkeitstheorie" von Prof. Dr. Harro Walk (Universität Stuttgart), dessen elektronische Form mir freundlicherweise von Priv.-Doz. Dr. Jürgen Dippon zur Verfügung gestellt wurde.

1 Grundlagen aus der Maßtheorie

1.1 Definition. Sei Ω eine nichtleere Menge. Ein System $\mathcal{A}$ von Teilmengen von Ω heißt eine **Algebra**, wenn gilt:

1) $\emptyset \in \mathcal{A}$,

2) $A \in \mathcal{A} \Longrightarrow A^c := \Omega \backslash A \in \mathcal{A}$,

3) $A, B \in \mathcal{A} \Longrightarrow A \cup B \in \mathcal{A}$.

Das System $\mathcal{A}$ heißt **σ-Algebra**, falls zusätzlich gilt

4) $A_n \in \mathcal{A} \quad (n = 1, 2 \ldots) \Rightarrow \bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{A}$.

1.2 Bemerkung. $\mathcal{A}$ σ-Algebra $\Longrightarrow$ $\mathcal{A}$ Algebra.

1.3 Lemma.

a) Sei $\mathcal{A}$ eine Algebra. D. g.:
$A_1, \ldots, A_m \in \mathcal{A} \Longrightarrow \bigcup\limits_{n=1}^{m} A_n \in \mathcal{A}, \quad \bigcap\limits_{n=1}^{m} A_n \in \mathcal{A}$
$A, B \in \mathcal{A} \Longrightarrow A \backslash B \in \mathcal{A}$.

b) Sei $\mathcal{A}$ σ-Algebra. D. g.:
$A_n \in \mathcal{A}\ (n \in \mathbb{N}) \Longrightarrow \bigcap\limits_{n=1}^{\infty} A_n \in \mathcal{A}$.

Beweis:

a) Es gilt $A \cap B = \Omega \backslash (A^c \cup B^c) \in \mathcal{A}$ nach 2) und 3), woraus auch $A \backslash B = A \cap B^c \in \mathcal{A}$ folgt, Rest ergibt sich mit Induktion.

b) Folgt aus $\bigcap\limits_{n=1}^{\infty} A_n = \left(\bigcup\limits_{n=1}^{\infty} A_n^c \right)^c$ nach de Morgan und 2), 4).

$\square$

Eine Algebra [σ-Algebra] enthält $\emptyset, \Omega$ und ist abgeschlossen gegenüber endlich [abzählbar] vielen üblichen Mengenoperationen.

1.4 Definition. Es sei Ω eine nichtleere Menge, $\mathcal{A}$ eine σ-Algebra in Ω. Das Paar $(\Omega, \mathcal{A})$ heißt Messraum; die Elemente $\in \mathcal{A}$ heißen $\mathcal{A}$ **messbare** Mengen.

1.5 Lemma.

a) $\mathcal{A}_\alpha$ σ-Algebra in Ω ($\alpha \in I \neq \emptyset$) $\Longrightarrow \bigcap\limits_{\alpha \in I} \mathcal{A}_\alpha$ σ-Algebra in Ω.

b) $\mathcal{C}$ Mengensystem in $\Omega \Longrightarrow$ es existiert eine (bzgl. Ω) kleinste $\mathcal{C}$ enthaltene σ-Algebra.

Entsprechend für Algebren.

Beweis: Siehe Integrationstheorie bzw. Maßtheorie. In b) betrachtet man

$$\bigcap_{\mathcal{A}\ \sigma-\text{Algebra in }\Omega\text{ mit }\mathcal{C}\subset\mathcal{A}} \mathcal{A}$$

□

1.6 Definition. $\mathcal{C}$ sei ein Mengensystem in Ω. Die kleinste der $\mathcal{C}$ enthaltenden σ-Algebren heißt die von $\mathcal{C}$ erzeugte σ-Algebra; $\mathcal{C}$ heißt ein **Erzeugersystem** dieser σ-Algebra. – Entsprechend für Algebren.
Bezeichnung: $\mathcal{F}(\mathcal{C})\ldots$ die von $\mathcal{C}$ erzeugte σ-Algebra.

1.7 Definition. $\mathcal{O}_n :=$ System der offenen Mengen des Euklidischen Raumes $\mathbb{R}^n; \mathcal{B}_n := \mathcal{F}(\mathcal{O}_n), \mathcal{B} = \mathcal{B}_1$. $\mathcal{B}_n$ wird als σ-Algebra der **Borelschen Mengen in** $\mathbb{R}^n$ bezeichnet.

$\mathcal{B}$ enthält alle

- offenen Mengen
- abgeschlossenen Mengen (da Komplement der offenen)
- Einpunktmengen (da abgeschlossen)
- höchstens abzählbaren Mengen (da abzählbare Vereinigungen von Einpunktmengen)

Es gilt $\mathcal{B}_n \underset{\neq}{\subset} \mathcal{P}(\mathbb{R}^n)$.

1.8 Satz. Das System $\mathcal{J}_n$ aller halboffenen Intervalle

$$(a,b] := \{(x_1,\ldots,x_n) \in \mathbb{R}^n : a_i < x_i \leq b_i\ (i=1,\ldots,n)\}$$

in $\mathbb{R}^n$ ($a = (a_1,\ldots,a_n), b = (b_1,\ldots,b_n) \in \mathbb{R}^n$ mit $a_i < b_i$) [oder auch das System der offenen Intervalle, der abgeschlossenen Intervalle, der Intervalle $(-\infty, a]$ oder der Intervalle $(-\infty, a)$] ist ein Erzeugersystem von $\mathcal{B}_n$.

Beweis: Siehe Integrationstheorie bzw. Maßtheorie. $\square$

1.9 Definition. $\mathcal{C}$ sei ein Mengensystem in Ω mit $\emptyset \in \mathcal{C}$. Eine Mengenfunktion $\mu : \mathcal{C} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty, -\infty\}$ heißt (mit $a + \infty = \infty$ $(a \in \mathbb{R})$, $\infty + \infty = \infty$ usw.)

a) ein **Inhalt** (content) auf $\mathcal{C}$, wenn

μ ist nulltreu, d.h. $\mu(\emptyset) = 0$,

μ positiv, d.h. $\underset{A\in\mathcal{C}}{\forall} \quad \mu(A) \geq 0$,

μ additiv, d.h. für alle paarweise disjunkten $A_n \in \mathcal{C}$ $(n = 1, 2, \ldots, m)$

$$\text{mit} \quad \sum_{n=1}^{m} A_n := \cup_{n=1}^{m} A_n \in \mathcal{C} \quad \text{gilt} \quad \mu\left(\sum_{n=1}^{m} A_n\right) = \sum_{n=1}^{m} \mu(A_n) \ .$$

b) ein **Maß** (measure) auf $\mathcal{C}$, wenn

μ ist nulltreu, d.h. $\mu(\emptyset) = 0$,

μ positiv, d.h. $\underset{A\in\mathcal{C}}{\forall} \quad \mu(A) \geq 0$,

μ σ-additiv, d.h. für alle paarweise disjunkten $A_n \in \mathcal{C}$ $(n = 1, 2, \ldots)$

$$\text{mit} \quad \sum_{n=1}^{\infty} A_n \in \mathcal{C} \quad \text{gilt} \quad \mu\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) \ .$$

"Natürliche" Definitionsbereiche für Inhalt und Maß sind Algebren bzw. σ-Algebren (dann Verzicht auf Voraussetzung $\sum\limits_{n=1}^{m} A_n \in \mathcal{C}$ bzw. $\sum\limits_{n=1}^{\infty} A_n \in \mathcal{C}$). Häufig werden nur diese Definitionsbereiche verwendet.

Klar: Jedes Maß ist ein Inhalt, da

$$\mu\left(\sum_{n=1}^{m} A_n\right) = \mu\left(\sum_{n=1}^{\infty} A_n\right) \ \text{mit } A_n = \emptyset \text{ für } n > m.$$

1.10 Definition. Ist $\mathcal{A}$ eine σ-Algebra in Ω, μ ein Maß auf $\mathcal{A}$, so heißt $(\Omega, \mathcal{A}, \mu)$ ein **Maßraum**.

Ist $(\Omega, \mathcal{A})$ ein Messraum, $Z = \{z_1, z_2, \ldots\}$ eine höchstens abzählbare Teilmenge von Ω, so heißt $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ mit

$$\mu(A) := \text{ Anzahl der } z_i \in A, A \in \mathcal{A},$$

ein **abzählbares Maß**.

1.11 Lemma. Sei $\mathcal{A}$ eine Algebra in Ω, μ ein Inhalt auf $\mathcal{A}$.

a) μ ist monoton, d.h. $[A, B \in \mathcal{A},\ A \subset B \Longrightarrow \mu(A) \leq \mu(B)]$.

b) μ ist subadditiv, d.h. $[A_1, \ldots, A_m \in \mathcal{A} \Longrightarrow \mu(\bigcup\limits_{n=1}^{m} A_n) \leq \sum\limits_{n=1}^{m} \mu(A_n)]$.

c) Ist μ Maß, dann ist μ sogar σ-subadditiv, d.h. $[A_n \in \mathcal{A}\ (n = 1, 2, \ldots)),\ \bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{A} \Longrightarrow \mu(\bigcup\limits_{n=1}^{\infty} A_n) \leq \sum\limits_{n=1}^{\infty} \mu(A_n)]$.

Beweis:

a) $B = B \backslash A \cup A \Longrightarrow \mu(B) = \mu(B \backslash A) + \mu(A) \geq \mu(A)$.

b)

$$\begin{aligned}
\mu\left(\bigcup_{n=1}^{m} A_n\right) &= \mu\left(A_1 \cup A_2 \backslash A_1 \cup A_3 \backslash (A_1 \cup A_2) \cup \ldots \cup A_m \backslash \left(\bigcup_{n=1}^{m-1} A_n\right)\right) \\
&= \mu(A_1) + \mu(A_2 \backslash A_1) + \ldots + \mu\left(A_m \backslash \left(\bigcup_{n=1}^{m-1} A_n\right)\right) \\
&\overset{a)}{\leq} \mu(A_1) + \mu(A_2) + \ldots + \mu(A_m).
\end{aligned}$$

c) Folgt analog wie b) aus der σ-Additivität des Maßes.

$\square$

1.12 Bemerkung. $\mathcal{A}$ Algebra in Ω, μ endlicher Inhalt auf $\mathcal{A}$.
$A, B \in \mathcal{A},\ A \subset B \Longrightarrow \mu(B \backslash A) = \mu(B) - \mu(A)\ \ldots$ Subtraktivität.

Bezeichnung [für Mengen A_n $(n = 1, 2, \ldots), A$]
$A_n \uparrow A : A_1 \subset A_2 \subset A_3 \subset \ldots,\ \bigcup\limits_{n=1}^{\infty} A_n = A$
$A_n \downarrow A : A_1 \supset A_2 \supset A_3 \supset \ldots$ und $\bigcap\limits_{n=1}^{\infty} A_n = A$

1.13 Satz. Sei $\mathcal{A}$ eine Algebra in Ω, μ ein **endlicher** Inhalt auf $\mathcal{A}$ (d.h. $\mu(\Omega) < \infty$). Dann sind die folgenden Aussagen äquivalent:

a) μ σ-additiv

b) μ stetig von unten,
d.h. $[A_n \in \mathcal{A},\ A_n \uparrow A \in \mathcal{A} \Longrightarrow \mu(A_n) \to \mu(A)]$

c) μ stetig von oben,
d.h. $[A_n \in \mathcal{A},\ A_n \downarrow A \in \mathcal{A} \Longrightarrow \mu(A_n) \to \mu(A)]$

d) μ $\emptyset$-stetig,
d.h. $[A_n \in \mathcal{A},\ A_n \downarrow \emptyset \Longrightarrow \mu(A_n) \to 0]$.

Auch ohne die Voraussetzung der Endlichkeit von μ gilt a) $\Longleftrightarrow$ b).

Beweis: Siehe Integrationstheorie bzw. Maßtheorie. $\square$

1.14 Definition. Sei $\mathcal{A}$ eine σ-Algebra in Ω. Eine Abbildung $P : \mathcal{A} \to \mathbb{R}$ heißt ein **Wahrscheinlichkeitsmaß** (W-Maß) auf $\mathcal{A}$, wenn

$P(\emptyset) = 0$,

P positiv, d.h. $\forall A \in \mathcal{A} : P(A) \geq 0$,

P σ-additiv, d.h. für alle paarweise disjunkten $A_n \in \mathcal{A}$ ($n = 1, 2, \ldots$) gilt

$$P\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n),$$

P normiert, d.h. $P(\Omega) = 1$.

$(\Omega, \mathcal{A}, P)$ heißt dann ein **Wahrscheinlichkeitsraum** (W-Raum).

Ein Wahrscheinlichkeitsraum ist ein normierter Maßraum.

Ist $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, so bezeichnet man die Grundmenge Ω auch als Merkmalraum oder als Stichprobenraum, die Mengen $A \in \mathcal{A}$ als Ereignisse, $P(A)$ als Wahrscheinlichkeit von A, die Elemente $\omega \in \Omega$ bzw. die 1-Punkt-Mengen $\{\omega\} \subset \Omega$ (nicht notwendig $\in \mathcal{A}$) als Elementarereignisse (auch Realisierungen).

Ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Ω endlich, $\mathcal{A} = \mathcal{P}(\Omega)$, $P(\{\omega\}) = \frac{1}{|\Omega|}$ $(\omega \in \Omega)$ ($|\Omega|$ = Anzahl der Elemente in Ω) heißt **Laplacescher Wahrscheinlichkeitsraum**:

$$\text{Für } A \in \mathcal{A} \text{ gilt } P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der “günstigen” Fälle}}{\text{Anzahl der “möglichen” Fälle}}$$

Es sei $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathcal{B})$ und $(p_k)_{k \in \mathbb{N}_0}$ ($\mathbb{N}_0 := \{0, 1, \ldots\}$) eine Zahlenfolge mit $p_k \geq 0$ $(k \in \mathbb{N}_0)$, $\sum\limits_{k=0}^{\infty} p_k = 1$. Dann ist $P : \mathcal{B} \to \mathbb{R}$ mit

$$P(A) := \sum_{k \in A \cap \mathbb{N}_0} p_k, \;\; A \in \mathcal{A},$$

ein W-Maß. (p_k) heißt **Zähldichte** zu P.

Ist speziell $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$, $(k \in \mathbb{N}_0)$ bei festem $\lambda > 0$, so heißt P **Poisson-Verteilung mit Parameter $\boldsymbol{\lambda}$** … Bezeichnung $\pi(\lambda)$.

1.15 Satz (1. Lemma von Borel und Cantelli).
Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$; $A_n \in \mathcal{A}$ $(n \in \mathbb{N})$. Mit

$$\begin{aligned} \overline{\lim}\, A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k &= \{\omega \in \Omega \mid \text{ existieren unendlich viele } n \in \mathbb{N} \text{ mit } \omega \in A_n\} \\ &= \text{“Ereignis, dass unendlich viele } A_n \text{ eintreten”} \end{aligned}$$

gilt dann:

$$\sum_{n=1}^{\infty} P(A_n) < \infty \quad \Longrightarrow \quad P(\overline{\lim}\, A_n) = 0, \quad \text{d.h. P-f.s. gilt: } A_n \text{ tritt nur endlich oft auf.}$$

Beispiel: Lernen aus Erfahrung

Student macht bei n-ter Aufgabe Fehler mit Wk. $\frac{1}{n^2} = P(A_n)$.

D. g.: P (“Student macht unendlich viele Fehler”) $= 0$,

da $\sum\limits_{n=1}^{\infty} P(A_n) = \sum\limits_{n=1}^{\infty} \frac{1}{n^2} < \infty$.

Beweis von Satz 1.15:

Für jedes $n \in \mathbb{N}$ gilt wegen $\overline{\lim}\, A_n \subset \bigcup\limits_{k=n}^{\infty} A_n$.

$P\left(\overline{\lim}\, A_n\right) \le P\left(\bigcup\limits_{k=n}^{\infty} A_n\right) \le \sum\limits_{k=n}^{\infty} P(A_k) \to 0 \ (n \to \infty)$, da $\sum\limits_{k=1}^{\infty} P(A_k) < \infty$. $\square$

1.16 Satz (Fortsetzungs- und Eindeutigkeitssatz). Sei P ein σ-endliches [d.h. es existieren $A_n \in \mathcal{A}$ mit $A_n \uparrow \Omega$ und $P(A_n) < \infty$ für alle n] Maß auf einer Algebra $\mathcal{A}$ in Ω. Dann existiert genau ein Maß P^* auf $\mathcal{F}(\mathcal{A})$, das P fortsetzt [d.h. $\underset{A \in \mathcal{A}}{\forall}$ $P^*(A) = P(A)$].

Beweis: Siehe Integrationstheorie bzw. Maßtheorie. $\square$

1.17 Bemerkung. P^* in Satz 1.16 ist gegeben durch

$$P^*(A) = \inf\{\sum_{n=1}^{\infty} P(B_n) \mid B_n \in \mathcal{A} \ \ (n = 1, 2, \ldots) \text{ mit } A \subset \bigcup_n B_n\}, \quad A \in \mathcal{F}(\mathcal{A})\,.$$

1.18 Bemerkung. Es ist im allgemeinen nicht möglich, P fortzusetzen von $\mathcal{A}$ auf $\mathcal{P}(\Omega)$.

1.19 Satz. Es gibt genau ein Maß λ auf $\mathcal{B}_n$, das jedem beschränkten (nach rechts halboffenen) Intervall in $\mathbb{R}^n$ seinen elementargeometrischen Inhalt zuordnet.

Beweis: Siehe Integrationstheorie bzw. Maßtheorie. $\square$

1.20 Definition. Das Maß λ in Satz 1.19 heißt **Lebesgue–Borel–Maß** (L-B-Maß) in $\mathbb{R}^n$.

1.21 Definition. Eine Funktion $F : \mathbb{R} \to \mathbb{R}$, die monoton wachsend (im Sinne von nicht fallend) und rechtsseitig stetig ist mit

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

heißt (eindimensionale) **Verteilungsfunktion** (VF).

1.22 Satz. Zu jedem W-Maß P auf $\mathcal{B}$ gibt es genau eine VF $F : \mathbb{R} \to \mathbb{R}$, so dass

$$(*) \qquad F(b) - F(a) = P((a,b]), \quad a \leq b, \quad a, b \in \mathbb{R}$$

gilt, und umgekehrt ... Bijektion $P \longleftrightarrow F$. Hierbei $F(x) = P((-\infty, x]), \quad x \in \mathbb{R}$.

Beweis. Siehe Integrationstheorie bzw. Maßtheorie. $\square$

Beispiel: $\Phi : \mathbb{R} \to [0,1]$ mit

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, \quad x \in \mathbb{R},$$

ist eine VF. Das zugehörige W-Maß auf $\mathcal{B}$ ist die sogenannte **standardisierte Normalverteilung** ... Bezeichnung $\boldsymbol{N(0,1)}$.
Verallgemeinerung: **Normalverteilung (Gauß-Verteilung)** $N(a, \sigma^2)$ ($a, \sigma \in \mathbb{R}$, $\sigma > 0$) mit durch

$$F(x) := \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt, \quad x \in \mathbb{R},$$

gegebener VF F.

1.23 Bemerkung. Definition 1.21 und Satz 1.22 lassen sich auf $\mathcal{B}_n$ und $\mathbb{R}^n$ als Definitionsbereiche von P bzw. F verallgemeinern statt $\mathcal{B}$ bzw. $\mathbb{R}$. Dabei ist die VF F zu einem W-Maß P auf $\mathcal{B}_n$ gegeben durch

$$F(x_1, \ldots, x_n) = P(\{(u_1, \ldots, u_n) \in \mathbb{R}^n \mid u_1 \leq x_1, \ldots, u_n \leq x_n\})$$

für $(x_1, \ldots, x_n) \in \mathbb{R}^n$.

1.24 Definition. Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $A \in \mathcal{A}$, $B \in \mathcal{A}$ mit $P(B) > 0$. Dann heißt

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

die **bedingte Wahrscheinlichkeit** von A unter der Bedingung B.

1.25 Satz. Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum.

a) Bei festem $B \in \mathcal{A}$ mit $P(B) > 0$ ist $P(\cdot \mid B)$ ein W-Maß auf $\mathcal{A}$, das auf B konzentriert ist [d.h. $P(B \mid B) = 1$].

b) Für $A, B \in \mathcal{A}$ mit $P(A) > 0$, $P(B) > 0$ gilt

$$P(A \mid B) = P(B \mid A) \cdot \frac{P(A)}{P(B)}\,.$$

c) Für $A_n \in \mathcal{A}$ $(n = 1, \ldots, m)$ mit $P(\bigcap\limits_{n=1}^{m-1} A_n) > 0$ gilt

$$P(\bigcap_{n=1}^{m} A_n) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2) \cdot \ldots \cdot P(A_m \mid \bigcap_{n=1}^{m-1} A_n)\,.$$

Beweis: Vgl. Einführung in die Stochastik. □

1.26 Satz. Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $\{B_n;\ n \in I\}$ eine höchstens abzählbare Familie paarweise disjunkter Ereignisse $B_n \in \mathcal{A}$ mit $P(B_n) > 0$ $(n \in I)$ und $\sum\limits_{n \in I} B_n = \Omega$. Dann gilt

a) $P(A) = \sum\limits_{n \in I} P(B_n) P(A \mid B_n)$ $(A \in \mathcal{A})$

(**Formel von der totalen Wahrscheinlichkeit**),

b) $P(B_k \mid A) = \dfrac{P(B_k) \cdot P(A \mid B_k)}{\sum\limits_{n \in I} P(B_n) P(A \mid B_n)}$, $k \in I$ $(A \in \mathcal{A}$ mit $P(A) > 0)$

(**Formel von Bayes**).

Beweis: Vgl. Einführung in die Stochastik. □

Bezeichnungen [Maßraum $(\Omega, \mathcal{A}, \mu)$, Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$]:
μ-fast überall (μ-f.ü.) in Ω bzw. P-fast sicher (P-f.s.) in Ω … überall bis auf eine Menge $\in \mathcal{A}$ vom μ-Maß bzw. P-Maß Null [oder eine Teilmenge hiervon].
L-fast überall (L-f.ü.) in $\mathbb{R}^n$ … überall bis auf eine Menge $\in \mathcal{B}_n$ vom L-B-Maß Null [oder eine Teilmenge hiervon].

2 Zufallsvariablen und Verteilungen

Beispiel: Werfen zweier echter Würfel.
Wie groß ist die Wahrscheinlichkeit, dass die Summe der Augenzahlen größer oder gleich 10 ist?
Stochastische Modellierung:

$$\Omega = \{(\omega_1, \omega_2) : \omega_1, \omega_2 \in \{1, \ldots, 6\}\}, \mathcal{A} = \mathcal{P}(\Omega), P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{36}.$$

Die gesuchte Wahrscheinlichkeit ist

$$P(\{(4,6), (5,5), (5,6), (6,4), (6,5), (6,6)\}) = \frac{6}{36} = \frac{1}{6}.$$

Statt der Werte der beiden Würfel betrachte die Summe der Augenzahlen

$$X((\omega_1, \omega_2)) = \omega_1 + \omega_2.$$

Wahrscheinlichkeitsraum für Zufallsexperiment mit Ergebnis $X((\omega_1, \omega_2))$:

$$\Omega' = \{2, 3, \ldots, 12\},\ \mathcal{A}' = \mathcal{P}(\Omega) \text{ und}$$

$$P_X : \mathcal{A}' \to \mathbb{R}$$

definiert durch

$$P_X(A') = P(\{\omega \in \Omega | X(\omega) \in A'\}) = P(X^{-1}(A'))$$

Also ist die gesuchte Wahrscheinlichkeit gleich:

$$\begin{aligned} P_X(\{10, 11, 12\}) &= P(X^{-1}(\{10, 11, 12\}) \\ &= P(\{(\omega_1, \omega_2) \in \Omega : \omega_1 + \omega_2 \in \{10, 11, 12\}\} \\ &= (\{(4,6), (5,5), (5,6), (6,4), (6,5), (6,6)\} \\ &= \tfrac{6}{36} = \tfrac{1}{6}. \end{aligned}$$

Allgemein: Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbf{P})$, σ-Algebra $\mathcal{A}'$ in $\Omega' \neq \emptyset$, Abbildung $X : \Omega \to \Omega'$. Wird ω gemäß $(\Omega, \mathcal{A}, \mathbf{P})$ zufällig erzeugt, so ist die Wahrscheinlichkeit, dass $X(\omega)$ in $A' \in \mathcal{A}'$ liegt, gegeben durch

$$P_X(A') := P(X^{-1}(A')).$$

Damit diese wohldefiniert ist für alle $A' \in \mathcal{A}'$, muss gelten:

$$X^{-1}(A') \in \mathcal{A} \quad (A' \in \mathcal{A}').$$

Abbildungen X mit dieser Eigenschaft heißen **Zufallsvariablen**.

2.1 Messbare Abbildungen und Zufallsvariablen

2.1 Definition. Zwei nichtleere Mengen Ω, Ω'; $A' \subset \Omega'$; Abbildung $X : \Omega \to \Omega'$. Die Menge

$$\{\omega \in \Omega \mid X(\omega) \in A'\} =: X^{-1}(A') =: [X \in A']$$

(in Ω) heißt das **Urbild** von A' bezüglich der Abbildung X; die somit definierte Abbildung $X^{-1} : \mathcal{P}(\Omega') \to \mathcal{P}(\Omega)$ heißt **Urbildfunktion** (zu unterscheiden von einer inversen Funktion!).

Bezeichnung. Sei $X : \Omega \to \Omega'$, $\mathcal{C}'$ Mengensystem in Ω'. $\underbrace{X^{-1}(\mathcal{C}')}_{\text{ein Mengensystem in } \Omega.} := \{X^{-1}(A') \mid A' \in \mathcal{C}'\}$.

2.2 Satz. Sei $X : \Omega \to \Omega'$.

a) X^{-1} und die Mengenoperationen $\cup$, $\sum$, $\cap$, ^{c}, $\setminus$ sind vertauschbar;
z.B. $X^{-1}(\bigcup_{\alpha\in I} A'_\alpha) = \bigcup_{\alpha\in I} X^{-1}(A'_\alpha)$ $\quad (A'_\alpha \subset \Omega', \quad \alpha \in$ Indexbereich $I)$.

b) $X^{-1}(\emptyset) = \emptyset$; $\quad X^{-1}(\Omega') = \Omega$; $\quad A' \subset B' \subset \Omega' \Longrightarrow X^{-1}(A') \subset X^{-1}(B')$.

c) $\mathcal{A}'$ σ-Algebra in $\Omega' \Longrightarrow X^{-1}(\mathcal{A}')$ σ-Algebra in Ω.

d) $\mathcal{C}' \subset \mathcal{P}(\Omega') \Longrightarrow X^{-1}(\mathcal{F}_{\Omega'}(\mathcal{C}')) = \mathcal{F}_\Omega(X^{-1}(\mathcal{C}'))$.

Beweis:

a) , b) elementar

c)
- $\emptyset \in X^{-1}(\mathcal{A}')$ da $\emptyset = X^{-1}(\emptyset)$ und $\emptyset \in \mathcal{A}'$
- Sei $A = X^{-1}(A') \in X^{-1}(\mathcal{A}')$ mit $A' \in \mathcal{A}'$.
  D. g.
  $A^c = \Omega \setminus X^{-1}(A') \overset{a)}{=} X^{-1}(\Omega' \setminus A') \in X^{-1}(\mathcal{A}')$
  da $\Omega' \setminus A' \in \mathcal{A}'$
- Sei $A_n = X^{-1}(A'_n) \in X^{-1}(\mathcal{A}')$ mit $A'_n \in \mathcal{A}'$ $(n \in \mathbb{N})$
  D. g.
  $$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} X^{-1}(A'_n) \overset{a)}{=} X^{-1}\left(\bigcup_{n=1}^{\infty} A'_n\right) \in X^{-1}(\mathcal{A}')$$
  da $\bigcup_{n=1}^{\infty} A'_n \in \mathcal{A}'$.

d) Zu zeigen:

$$X^{-1}(\mathcal{F}_{\Omega'}(\mathcal{C}')) = \mathcal{F}_\Omega(X^{-1}(\mathcal{C}'))$$

"$\supset$" Folgt aus $X^{-1}(\mathcal{C}') \subset X^{-1}(\mathcal{F}_{\Omega'}(\mathcal{C}'))$ und $X^{-1}(\mathcal{F}_{\Omega'}(\mathcal{C}'))$ σ-Algebra nach c).
"$\subset$" Man sieht leicht:

$$\overline{\mathcal{A}} := \{\overline{A} \subset \Omega' : X^{-1}(\overline{A}) \in \mathcal{F}_{\Omega}(X^{-1}(\mathcal{C}'))\}$$

ist σ-Algebra in Ω' mit $\mathcal{C}' \subset \overline{\mathcal{A}}$.
$\Rightarrow\ \mathcal{F}_{\Omega'}(\mathcal{C}') \subset \overline{\mathcal{A}} \Rightarrow X^{-1}(\mathcal{F}_{\Omega'}(\mathcal{C}')) \subset \mathcal{F}_{\Omega}(X^{-1}(\mathcal{C}'))$.

$\square$

2.3 Definition. Messräume $(\Omega, \mathcal{A})$, $(\Omega', \mathcal{A}')$. Die Abbildung $X : \Omega \to \Omega'$ heißt $\mathcal{A}$-$\mathcal{A}'$-messbar [kurz: **messbar**], wenn gilt:

$$\underset{A' \in \mathcal{A}'}{\forall} X^{-1}(A') \in \mathcal{A}, \quad \text{d.h. } X^{-1}(\mathcal{A}') \subset \mathcal{A},$$

d.h. Urbilder von messbaren Mengen in Ω' sind messbare Mengen in Ω.
In diesem Falle Schreibweise X: $(\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$.

2.4 Bemerkung. Nach Satz 2.2c ist $X^{-1}(\mathcal{A}')$ die kleinste der σ-Algebren $\mathcal{A}$ in Ω mit $\mathcal{A} - \mathcal{A}'$-Messbarkeit von X.

2.5 Definition. Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $(\Omega', \mathcal{A}')$ ein Messraum. Die Abbildung $X : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ heißt **Zufallsvariable** (ZV) auf $(\Omega, \mathcal{A}, P)$ [mit Zustandsraum Ω'] (ausführlich: $(\Omega', \mathcal{A}')$-ZV auf $(\Omega, \mathcal{A}, P)$).
$X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ heißt **reelle ZV** (oft auch nur ZV);
$X : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ heißt **erweitert-reelle ZV**;
$X : (\Omega, \mathcal{A}) \to (\mathbb{R}^n, \mathcal{B}_n)$ heißt n-**dimensionaler Zufallsvektor**.
$X(\omega)$ für ein $\omega \in \Omega$ heißt eine **Realisierung** der ZV X.
Bezeichnung $\overline{\mathcal{B}} := \{B,\ B \cup \{+\infty\},\ B \cup \{-\infty\},\ B \cup \{-\infty, +\infty\} \mid B \in \mathcal{B}\}$.

2.6 Satz. Messräume $(\Omega, \mathcal{A})$, $(\Omega', \mathcal{A}')$; $X : \Omega \to \Omega'$; $\mathcal{C}'$ Erzeugersystem von $\mathcal{A}'$. Dann gilt:

$$X \text{ messbar } \iff \underbrace{(\underset{C' \in \mathcal{C}'}{\forall} X^{-1}(C') \in \mathcal{A})}_{\text{d.h. } X^{-1}(\mathcal{C}') \subset \mathcal{A}}$$

Beweis: Nach Satz 2.2 d) gilt:

$$X^{-1}(\mathcal{C}') \subset \mathcal{A} \quad \Longrightarrow \quad X^{-1}(\mathcal{F}_{\Omega'}(\mathcal{C}')) = \mathcal{F}_{\Omega}(X^{-1}(\mathcal{C}')) \subset \mathcal{A},$$

da $\mathcal{A}$ σ-Algebra. $\square$

2.7 Satz. Messraum $(\Omega, \mathcal{A})$, $X : \Omega \to \overline{\mathbb{R}}$. Dann gilt

$$\begin{aligned} X \ \mathcal{A}\text{-}\overline{\mathcal{B}}\text{-messbar} &\iff \underset{\alpha\in\mathbb{R}}{\forall} [X \leq \alpha] \in \mathcal{A} \\ &\iff \underset{\alpha\in\mathbb{R}}{\forall} [X < \alpha] \in \mathcal{A} \\ &\iff \underset{\alpha\in\mathbb{R}}{\forall} [X \geq \alpha] \in \mathcal{A}. \end{aligned}$$

Hierbei:

$$[X \leq \alpha] = [X \in [-\infty, \alpha]] = X^{-1}([-\infty, \alpha]).$$

Beweis: Folgt aus Satz 2.6 unter Verwendung der Tatsache, dass z.B.

$$\{[-\infty, \alpha] \ : \ \alpha \in \mathbb{R}\}$$

ein Erzeuger von $\bar{\mathcal{B}}$ ist, vgl. Integrationstheorie bzw. Maßtheorie. □

2.8 Korollar. Für zwei Abbildungen $X, Y : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ gilt

$$\underbrace{[X < Y]}_{:= \{\omega \in \Omega \mid X(\omega) < Y(\omega)\}}, \quad [X \leq Y], \quad [X = Y], \quad [X \neq Y] \in \mathcal{A}.$$

Beweis:

- $[X < Y] = \bigcup\limits_{r\in\mathbb{Q}} [X < r] \cap [r < Y]$
- $[X \leq Y] = [Y < X]^c$
- $[X = Y] = [X \leq Y] \cap [Y \leq X]$
- $[X \neq Y] = [X = Y]^c$

□

2.9 Satz. Sei $\emptyset \neq A \subset \mathbb{R}^m$, $A \cap \mathcal{B}_m := \{A \cap B \mid B \in \mathcal{B}_m\}$.
Jede stetige Abbildung $f : A \to \mathbb{R}^n$ ist $A \cap \mathcal{B}_m$-$\mathcal{B}_n$-messbar.
Beweis: Sei $S_r(x)$ die offene Kugel um x mit Radius $r > 0$. Sei O eine beliebige offene Menge. Da f stetig ist und O mit jedem Punkt auch noch eine ganze Umgebung des Punktes enthält existiert für jedes $x \in f^{-1}(O)$ ein $\delta_x > 0$ mit

$$f(S_{\delta_x}(x) \cap A) \subset O.$$

Also gilt (da die Vereinigung offener Mengen immer selbst offen ist):

$$f^{-1}(O) = \bigcup_{x\in f^{-1}(O)} S_{\delta_x}(x) \cap A = A \cap \left(\bigcup_{x\in f^{-1}(O)} S_{\delta_x}(x) \right) \in A \cap \mathcal{B}_m.$$

Mit Satz 2.6 folgt die Behauptung. $\square$

2.10 Satz. $X : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$, $Y : (\Omega_2, \mathcal{A}_2) \to (\Omega_3, \mathcal{A}_3)$.
Die zusammengesetzte Abbildung $Y \circ X : \Omega_1 \to \Omega_3$ ist dann $\mathcal{A}_1$-$\mathcal{A}_3$-messbar.
Beweis: Klar, da $(Y \circ X)^{-1}(A) = X^{-1}(Y^{-1}(A))$. $\square$

2.11 Korollar. Für $X : (\Omega, \mathcal{A}) \to (\mathbb{R}^m, \mathcal{B}_m)$ und eine stetige Abbildung $g : \mathbb{R}^m \to \mathbb{R}^n$ ist $g \circ X : \Omega \to \mathbb{R}^n$ $\mathcal{A}$-$\mathcal{B}_n$-messbar.
Beweis: Folgt unmittelbar aus den Sätzen 2.9 und 2.10. $\square$

2.12 Satz. Messraum $(\Omega, \mathcal{A})$.

a) Sei $X_n : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}) \quad (n = 1, \ldots, m)$.
Dann ist $Y : \Omega \to \mathbb{R}^m$ mit $Y(\omega) := (X_1(\omega)), \ldots, X_m(\omega))$, $\omega \in \Omega$, $\mathcal{A}$-$\mathcal{B}_m$-messbar, und für $g : (\mathbb{R}^m, \mathcal{B}_m) \to (\mathbb{R}, \mathcal{B})$ ist $g \circ Y : \Omega \to \mathbb{R}$ $\mathcal{A}$-$\mathcal{B}$-messbar.

b) Seien $X_{1,2} : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$.
Dann sind $\alpha X_1 + \beta X_2 \quad (\alpha, \beta \in \mathbb{R})$, $X_1 X_2$, $\dfrac{X_1}{X_2}$ (falls existent) $\mathcal{A}$-$\mathcal{B}$-messbar.

Beweis: Siehe Integrationstheorie bzw. Maßtheorie. $\square$

2.13 Satz. Messraum $(\Omega, \mathcal{A})$; $X_n : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}}) \quad (n = 1, 2, \ldots)$.
Dann sind $\inf_n X_n$, $\sup_n X_n$, $\overline{\lim_n} X_n$, $\underline{\lim_n} X_n$ $\mathcal{A}$-$\overline{\mathcal{B}}$-messbar, ebenso $\lim_n X_n$ (falls existent).
Beweis: Siehe Integrationstheorie bzw. Maßtheorie. $\square$

2.2 Bildmaße und Verteilungen

2.14 Satz. Messräume $(\Omega, \mathcal{A})$, $(\Omega', \mathcal{A}')$; Abbildung $X : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$; Maß μ auf $\mathcal{A}$. Durch

$$\begin{aligned} \mu_X(A') &:= \mu(X^{-1}(A')) \\ &= \mu(\{\omega \in \Omega \mid X(\omega) \in A'\}) =: \mu[X \in A']; \quad A' \in \mathcal{A}', \end{aligned}$$

wird ein Maß (das sogenannte **Bildmaß**) μ_X auf $\mathcal{A}'$ definiert.
Ist μ ein W-Maß auf $\mathcal{A}$, dann ist μ_X ein W-Maß auf $\mathcal{A}'$.

Beweis: 0) μ_X ist wohldefiniert, da $X^{-1}(A') \in \mathcal{A}$ für $A' \in \mathcal{A}'$ gilt.

1. $\mu_X(\emptyset) = \mu(X^{-1}(\emptyset)) = \mu(\emptyset) = 0$

2. $\mu_X(A') = \mu(X^{-1}(A')) \geq 0$, da $\mu(A) \geq 0 \ \ (A \in \mathcal{A})$.

3. Sind $A'_1, A'_2, \ldots \in \mathcal{A}'$ paarweise disjunkt, so sind auch $X^{-1}(A'_n), X^{-1}(A'_2), \ldots \in \mathcal{A}$ paarweise disjunkt, und da μ Maß ist, gilt:

$$\begin{aligned}
\mu_X\left(\bigcup_{n=1}^{\infty} A'_n\right) &= \mu\left(X^{-1}\left(\bigcup_{n=1}^{\infty} A'_n\right)\right) && \text{(nach Definition)} \\
&= \mu\left(\bigcup_{n=1}^{\infty} X^{-1}(A'_n)\right) && \text{(nach Satz 2.2 a))} \\
&= \sum_{n=1}^{\infty} \mu(X^{-1}(A'_n)) && \text{(da } \mu \ \sigma\text{-additiv)} \\
&= \sum_{n=1}^{\infty} \mu_X(A'_n) && \text{(nach Definition)} .
\end{aligned}$$

Ist μ Wahrscheinlichkeitmaß, so gilt weiter:

$$\mu_X(\Omega') = \mu(X^{-1}(\Omega')) = \mu(\Omega) = 1$$

$\square$

2.15 Definition. Sei X eine $(\Omega', \mathcal{A}')$-ZV auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Das W-Maß P_X im Bild-Wahrscheinlichkeitsraum $(\Omega', \mathcal{A}', P_X)$ heißt **Verteilung** der ZV X.
Sprechweise: Die ZV X liegt in $A' \in \mathcal{A}'$ bzw. nimmt Werte in $A' \in \mathcal{A}'$ an mit Wahrscheinlichkeit $P_X(A') = P[X \in A']$.
$P[X \in A'] = 1 \ldots P$-fast sicher (P-f.s.) liegt X in A'.

2.16 Bemerkung. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, Messräume $(\Omega', \mathcal{A}')$, $(\Omega'', \mathcal{A}'')$; $X : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$, $Y : (\Omega', \mathcal{A}') \to (\Omega'', \mathcal{A}'')$. Dann gilt

$$P_{Y \circ X} = (P_X)_Y \, .$$

Begründung:

$$\begin{aligned}
P_{Y \circ X}(A'') &= P((Y \circ X)^{-1}(A'')) && \text{(nach Definition)} \\
&= P(X^{-1}(Y^{-1}(A''))) \\
&= P_X(Y^{-1}(A'')) && \text{(nach Definition)} \\
&= (P_X)_Y(A'') && \text{(nach Definition)}
\end{aligned}$$

$\square$

2.17 Definition. Messräume $(\Omega_i, \mathcal{A}_i)$ $(i = 1, \ldots, n)$.
Die **Produkt-$\boldsymbol{\sigma}$-Algebra** $\bigotimes_{i=1}^{n} \mathcal{A}_i$ wird definiert als die von dem Mengensystem $\left\{ \prod_{i=1}^{n} A_i \mid A_i \in \mathcal{A}_i \ (i = 1, \ldots, n) \right\}$ in

$$\prod_{i=1}^{n} \Omega_i := \{(\omega_1, \ldots, \omega_n) \, : \, \omega_i \in \Omega_i \ (i = 1, \ldots, n)\}$$

erzeugte σ-Algebra.
$\bigotimes_{i=1}^{n}(\Omega_i, \mathcal{A}_i) := \left(\prod_{i=1}^{n}\Omega_i,\ \bigotimes_{i=1}^{n}\mathcal{A}_i\right)$ heißt **Produkt-Messraum**.

Beispiel: Mit Identifizierung von $\mathbb{R}^n \times \mathbb{R}^m$ und $\mathbb{R}^{n+m}$ gilt

$$\mathcal{B}_n \otimes \mathcal{B}_m = \mathcal{B}_{n+m},$$

vgl. Integrationstheorie bzw. Maßtheorie.

2.18 Bemerkung. Abbildung $X_i : (\Omega, \mathcal{A}) \to (\Omega_i, \mathcal{A}_i) \quad (i = 1, \ldots, n)$. Die Abbildung $X : \Omega \to \prod_{i=1}^{n}\Omega_i$ mit

$$X(\omega) := (X_1(\omega), \ldots, X_n(\omega)),\ \omega \in \Omega$$

ist $\mathcal{A}$-$\bigotimes_{i=1}^{n}\mathcal{A}_i$-messbar.

Begründung: Behauptung folgt aus Satz 2.6 unter Beachtung von

$$X^{-1}\left(\prod_{i=1}^{n} A_n\right) = \bigcap_{i=1}^{n} X_i^{-1}(A_i) \in \mathcal{A}$$

für alle $A_i \in \mathcal{A}_i$ □

2.19 Bemerkung. Messräume $(\Omega_i, \mathcal{A}_i)$ $(i = 1, \ldots, n)$. Ein W-Maß auf $\bigotimes_{i=1}^{n}\mathcal{A}_i$ ist eindeutig festgelegt durch seine Werte auf dem Mengensystem $\left\{\prod_{i=1}^{n} A_i \mid A_i \in \mathcal{A}_i\ (i = 1, \ldots, n)\right\}$.

Begründung: Folgt aus allgemeiner Form der Fortsetzungssätze aus der Integrationstheorie bzw. Maßtheorie unter Beachtung der Tatsache, daß das Wahrscheinlichkeitsmaß als endliches Maß auf $\{\prod_{i=1}^{n} A_i : A_i \in \mathcal{A}_i\}$ immer σ-endlich ist. □

2.20 Definition. Seien X_i $(\Omega_i, \mathcal{A}_i)$-ZVn auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ $(i = 1, \ldots, n)$. Die Verteilung P_X der ZV $X := (X_1, \ldots, X_n)$ auf $(\Omega, \mathcal{A}, P)$ — erklärt durch

$$P_X(A) := P\left[(X_1, \ldots, X_n) \in A\right]$$

für $A \in \bigotimes_{i=1}^{n}\mathcal{A}_i$ oder auch nur für $A = \prod_{i=1}^{n} A_i$ $(A_i \in \mathcal{A}_i,\ i = 1, \ldots, n)$ — heißt die **gemeinsame Verteilung** der ZVn X_i. Die Verteilungen P_{X_i} — erklärt durch

$$P_{X_i}(A_i) := P[X_i \in A_i] = P[X \in \Omega_1 \times \ldots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \ldots \times \Omega_n]$$

für $A_i \in \mathcal{A}_i$ — heißen **Randverteilungen** von $P_X \quad (i = 1, \ldots, n)$.

Wichtiger Spezialfall in Definition 2.20: $(\Omega_i, \mathcal{A}_i) = (\mathbb{R}, \mathcal{B}) \quad (i = 1, \dots, n)$.

2.21 Bemerkung. Bezeichnungen wie in Definition 2.20.

a) Die Verteilung P_X ist ohne Zusatzvoraussetzung durch ihre Randverteilungen **nicht** eindeutig festgelegt.

b) Ist $\pi_i : \prod_{k=1}^{n} \Omega_k \to \Omega_i$ die (messbare!) Projektionsabbildung

$$(\omega_1, \dots, \omega_n) \to \omega_i,$$

so gilt $P_{X_i} = (P_X)_{\pi_i}$ $(i = 1, \dots, n)$.

Begründung:

a) Siehe Übungen

b) $(P_X)_{\pi_i} = P_{\pi_i \circ X} = P_{X_i}$ nach Bemerkung 2.16. □

2.22 Bemerkung.

a) Sei Q ein W-Maß auf $(\mathbb{R}^n, \mathcal{B}_n)$. Dann existieren reelle ZVn $X_1, \dots, X_n$ auf einem geeigneten Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ so, dass ihre gemeinsame Verteilung mit Q übereinstimmt:
$(\Omega, \mathcal{A}, P) := (\mathbb{R}^n, \mathcal{B}_n, Q)$, $X_i : \Omega \to \mathbb{R}$ wird definiert als die Projektionsabbildung

$$(x_1, \dots, x_n) \to x_i \quad (i = 1, \dots, n);$$

hierbei ist also $X := (X_1, \dots, X_n)$ die auf $\mathbb{R}^n$ definierte identische Abbildung.

b) Möglichkeit der unmittelbaren Verallgemeinerung von a) auf einen Produkt-Meßraum $\left(\prod_{i=1}^{n} \Omega_i, \bigotimes_{i=1}^{n} \mathcal{A}_i\right)$ statt $(\mathbb{R}^n, \mathcal{B}_n)$.

c) Sonderfall zu b): Ist $(\Omega, \mathcal{A}, Q)$ ein Wahrscheinlichkeitsraum, $X : \Omega \to \Omega$ die identische Abbildung, so gilt $P_X = Q$. **Jedes W-Maß lässt sich somit als eine Verteilung auffassen (und** — definitionsgemäß — **umgekehrt)**.

Begründung von a):
Sei $P = Q$ und $X_i = \pi_i$ mit π_i wie in Bemerkung 2.21 b). Dann gilt

$$\begin{aligned} & X : \Omega \to \Omega \\ & X((\omega_1, \dots, \omega_n)) = (\pi_1(\omega_1, \dots, \omega_n), \dots, \pi_n(\omega_1, \dots, \omega_n)) \\ & \qquad = (\omega_1, \dots, \omega_n) \end{aligned}$$

und damit

$$P_X(A) = Q(X^{-1}(A)) = Q(A).$$

□

3 Erwartungswerte und Dichten

Beispiel: Glücksrad mit 64 Feldern, davon 1 rotes, 8 blaue und 55 weiße.
Als Gewinn erhält man bei

- rotem Feld: EURO 8
- blauem Feld: EURO 4
- weißem Feld: EURO 0

Wie groß ist der Gewinn bei Drehen an diesem Glücksrad "im Mittel"?
Glücksrad wird N-mal gedreht, N_r bzw. N_b bzw. N_w sei die Anzahl des Auftretens eines roten bzw. blauen bzw. weißen Feldes.
Ausgezahlter mittlerer Gewinn:

$$\frac{8 \cdot N_r + 4 \cdot N_b + 0 \cdot N_w}{N} = 8 \cdot \frac{N_r}{N} + 4 \cdot \frac{N_b}{N} + 0 \cdot \frac{N_w}{N}.$$

Für N "groß" strebt das ganze gegen:

$$8 \cdot \frac{1}{64} + 4 \cdot \frac{8}{64} + 0 \cdot \frac{55}{64} = \frac{40}{64} = \frac{5}{8}$$

Verallgemeinerung: X sei reelle ZV, die die Werte $x_1, \ldots, x_k$ mit den Wahrscheinlichkeiten $p_1, \ldots, p_K \geq 0$ annimmt, weiter $p_1 + \ldots + p_K = 1$.
Mittlerer Wert von X ist

$$EX = \sum_{k=1}^{N} x_k \cdot p_k = \sum_{k=1}^{N} x_k \cdot P[X = x_k].$$

Hierbei werden die auftretenden Werte mit den Wahrscheinlichkeiten gewichtet gemittelt.
Beispiel: (Kontinuierliche Verallgemeinerung des obigen Beispiels)
X sei normalverteilte Zufallsvariable, d. h. P_X hat die Verteilungsfunktion

$$F(x) = \int\limits_{-\infty}^{x} \underbrace{\frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(t-a)^2}{2\sigma^2}}}_{\uparrow} \, dt$$

sogenannte Dichtefunktion,
Wert entspricht den Wk.
$p_1, \ldots, p_K$ von oben.

Dann setzen wird

$$\begin{aligned}
EX &= \int_{-\infty}^{\infty} t \cdot \tfrac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(t-a)^2}{2\sigma^2}} dt \\
&= \int\limits_{-\infty}^{\infty} (t-a) \cdot \tfrac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(t-a)^2}{2\sigma^2}} dt + a \cdot \int\limits_{-\infty}^{\infty} \tfrac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(t-a)^2}{2\sigma^2}} dt \\
&= 0 + a \cdot 1 = a,
\end{aligned}$$

wobei beim dritten Gleichheitszeichen verwendet wurde, dass der erste Integrand punktsymmetrisch bzgl. $t = a$, ist, und das zweite Integral Eins ergibt (wegen $F(x) \to 1$ $(x \to \infty)$).

Allgemeiner

Gegeben seien ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und eine reelle ZV X auf $(\Omega, \mathcal{A}, P)$ mit Verteilung P_X (auf $\mathcal{B}$) und VF $F : \mathbb{R} \to \mathbb{R}$.

Bezeichnungen:

$$X^+(\omega) := \begin{cases} X(\omega), & \text{falls } X(\omega) > 0 \\ 0 & \text{sonst} \end{cases} \qquad X^-(\omega) := \begin{cases} -X(\omega), & \text{falls } X(\omega) < 0 \\ 0 & \text{sonst.} \end{cases}$$

Der Erwartungswert EX der ZV X gibt den "mittleren Wert" von X bezüglich P an, die Varianz $V(X)$ gibt die "Schwankung" der ZV X als "mittleren Wert" ihrer quadratischen Abweichung von EX an.

3.1 Definition. Sei X reelle ZV auf W-Raum $(\Omega, \mathcal{A}, P)$.
Dann heißt

$$EX := \int_\Omega X(\omega) P(d\omega) := \int_\Omega X dP$$

Erwartungswert (kurz: EW) von X.
Hierbei wird das obige Maßintegral wie folgt definiert:
0. Schritt: Sei $X \geq 0$, und X nehme nur endlich viele Werte an. Dann existiert Darstellung $X = \sum\limits_{i=1}^{N} \alpha_i \chi_{A_i}$ mit $\alpha_i \in \mathbb{R}$, $A_i \in \mathcal{A}$ $(i = 1, \dots, N)$, paarweise disjunkt, $\sum\limits_{i=1}^{N} A_i = \Omega$, und wir setzen:

$$EX := \int_\Omega X\, dP := \sum_{i=1}^{N} \alpha_i P(A_i)$$

1. Schritt: Sei $X \geq 0$. Dann existieren ZVn $X_n \geq 0$, die jeweils nur endlich viele Werte annehmen mit $X_n(\omega) \uparrow X(\omega)$ $(n \to \infty)$, $\omega \in \Omega$ (!), und wir setzen:

$$EX := \int_\Omega X\, dP := \lim_{n \to \infty} EX_n$$

2. Schritt: Seien EX^+, EX^- nicht beide ∞. Dann setzen wir:

$$EX := \int_\Omega X\, dP := EX^+ - EX^-$$

3.2 Bemerkung. (zu Definition 3.1)

a) Die einzelnen Schritte sind sinnvoll (vgl. Integrationstheorie bzw. Maßtheorie).

b) Der Begriff $\int_\Omega X dP$ in Definition 3.1 lässt sich unmittelbar verallgemeinern auf den Fall eines Maßraumes $(\Omega, \mathcal{A}, \mu)$ und liefert $\int_\Omega X(\omega)\, \mu(d\omega) =: \int_\Omega X\, d\mu$.

   Wichtiger Spezialfall: $(\Omega, \mathcal{A}) = (\mathbb{R}^n, \mathcal{B}_n)$ oder etwas allgemeiner $\Omega = B$ mit $B \in \mathcal{B}_n$, $\mathcal{A}$ = σ-Algebra der Borelschen Mengen in B; $\mu = \lambda :=$ (Restriktion des) L-B-Maß(es) auf $\mathcal{A}$. $\int_{\mathbb{R}^n} X\, d\lambda$ bzw. $\int_B X\, d\lambda$ wird als **Lebesgue–Integral** bezeichnet. Im Falle $n = 1$ Schreibweise $\int_\mathbb{R} X(x)\, dx$ bzw. $\int_B X(x)\, dx$.

c) Sei X eine erweitert-reelle ZV auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit

$$P[|X| = \infty] = 0; \;\; Y(\omega) := \begin{cases} X(\omega), & \text{falls } |X(\omega)| < \infty, \;\; \omega \in \Omega, \\ 0 & \text{sonst.} \end{cases}$$

   Man definiert $EX := EY$, falls EY existiert.

3.3 Bemerkung. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Für $A \in \mathcal{A}$ gilt $P(A) = E\chi_A$, da $\mathcal{X}_A = 1 \cdot \mathcal{X}_A + 0 \cdot \mathcal{X}_{A^c}$ und damit nach Schritt 0 der Definition des Integrals gilt

$$E\mathcal{X}_A = 1 \cdot P(A) + 0 \cdot P(A^c) = P(A).$$

3.4 Definition. Existiert für die reelle ZV X auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ ein endlicher Erwartungswert EX, so heißt X integrierbar (bezüglich P). Analog für $\int_\Omega X\, d\mu$ in Bemerkung 3.2b.

3.5 Lemma. Für eine reelle ZV $X \geq 0$ mit VF F gilt

$$EX = \int_0^\infty (1 - F(x))\, dx\, .$$

Beweis: Nach dem Satz von Fubini (vgl. Integrationstheorie bzw. Maßtheorie) gilt:

$$\int_0^\infty (1 - F(t))dt = \int_0^\infty P[X > t]dt = \int_0^\infty \int_\Omega \mathcal{X}_{(t,\infty)}(X)dPdt$$

$$= \int_\Omega \int_0^\infty \mathcal{X}_{(t,\infty)}(X)dtdP = \int_\Omega XdP = EX,$$

wobei die zweitletzte Gleichheit aus

$$\mathcal{X}_{(t,\infty)}(X) = \begin{cases} 1 & \text{für} \quad t \leq X \\ 0 & \text{für} \quad t > X \end{cases}$$

folgt. $\square$

3.6 Satz. Sei X eine reelle ZV auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$.

a) X integrierbar $\Longleftrightarrow X^+$ und X^- integrierbar $\Longleftrightarrow |X|$ integrierbar,

b) existiert reelle integrierbare ZV $Y \geq 0$ mit $|X| \leq Y$ P-f.s. $\Longrightarrow X$ integrierbar,

c) X integrierbar, $\underbrace{Y}_{\text{reelle ZV}} = X$ P-f.s. $\Longrightarrow$ existiert $EY = EX$.

Beweis: Siehe Integrationstheorie bzw. Maßtheorie. $\square$

3.7 Satz. Seien X, Y reelle integrierbare ZVn auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$.

a) Es existiert $E(X+Y) = EX + EY$.

b) Es existiert $E(\alpha X) = \alpha EX$ $(\alpha \in \mathbb{R})$.

c) $X \geq Y \Longrightarrow EX \geq EY$.

d) $|EX| \leq E|X|$.

e) $X = Y$ P-f.s. $\Longrightarrow EX = EY$.

f) $X \geq 0, \quad EX = 0 \Longrightarrow X = 0$ P-f.s.

Beweis:

a) – d) Siehe Integrationstheorie bzw. Maßtheorie.

e)

$$\begin{aligned} EX &= E(X \cdot \mathcal{X}_{[X=Y]}) + E(X \cdot \mathcal{X}_{[X \neq Y]}) \\ &= E(X \cdot \mathcal{X}_{[X=Y]}) + 0 \quad \text{(nach Satz 3.6c), da } X \cdot \mathcal{X}_{[X \neq Y]} = 0 \text{ P-f.s.)} \\ &= E(Y \cdot \mathcal{X}_{[X=Y]}) \\ &= EY \qquad \text{(analog oben).} \end{aligned}$$

f) Sei $X \geq 0$ und $EX = 0$.
Setze $A = [X > 0]$ und $A_n = [X \geq \frac{1}{n}]$.

Dann gilt: $A_n \uparrow A$
und daher

$$\begin{aligned} P[X > 0] &= P(A) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} P[X \geq \tfrac{1}{n}] \\ &\leq \lim_{n\to\infty} E(n \cdot X) \quad (\text{da } \mathcal{X}_{[X \geq \frac{1}{n}]} \leq n \cdot X) \\ &= \lim_{n\to\infty} n \cdot EX = \lim_{n\to\infty} n \cdot 0 = 0. \end{aligned}$$

$\rightsquigarrow X = 0$ f.s. $\square$

3.8 Satz (Satz von der monotonen Konvergenz; B. Levi).
Für reelle ZVn X_n auf $(\Omega, \mathcal{A}, P)$ mit $X_n \geq 0 \quad (n \in \mathbb{N})$ und $X_n \uparrow X \quad (n \to \infty)$ existiert EX und es gilt

$$EX = \lim_n EX_n.$$

Hierbei $EX := \infty$, falls $P[X = \infty] > 0$. – Analog für $\int_\Omega X_{(n)} d\mu$ gemäß Bemerkung 3.2b.
Beweis: Siehe Integrationstheorie bzw. Maßtheorie. $\square$

Deutung: Nochmaliger Grenzübergang führt nicht zu neuem Integralbegriff.

Spezialfall des Satzes 3.8:
Nichtnegative ZVen X_n. D. g.:

$$E\left(\sum_{n=1}^{\infty} X_n\right) \overset{Satz\,3.8}{=} \lim_{N\to\infty} E\left(\sum_{n=1}^{N} X_n\right) \overset{Satz\,3.7}{=} \sum_{n=1}^{\infty} EX_n.$$

Weiterer wichtiger Grenzwertsatz:
3.9 Satz (Satz von der majorisierten Konvergenz).
Reelle Zven X_n, X und Y mit $X_n(\omega) \to X(\omega)$ für alle $\omega \in \Omega$ (oder $X_n(\omega) \to X(\omega)$ für P-f.a. $\omega \in \Omega$), $|X_n(\omega)| \leq Y(\omega)$ für alle $\omega \in \Omega$ und $EY < \infty$.
Dann existiert EX und es gilt

$$EX = \lim_{n\to\infty} EX_n.$$

Satz gilt analog auch für Maßintegrale.
Beweis: Siehe Integrationstheorie bzw. Maßtheorie. $\square$

3.10 Definition. Sei X ein n-dimensionaler Zufallsvektor mit Verteilung P_X auf $\mathcal{B}_n$ und VF $F : \mathbb{R}^n \to \mathbb{R}$.

a) Nimmt (eventuell nach Vernachlässigung einer P-Nullmenge in Ω) X höchstens abzählbar viele Werte an, so ist P_X eine sogenannte **diskrete W-Verteilung**.
Ist X eine reelle ZV und P_X auf $\mathbb{N}_0$ konzentriert (d.h. $P_X(\mathbb{N}_0) = 1$), so heißt die Folge (p_k) mit $p_k := P_X(\{k\}) = P[X = k]$, $k \in \mathbb{N}_0$, (wobei $\sum p_k = 1$) **Zähldichte** von X bzw. P_X.

b) Jede Funktion $f : (\mathbb{R}^n, \mathcal{B}_n) \to (\mathbb{R}, \mathcal{B})$ mit

$$f(x) \geq 0 \ (x \in \mathbb{R}^n) \text{ und } \int_{\mathbb{R}^n} f(x)dx = 1$$

heißt **Dichte**.
Gilt

$$P_X(B) = \int_B f(x)dx := \int f(x) \cdot \mathcal{X}_B(x)dx \quad (B \in \mathcal{B}_n),$$

so heißt f **Dichte von** P_X **bzw.** X **bzw.** F. In diesem Fall heißen P_X und F **totalstetig**.

3.11 Bemerkung.

a) F totalstetig $\Rightarrow$ F stetig.

Begründung: *oBdA* $n = 1$. Sei f Dichte von P_X.
Dann gilt für $(x_l)_l$ mit $x_l \to x$ $(l \to \infty)$:

$$f(t) \cdot \mathcal{X}_{(-\infty,x_l]}(t) \to f(t) \cdot \mathcal{X}_{(-\infty,x]}(t)$$

für L-f.a. $t \in \mathbb{R}$ (sogar für alle $t \in \mathbb{R} \setminus \{x\}$), und $f(t) \cdot \mathcal{X}_{(-\infty,x_l]}(t)$ ist betragsmäßig durch die integrierbare Majorante f beschränkt.

Nach dem Satz von der majorisierten Konvergenz gilt dann

$$\begin{aligned}
\lim_{l\to\infty} F(x_l) &= \lim_{l\to\infty} P_X((-\infty, x_l]) \\
&= \lim_{l\to\infty} \int_{(-\infty,x_l]} f(t)dt \\
&= \lim_{l\to\infty} \int f(t) \cdot \mathcal{X}_{(-\infty,x_l]}(t)dt \\
&= \int f(t) \cdot \mathcal{X}_{(-\infty,x]}(t)dt \qquad \text{(majorisierte Konvergenz)} \\
&= P_X((-\infty, x]) \\
&= F(x).
\end{aligned}$$

□

b) Besitzt die n-dimensionale VF $F : \mathbb{R}^n \to \mathbb{R}$ eine Dichtefunktion $f : \mathbb{R}^n \to \mathbb{R}_+$, so existiert L-f.ü. $\dfrac{\partial^n F}{\partial x_1 \ldots \partial x_n}$, und L-f.ü. — insbesondere an den Stetigkeitsstellen von f — gilt $\dfrac{\partial^n F}{\partial x_1 \ldots \partial x_n} = f$.

Begründung für $n = 1$**:**

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{h} \int_x^{x+h} f(t)dt \to f(x) \quad (h \to 0)$$

für L-f.a. x (nach dem Dichtetheorem von Lebesgue). Trivialerweise gilt die obige Grenzwertbeziehung in allen Stetigkeitspunkten von f. □

c) f Dichte von $P_X \iff \forall x \in \mathbb{R}^n :\ F(x) = \int_{(-\infty,x]} f(t)dt$.

Begründung: "$\Rightarrow$" folgt durch Wahl von $B = (-\infty, x]$ in Def. 3.10 b)
"$\Leftarrow$" Nach Voraussetzung gilt

$$P_X((-\infty, x]) = \int_{(-\infty,x]} f(t)dt$$

für alle $x \in \mathbb{R}^n$. Da $\{(-\infty, x] : x \in \mathbb{R}^n\}$ ein Erzeugersystem für $\mathcal{B}_n$ ist, folgt die Behauptung mit dem Fortsetzungssatz für Maße. $\square$

3.12 Satz. Sei X eine reelle ZV mit Verteilung P_X.

a) Ist P_X auf $\mathbb{N}_0$ konzentriert mit Zähldichte (p_k), dann gilt

$$EX = \sum_k k \cdot p_k \, .$$

b) Besitzt X eine Dichtefunktion $f : \mathbb{R} \to \mathbb{R}_+$, so existiert das Lebesgue-Integral $\int_\mathbb{R} xf(x)\,dx =: \int_\mathbb{R} xf(x)\,\lambda(dx)$ [λ L-B-Maß auf $\mathcal{B}$] genau dann, wenn EX existiert, und es gilt hierbei

$$EX = \int_\mathbb{R} xf(x)\,dx \, .$$

Beweis: Folgt aus Satz 3.13 unten. $\square$

Beispiel zu Satz 3.12. ZV X auf $(\Omega, \mathcal{A}, P)$.

a) X sei $b(n,p)$-verteilt, d.h. $P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$ $(k = 0, 1, \ldots, n)$ mit $n \in \mathbb{N}$, $p \in [0,1]$. Dann gilt

$$EX = \sum_{k=0}^{n} k \cdot \binom{n}{k} \cdot p^k (1-p)^{n-k}.$$

Nach dem binomischen Lehrsatz gilt für die sogenannte erzeugende Funktion g von X:

$$g(s) = \sum_{k=0}^{n} \binom{n}{k} \cdot p^k (1-p)^{n-k} \cdot s^k = (ps + (1-p))^n.$$

Ableiten aller drei Ausdrücke liefert

$$g'(s) = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} \cdot s^{k-1} = n \cdot (ps + (1-p))^{n-1} \cdot p,$$

woraus folgt

$$\underline{EX} = g'(1) = n \cdot (p \cdot 1 + (1-p))^{n-1} \cdot p = \underline{n \cdot p.}$$

b) X sei $\pi(\lambda)$-verteilt, d.h. $P[X = k] = e^{-\lambda} \dfrac{\lambda^k}{k!}$ $(k \in \mathbb{N}_0)$, mit $\lambda > 0$. Dann gilt

$$\begin{aligned} EX &= \sum_{k=0}^{\infty} k \cdot \tfrac{\lambda^k}{k!} \cdot e^{-\lambda}. \\ &= \lambda \cdot e^{-\lambda} \cdot \sum_{k=1}^{\infty} \tfrac{\lambda^{k-1}}{k-1}! \\ &= \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda. \end{aligned}$$

c) X sei $N(a,\sigma^2)$-verteilt, d.h. X habe Dichtefunktion $f : \mathbb{R} \to \mathbb{R}_+$ mit

$$f(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} \quad x \in \mathbb{R}\,,$$

mit $a \in \mathbb{R}$, $\sigma > 0$. Dann gilt $EX = a$ (vgl. Einführung in die Stochastik).

3.13 Satz. X sei eine reelle ZV mit Verteilung P_X und VF F; $g : (\mathbb{R},\mathcal{B}) \to (\mathbb{R},\mathcal{B})$. $E(g \circ X)$ existiert genau dann, wenn $\int_{\mathbb{R}} g\, dP_X$ existiert, und es gilt hierbei nach dem Transformationssatz für Integrale

$$E(g \circ X) = \int_{\mathbb{R}} g\, dP_X =: \int_{\mathbb{R}} g\, dF =: \int_{\mathbb{R}} g(x) \underbrace{dF(x)}_{F(dx)},$$

sowie

$$E(g \circ X) = \begin{cases} \sum\limits_k g(k) p_k & \text{unter der Voraussetzung von Satz 3.12a.} \\ \underbrace{\int\limits_{\mathbb{R}} g(x) f(x)\, dx}_{\text{Lebesgue-Integral}} & \text{unter der Voraussetzung von Satz 3.12b.} \end{cases}$$

Insbesondere gilt (mit $g = \chi_A$, wobei $A \in \mathcal{B}$)

$$P[X \in A] = P_X(A) = \begin{cases} \sum\limits_{k \in A \cap \mathbb{N}_0} p_k & \text{unter Vorauss. von Satz 3.12a} \\ \underbrace{\int_{\mathbb{R}} \chi_A(x) f(x)\, dx}_{\text{Lebesgue-Integral}} =: \int_A f(x)\, dx & \text{unter Vorauss. von Satz 3.12b} \end{cases}$$

Bemerkung: Satz 3.12 folgt aus Satz 3.13 mit $g(x) = x$.

Beweis von Satz 3.13:

a) **Fall 0:** $g = \mathcal{X}_A$ für ein $A \in \mathcal{B}_n$.
D. g.

$$\begin{aligned} E\mathcal{X}_A \circ X &= E\mathcal{X}_{X^{-1}(A)} \overset{Def.Integral}{=} P(X^{-1}(A)) \\ &\overset{Def.P_X}{=} P_X(A) \overset{Def.Integral}{=} \int \mathcal{X}_A dP_X. \end{aligned}$$

Fall 1: g nicht-negativ einfach.
Folgt mit Linearität des Erwartungswertes aus Fall 0.

Fall 2: g nichtnegativ.
Wähle nichtnegative einfache g_n mit $g_n \uparrow g$.

Dann sind $g_n \circ X$ nichtnegativ einfach mit $g_n \circ X \uparrow g \circ X$, und mit der Definition des Integrals folgt:

$$Eg \circ X \overset{Def.}{=} \lim_{n\to\infty} Eg_n \circ X \overset{Fall\ 1}{=} \lim_{n\to\infty} \int g_n dP_X \overset{Def.}{=} \int g dP_X.$$

Fall 3: g messbar

Dann gilt $g = g^+ - g^-$ und $(g \circ X)^+ = g^+ \circ X, (g \circ X)^- = g^- \circ X$, woraus folgt

$$\begin{aligned} E(g \circ X) &\overset{Def.}{=} E(g \circ X)^+ - E(g \circ X)^- \\ &\overset{s.o.}{=} E(g^+ \circ X) - E(g^- \circ X) \\ &\overset{Fall\ 2}{=} \int g^+ dP_X - \int g^- dP_X \\ &\overset{Def.}{=} \int g dP_X \end{aligned}$$

b) Nach dem Analogon zu Satz 3.7e) für Maßintegrale und nach dem Satz von der monotonen Konvergenz gilt

$$\begin{aligned} \int g^+ dP_X &= \int g^+ \mathcal{X}_{\mathbb{N}_0} dP_X \\ &\quad (\text{da } g^+ = g^+ \cdot \mathcal{X}_{\mathbb{N}_0}\ P_X\text{-f.ü.}) \\ &= \int \sum_{k=0}^{\infty} g^+(k) \cdot \mathcal{X}_{\{k\}} dP_X \\ &= \sum_{k=0}^{\infty} \int g^+(k) \cdot \mathcal{X}_{\{k\}} dP_X \\ &= \sum_{k=0}^{\infty} g^+(k) \cdot \int \mathcal{X}_{\{k\}}(x) dP_X(x) \\ &= \sum_{k=0}^{\infty} g^+(k) \cdot P[X = k] \end{aligned}$$

und analog

$$\int g^- dP_X = \sum_{k=0}^{\infty} g^-(k) \cdot P[X = k]$$

woraus folgt

$$\begin{aligned} \int g\, dP_X &= \int g^+ dP_X - \int g^- dP_X = \sum_{k=0}^{\infty} (g^+(k) - g^-(k)) \cdot P[X = k] \\ &= \sum_{k=0}^{\infty} g(k) \cdot P[X = k]. \end{aligned}$$

c) Für $g = \mathcal{X}_A$ gilt nach Definition der Dichte:

$$\begin{aligned} \int g dP_X &= \int \mathcal{X}_A dP_X = P_X(A) \\ &= \int_A f(x) dx = \int \mathcal{X}_A(x) \cdot f(x) dx \\ &= \int g(x) \cdot f(x) dx. \end{aligned}$$

Analog zu a) folgt daraus die Behauptung im allgemeinen Fall (wobei in Fall 2 der Satz von der monotonen Konvergenz angewendet wird). □

Beispiel Ein Punkt P wird rein zufällig aus der Einheitskreisscheibe ausgewählt.
Wie groß ist der zufällige Abstand vom Ursprung im Mittel?

Zur stochastischen Modellierung betrachten wir einen 2-dimensionalen Zufallsvektor (X_1, X_2) auf W-Raum $(\Omega, \mathcal{A}, P)$, wobei $P_{(X_1,X_2)}$ konzentriert ist auf

$$K = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$$

und

$$P_{(X_1,X_2)}\big|_K = \frac{1}{\pi} \cdot m\big|_K \ \text{ mit } \ m = LB - \text{Maß auf } \mathcal{B}_2.$$

Gesucht ist dann

$$E\left(\sqrt{X_1^2 + X_2^2}\right).$$

1. Lösungsweg: Verteilung von $Y = \sqrt{X_1^2 + X_2^2}$ ermitteln.
F sei die Verteilungsfunktion von Y.
Dann gilt für $0 \leq y \leq 1$:

$$\begin{aligned} F(y) &= P(Y \leq y) \\ &= P(X_1^2 + X_2^2 \leq y^2) \\ &= P_{(X_1,X_2)}(\underbrace{\{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq y^2\}}_{\text{Kreis um } 0 \text{ mit Radius } y}) \\ &= \tfrac{1}{\pi} \cdot \pi y^2 = y^2, \end{aligned}$$

woraus folgt

$$F(y) = \begin{cases} 0 & , \quad y < 0 \\ y^2 & , \quad 0 \leq y \leq 1 \\ 1 & , \quad y > 1 \end{cases}$$

Zu F existiert Diche f mit

$$f(y) = F'(y) = \begin{cases} 0 & , \quad y < 0 \text{ oder } y > 1 \\ 2y & , \quad 0 \leq y \leq 1 \end{cases}$$

(denn man sieht sofort, dass die angegebene Funktion

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

erfüllt, vgl. Bemerkung 3.11 c)). Damit folgt

$$\begin{aligned} E\sqrt{X_1^2 + X_2^2} &= EY \overset{Satz\ 3.12b)}{=} \int_{-\infty}^{\infty} y \cdot f(y)dy \\ &= \int_0^1 2y^2 dy = \tfrac{2}{3}. \end{aligned}$$

2. Lösungsweg: Anwendung von Satz 3.13
(X_1, X_2) hat Dichte

$$f(x) = \begin{cases} \frac{1}{\pi} & , \quad x \in K \\ 0 & , \quad \text{sonst} \end{cases}$$

Also gilt nach Satz 3.13:

$$\begin{aligned} E\sqrt{X_1^2 + X_2^2} &= \int_K \sqrt{x_1^2 + x_2^2} \cdot \tfrac{1}{\pi} d(x_1, x_2) \\ &\overset{Polarkoordinaten}{=} \int_0^{2\pi} \int_0^1 r \cdot \tfrac{1}{\pi} \cdot r \; dr d\varphi \\ &= \int_0^1 2r^2 dr = \tfrac{2}{3}. \end{aligned}$$

3.14 Bemerkung. $\int_A f(x)\,dx$ in Satz 3.13 ist gleich dem L-B-Maß der Ordinatenmenge $\{(x, y) \in \mathbb{R}^2 \mid x \in A,\, 0 \leq y \leq f(x)\}$ von $f|_A$ in $\mathbb{R}^2$.

3.15 Lemma. X sei eine reelle ZV; $0 \leq \alpha < \beta < \infty$. Dann gilt

$$E|X|^\beta < \infty \Longrightarrow E|X|^\alpha < \infty\,.$$

Beweis:
$|X|^\alpha \leq 1 + |X|^\beta \Rightarrow E(|X|^\alpha) \leq 1 + E(|X|^\beta)$ □

3.16 Definition. X sei eine reelle ZV; $k \in \mathbb{N}$. Im Falle der Existenz heißt EX^k das ***k*-te Moment** von X und $E(X - EX)^k$ das ***k*-te zentrale Moment** von X. Ist X integrierbar, dann heißt $V(X) := E(X - EX)^2$ die **Varianz** von X und $\sigma(X) := \sqrt{V(X)}$ die **Streuung** von X.

3.17 Satz. X sei eine reelle integrierbare ZV.

a) $V(X) = EX^2 - (EX)^2$.

b) $V(aX + b) = a^2 V(X)$ $(a,\ b \in \mathbb{R})$.

Beweis:

a)

$$\begin{aligned} V(X) &= E(X^2 - 2X \cdot EX + (EX)^2) \\ &= E(X^2) - 2 \cdot EX \cdot EX + (EX)^2 \\ &= E(X^2) - (EX)^2. \end{aligned}$$

b)

$$\begin{aligned} V(a \cdot X + b) &= E[(a \cdot X + b - E(a \cdot X + b))^2] \\ &= E[(a \cdot X + b - a \cdot EX - b)^2] \\ &= E[a^2 \cdot (X - EX)^2] \\ &= a^2 \cdot V(X) \end{aligned}$$

□

Beispiel zu Satz 3.17.

a) X sei $b(n,p)$-verteilt mit $n \in \mathbb{N}$, $p \in [0,1]$. Dann gilt $V(x) = np(1-p)$.

Begründung: Es gelte

$$P[X = k] = \binom{n}{k} \cdot p^k (1-p)^{n-k} \ (k \in \{0, \ldots, n\}).$$

Setze

$$g(s) = \sum_{k=0}^{n} \binom{n}{k} \cdot p^k (1-p)^{n-k} \cdot s^k = (ps + 1 - p)^n.$$

Zweimaliges Ableiten aller drei Ausdrücke liefert

$$\begin{aligned} g''(s) &= \sum_{k=0}^{n} k(k-1) \cdot \tbinom{n}{k} \cdot p^k (1-p)^{n-k} \cdot s^{k-2} \\ &= n \cdot (n-1) \cdot (ps + 1 - p)^{n-2} \cdot p^2, \end{aligned}$$

also

$$\begin{aligned} g''(1) &= \sum_{k=0}^{n} k \cdot (k-1) \tbinom{n}{k} \cdot p^k (1-p)^{n-k} \\ &= n \cdot (n-1) \cdot (p + 1 - p)^{n-2} \cdot p^2 \\ &= n \cdot (n-1) \cdot p^2. \end{aligned}$$

Damit

$$\begin{aligned} V(X) &= E(X^2) - (EX)^2 = E(X \cdot (X-1)) + EX - (EX)^2 \\ &= n \cdot (n-1) \cdot p^2 + n \cdot p - (n \cdot p)^2 \\ &= n^2 \cdot p^2 - n \cdot p^2 + n \cdot p - n^2 \cdot p^2 \\ &= n \cdot (p - p^2) = n \cdot p \cdot (1-p). \end{aligned}$$

b) X sei $\pi(\lambda)$-verteilt mit $\lambda > 0$. Dann gilt $V(X) = \lambda$.

Begründung:
Sei X $\pi(\lambda)$-verteilt, also

$$P[X = k] = \frac{\lambda^k}{k!} \cdot e^{-\lambda} \ (k \in \mathbb{N}_0).$$

Dann gilt

$$\begin{aligned}
E(X\cdot(X-1)) &= \sum_{k=0}^{\infty} k\cdot(k-1)\cdot\tfrac{\lambda^k}{k!}\cdot e^{-\lambda}\\
&= \sum_{k=2}^{\infty}\tfrac{\lambda^k}{(k-2)!}\cdot e^{-\lambda} = \lambda^2\cdot e^{-\lambda}\cdot\sum_{k=2}^{\infty}\tfrac{\lambda^{k-2}}{(k-2)!}\\
&= \lambda^2\cdot e^{-\lambda}\cdot e^{\lambda} = \lambda^2,
\end{aligned}$$

woraus folgt

$$\begin{aligned}
V(X) &= E(X\cdot(X-1)) + EX - (EX)^2\\
&= \lambda^2+\lambda-(\lambda)^2=\lambda.
\end{aligned}$$

c) X sei $N(a,\sigma^2)$-verteilt mit $a\in\mathbb{R}$, $\sigma>0$.
Dann gilt $E(X-a)^{2k-1}=0$, $E(X-a)^{2k}=\sigma^{2k}\prod\limits_{j=1}^{k}(2j-1)\quad(k\in\mathbb{N})$.

(vgl. Übungen).

3.18 Satz. X sei eine reelle ZV auf einem Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},P)$. Dann gilt für jedes $\varepsilon>0$, $r>0$:

$$P[|X|\geq\varepsilon]\leq\varepsilon^{-r}E|X|^r$$

(Markoffsche Ungleichung).

Für $r=2$ erhält man die sog. **Tschebyschevsche Ungleichung**, die bei integrierbarem X auch in der Variante

$$P[|X-EX|\geq\varepsilon]\leq\varepsilon^{-2}V(X)$$

angegeben wird.

Beweis.

a)

$$\begin{aligned}
P(|X|\geq\varepsilon) &= \int\limits_{\{\omega:|X(\omega)|\geq\varepsilon\}} 1\ dP\\
&\leq \int\limits_{\{\omega:|X(\omega)|\geq\varepsilon\}} \tfrac{|X(\omega)|^r}{\varepsilon^r}\ dP\\
&\quad\text{(da Integrand auf dem Integrationsbereich } \geq 1 \text{ ist)}\\
&\leq \int\limits_{\Omega}\tfrac{|X(\omega)|^r}{\varepsilon^r}\ dP\\
&\quad\text{(da Integrand nichtnegativ ist)}\\
&= \varepsilon^{-r}\cdot E(|X|^r).
\end{aligned}$$

b) Mit $Y = |X - EX|$ und $r = 2$ folgt aus a):

$$\begin{aligned} P(|X - EX| \geq \varepsilon) &= P(|Y| \geq \varepsilon) \\ &\leq \varepsilon^{-2} \cdot E(|Y|^2) \\ &= \tfrac{V(X)}{\varepsilon^2} \end{aligned}$$

$\square$

4 Unabhängigkeit

W-Raum $(\Omega, \mathcal{A}, P), A, B \in \mathcal{A}$ mit $P(A) > 0$ und $P(B) > 0$.
Frage: Wann beeinflussen sich A und B gegenseitig nicht?
Naheliegende Forderung dafür:
Wenn gilt

$$P(A|B) = P(A) \text{ und } P(B|A) = P(B).$$

Äquivalent, und auch im Falle $P(A) = 0$ oder $P(B) = 0$ formulierbar:

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

Beispiel: Werfen zweier echter Würfel.
A = "Erste Augenzahl ist gerade"
B = "Zweite Augenzahl ist gerade"
C = "Summe der Augenzahlen ist ungerade".
Dann ist

$$P(A) = P(B) = P(C) = \frac{1}{2}$$

und

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4},$$

also gilt hier

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

4.1 Definition. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Eine Familie $\{A_i \mid i \in I\}$ von Ereignissen $A_i \in \mathcal{A}$ heißt **unabhängig** (ausführlich: stochastisch unabhängig bzgl. P), falls für jede nichtleere endliche Menge $K \subset I$ gilt

$$P\left(\bigcap_{k \in K} A_k\right) = \prod_{k \in K} P(A_k)\,.$$

4.2 Definition. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Eine Familie $\{X_i \mid i \in I\}$ von $(\Omega_i, \mathcal{A}_i)$ - ZVn auf $(\Omega, \mathcal{A}, P)$ heißt **unabhängig**, wenn gilt: Für jede nichtleere endliche Indexmenge $\{i_1, \ldots, i_n\} \subset I$ und jede Wahl von Mengen $A_{i_\nu} \in \mathcal{A}_{i_\nu}$ $(\nu = 1, \ldots, n)$ ist

$$P\left[X_{i_1} \in A_{i_1}, \ldots, X_{i_n} \in A_{i_n}\right] = \prod_{\nu=1}^{n} P\big[X_{i_\nu} \in A_{i_\nu}\big]\,.$$

Sprechweise: Unabhängigkeit der ZVn statt Unabhängigkeit der Familie der ZVn.

4.3 Lemma. Eine Familie von Ereignissen ist genau dann unabhängig, wenn jede endliche Teilfamilie unabhängig ist. Entsprechendes gilt für Zufallsvariablen.
Beweis: Folgt direkt aus den beiden Definitionen. □

4.4 Bemerkung zu Definition 4.1 bzw. Definition 4.2.
Paarweise Unabhängigkeit $\not\Longrightarrow$ Unabhängigkeit

Begründung:
Im Beispiel oben sind $\{A,B\},\{A,C\}$ und $\{B,C\}$ jeweils unabhängig, aber $\{A,B,C\}$ ist wegen

$$P(A\cap B\cap C)=P(\emptyset)=0\neq P(A)\cdot P(B)\cdot P(C)$$

nicht unabhängig. □

4.5 Bemerkung. Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},P)$; Ereignisse $A_i\in\mathcal{A}$, $i\in I$, mit (reellen) ZVn χ_{A_i} (Indikatorfunktionen) auf $(\Omega,\mathcal{A},P)$.

$$\{A_i\mid i\in I\}\text{ unabhängig }\Longleftrightarrow\{\chi_{A_i}\mid i\in I\}\text{ unabhängig}$$

Begründung:
"$\Rightarrow$" Wegen

$$[\mathcal{X}_{A_i}\in B]\in\{\emptyset,A_i,A_i^c,\Omega\}$$

genügt es zu zeigen:
Sind $\{A_i|i\in I\}$ unabhängig und $\overline{A}_i\in\{\emptyset,A_i,A_i^c,\Omega\}$, so ist auch $\{\overline{A}_i,|i\in I\}$ unabhängig. Da bei der Definition der Unabhängigkeit nur **endliche** Indexteilmengen in I betrachtet werden, können wir *oBdA* I endlich ansetzen. Dann genügt es zu zeigen: Aus $\{A_1,A_2,\ldots,A_n\}$ unabhängig und $\overline{A}_1\in\{\emptyset,A_i,A_i^c,\Omega\}$ folgt $\{\overline{A}_1,A_2,\ldots,A_n\}$ unabhängig. Dies ist trivial für $\overline{A}_1\in\{\emptyset,A_1,\Omega\}$.
Ist nun $\overline{A}_1=A_1^c$ so gilt wegen $\{A_1,\ldots,A_n\}$ unabhängig:

$$\begin{aligned}
&P(A_1^c\cap A_{i_2}\cap\ldots\cap A_{i_k})\\
&=P(A_{i_2}\cap\ldots\cap A_{i_k})-P(A_1\cap A_{i_2}\cap\ldots\cap A_{i_k})\\
&=P(A_{i_2})\cdot\ldots\cdot P(A_{i_k})-P(A_1)\cdot P(A_{i_2})\cdot\ldots\cdot P(A_{i_k})\\
&=(1-P(A_1))\cdot P(A_{i_2})\cdot\ldots\cdot P(A_{i_k})\\
&=P(A_1^c)\cdot P(A_{i_2})\cdot\ldots\cdot P(A_{i_k})\quad\rightsquigarrow\ \text{Beh.}
\end{aligned}$$

"$\Leftarrow$" Folgt aus

$$\begin{aligned}
&P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(\mathcal{X}_{A_{i_1}} \in \{1\}, \ldots, \mathcal{X}_{A_{i_k}} \in \{1\}) \\
&= P(\mathcal{X}_{A_{i_1}} \in \{1\}) \cdot \ldots \cdot P(\mathcal{X}_{A_{i_k}} \in \{1\}) \\
&\qquad (\text{da } \{\mathcal{X}_{A_i} | i \in I\} \text{ unabhängig}) \\
&= P(A_{i_1}) \cdot \ldots \cdot P(A_{i_k}).
\end{aligned}$$

□

4.6 Satz. Gegeben seien $(\Omega_i, \mathcal{A}_i)$ - ZVn X_i auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und Abb. $g_i : (\Omega_i, \mathcal{A}_i) \to (\Omega_i', \mathcal{A}_i')$, $i \in I$. Dann gilt:

$$\{X_i \mid i \in I\} \text{ unabhängig } \Longrightarrow \{g_i \circ X_i \mid i \in I\} \text{ unabhängig}$$

Beweis. Wegen Lemma 4.3 *oBdA* $I = \{1, \ldots, n\}$ und $\{i_1, \ldots, i_n\} = I$.

$$\begin{aligned}
&P[g_1 \circ X_1 \in A_1', \ldots, g_n \circ X_n \in A_n'] \\
&= P[X_1 \in g_1^{-1}(A_1'), \ldots, X_n \in g_n^{-1}(A_n')] \\
&= \prod_{i=1}^{n} P[X_i \in g_i^{-1}(A_i')] \quad (\text{da } \{X_i : i \in I\} \text{ unabhängig}) \\
&= \prod_{i=1}^{n} P[g_i \circ X_i \in A_i'].
\end{aligned}$$

□

Bsp. Zwei Leute füllen unbeeinflusst voneinander je einen Lottoschein aus. Dann sind das Quadrat des ersten Gewinns und die Wurzel des zweiten Gewinns unabhängig.

4.7 Satz. Gegeben seien eine unabhängige Familie $\{X_i \mid i \in I\}$ reeller ZVn auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und eine Zerlegung $I = \sum_{j \in J} I_j$ mit $|I_j| < \infty$. Das $|I_j|$-tupel $\{X_k \mid k \in I_j\}$ sei mit Y_j bezeichnet, $j \in J$.

a) Die Familie $\{Y_j \mid j \in J\}$ ist unabhängig.

b) Sind Abb. $g_j : (\mathbb{R}^{|I_j|}, \mathcal{B}_{|I_j|}) \to (\mathbb{R}, \mathcal{B})$, $j \in J$, gegeben, so ist (nach Satz 4.6) die Familie $\{g_j \circ Y_j \mid j \in J\}$ unabhängig.

4.8 Bemerkung. Satz 4.7 lässt sich auf $(\Omega_i, \mathcal{A}_i)$ - ZVn und Abb. $g_j : \bigotimes_{i \in I_j} (\Omega_i, \mathcal{A}_i) \to (\Omega_j', \mathcal{A}_j')$ verallgemeinern.

Beispiel:
Sind X, Y, Z unabhängige reelle Zven, dann sind auch

$$(X, Y), Z$$

und

$$X \cdot Y, Z^2$$

jeweils unabhängig.

Beweis von Satz 4.7 und Bemerkung 4.8
b) folgt aus a) und Satz 4.6.
In a) kann nach Lemma 4.3 *oBdA* $|J| < \infty$ angenommen werden. In diesem Fall ist auch $|I| < \infty$.
Zwecks Vereinfachung der Schreibweise wird der Beweis nur im Spezialfall $I = \{1, 2, 3, 4\}$, $J = \{1, 2\}$, $I_1 = \{1, 2\}$ und $I_2 = \{3, 4\}$ durchgeführt, der allgemeiner Fall folgt analog.
Also

$$Y_1 = (X_1, X_2), Y_2 = (X_3, X_4)$$

und zu zeigen ist

$$P[Y_1 \in A, Y_2 \in B] = P[Y_1 \in A] \cdot P[Y_2 \in B]$$

für alle $A \in \mathcal{A}_1 \otimes \mathcal{A}_2, B \in \mathcal{A}_3 \otimes \mathcal{A}_4$.

Fall 1: $A = A_1 \times A_2, B = A_3 \times A_4$ mit $A_i \in \mathcal{A}_i$ $(i \in I)$.
Dann gilt

$$\begin{aligned}
&P[Y_1 \in A, Y_2 \in B] \\
&= P[(X_1, X_2) \in A_1 \times A_2, (X_3, X_4) \in A_3 \times A_4] \\
&= P[X_1 \in A_1, X_2 \in A_2, X_3 \in A_3, X_4 \in A_4] \\
&= P[X_1 \in A_1] \cdot P[X_2 \in A_2] \cdot P[X_3 \in A_3] \cdot P[X_4 \in A_4] \\
&\qquad \text{(da } X_1, \ldots X_4 \text{ unabhängig sind)} \\
&= P[X_1 \in A_1, X_2 \in A_2] \cdot P[X_3 \in A_3, X_4 \in A_4] \\
&\qquad \text{(da } X_1, X_2 \text{ und } X_3, X_4 \text{ jeweils unabhängig sind, vgl. Lemma 4.3)} \\
&= P[Y_1 \in A] \cdot P[Y_2 \in B] \qquad \rightsquigarrow \text{Beh.}
\end{aligned}$$

Fall 2: $A \in \mathcal{A}_1 \otimes \mathcal{A}_2, B = A_3 \times A_4$ mit $A_3 \in \mathcal{A}_3$ und $A_4 \in \mathcal{A}_4$
zu zeigen:

$$P[Y_1 \in A, Y_2 \in B] = P[Y_1 \in A] \cdot P[Y_2 \in B].$$

Als Funktion von A sind beide Seiten (bei festgehaltenem B) endliche Maße. Daher folgt die Behauptung aus Fall 1 mit dem Eindeutigkeitssatz für endliche Maße.

Fall 3: $A \in \mathcal{A}_1 \otimes \mathcal{A}_2, B \in \mathcal{A}_3 \otimes \mathcal{A}_4$
Folgt analog zum Vorgehen bei Fall 2 aus Fall 2 und dem Eindeutigkeitssatz für endliche Maße. $\square$

4.9 Satz. Seien X_i $(i = 1, \ldots, n)$ reelle ZVn auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Verteilungsfunktionen F_i. Der Zufallsvektor $X := (X_1, \ldots, X_n)$ habe die n-dimensionale VF F.

a) $\{X_1, \ldots, X_n\}$ unabhängig $\iff (*) \underset{(x_1,\ldots,x_n)\in\mathbb{R}^n}{\forall} F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_i(x_i)\,.$

b) Existieren zu F_i $(i = 1, \ldots, n)$, F Dichten f_i bzw. f, dann gilt

$$(*) \iff \begin{cases} f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i) \\ \text{für } \underbrace{\text{Lebesgue - fast alle}}_{\lambda\text{-fast alle mit } \lambda \,=\, \text{L-B-Maß auf } \mathcal{B}_n} (x_1, \ldots, x_n) \in \mathbb{R}^n. \end{cases}$$

Beweis

a) "$\Rightarrow$" klar,wähle $A_i = (-\infty, x_i]$ in Definition 4.2
"$\Leftarrow$" Aus $(*)$ folgt:

$$P[(X_1, \ldots, X_n) \in I_1 \times \ldots \times I_n] = P[X_1 \in I_1] \cdot \ldots \cdot P[X_n \in I_n]$$

für alle Intervalle $I_j \in \{(-\infty, x] : x \in \mathbb{R}\}$.

Sukzessives Anwenden des Eindeutigkeitssatzes für endliche Maße liefert zuerst

$$P[(X_1, \ldots, X_n) \in B_1 \times I_2 \times \ldots \times I_n] = P[X_1 \in B_1] \cdot P[X_2 \in I_2] \cdot \ldots P[X_n \in I_n]$$

für alle $B_1 \in \mathcal{B}, I_2, \ldots, I_n \in \{(-\infty, x] : x \in \mathbb{R}\}$, und nach n-maligem Anwenden

$$P[(X_1, \ldots, X_n) \in B_1 \times B_2 \times \ldots \times B_n] = P[X_1 \in B_1] \cdot \ldots \cdot P[X_n \in B_n]$$

für alle $B_1, \ldots, B_n \in \mathcal{B}$, was zu zeigen war.

b) "$\Rightarrow$" Wendet man auf $(*)$ die Differentiation

$$\frac{\partial^n}{\partial x_1 \ldots \partial x_n}$$

L-f.ü. an, so folgt die Behauptung aus Bemerkung 3.11b).

"$\Leftarrow$" Es gelte

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i) \quad \text{für L-f.a. } (x_1, \ldots x_n) \in \mathbb{R}^n.$$

Dann gilt auch

$$\begin{aligned}
F(x_1,\dots,x_n) &= \int\limits_{(-\infty,x_1]\times\ldots\times(-\infty,x_n]} f(u_1,\dots,u_n)d(u_1,\dots,u_n) \\
&\overset{\text{Vor.}}{=} \int\limits_{(-\infty,x_1]\times\ldots\times(-\infty,x_n]} f_1(u_1)\cdot\ldots\cdot f_n(u_n)\; d(u_1,\dots,u_n) \\
&\overset{Fubini}{=} \int\limits_{-\infty}^{x_1}\ldots\int\limits_{-\infty}^{x_n} f_1(u_1)\cdot\ldots\cdot f_n(u_n)\; du_n\ldots du_1 \\
&= \int\limits_{-\infty}^{x_1} f_1(u_1)du_1\cdot\ldots\cdot\int\limits_{-\infty}^{x_n} f_n(u_n)du_n \\
&= F_1(x_1)\cdot\ldots\cdot F_n(x_n).
\end{aligned}$$

□

4.10 Satz. Seien $X_1,\dots,X_n$ unabhängige reelle ZVn auf einem Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},P)$ mit endlichen Erwartungswerten. Dann existiert

$$E(X_1\cdot\ldots\cdot X_n) = EX_1\cdot\ldots\cdot EX_n.$$

— Entsprechendes gilt für komplexe ZVn $X_k:(\Omega,\mathcal{A})\to(\mathbb{R}^2,\mathcal{B}_2)$, wobei

$$EX_k := E\text{ Re }X_k + iE\text{ Im }X_k \quad (k=1,\dots,n)$$

Beweis: *oBdA* $n=2$ und X_i reellwertig.
1. Weg
Im Falle $X_1=\mathcal{X}_{A_1}, X_2=\mathcal{X}_{A_2}$ mit A_1,A_2 unabhängig gilt

$$\begin{aligned}
E(X_1\cdot X_2) &= E(\mathcal{X}_{A_1\cap A_2}) = P(A_1\cap A_2) \\
&= P(A_1)\cdot P(A_2) \quad (\text{da } A_1,A_2 \text{ unabhängig}) \\
&= E\mathcal{X}_{A_1}\cdot E\mathcal{X}_{A_2} = EX_1\cdot EX_2.
\end{aligned}$$

Dann gemäß der schrittweisen Definition des Maßintegrals den allgemeinen Fall darauf zurückführen.
2. Weg: Eleganter, mit Maßtheorie.
ZVen $X_i:(\Omega,\mathcal{A},P)\to(\Omega_i,\mathcal{A}_i)\quad(i\in\{1,2\})$.
Produkt Messraum $(\Omega_1\times\Omega_2,\mathcal{A}_1\otimes\mathcal{A}_2)$, wobei $(\mathcal{A}_1\otimes\mathcal{A}_2)$ die kleinste σ-Algebra in $\Omega_1\times\Omega_2$ ist, die alle Mengen der Form $A_1\times A_2\ \ (A_i\in\mathcal{A}_i\ (i\in\{1,2\}))$ enthält.
Man kann zeigen:
Es existiert ein eindeutiges Maß Q auf $\mathcal{A}_1\otimes\mathcal{A}_2$ mit

$$Q(A_1\times A_2) = P_{X_1}(A_1)\cdot P_{X_2}(A_2)$$

für alle $A_1\in\mathcal{A}_2, A_2\in\mathcal{A}_2$.
Bezeichnung: Q ist das Produkt-Maß zu P_{X_1} und $P_{X_2}, Q=P_{X_1}\otimes P_{X_2}$.

Nun gilt:
(X_1, X_2) unabhängig

$$\overset{Def.}{\Leftrightarrow} \forall A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2: \quad P[X_1 \in A_1, X_2 \in A_2] = P[X_1 \in A_1] \cdot P[X_2 \in A_2]$$

$$\Leftrightarrow \forall A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2: \quad P_{(X_1,X_2)}(A_1 \times A_2) = (P_{X_1} \otimes P_{X_2})(A_1 \times A_2)$$

$$\Leftrightarrow P_{(X_1,X_2)} = P_{X_1} \otimes P_{X_2}$$

wobei die letzte Äquivalenz aus der Eindeutigkeit des Produkt-Maßes folgt.

In anderen Worten:
Im Falle der Unabhängigkeit von X_1 und X_2 ist die gemeinsame Verteilung $P_{(X_2,X_2)}$ also gerade gleich dem Produkt der Randverteilungen P_{X_1} und P_{X_2}.

Weiter gilt:
Satz von Fubini (für Produkt W-Maße, analog für Produkte von σ-endlichen Maßen).

a) $f : (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \to (\mathbb{R}_+, \mathcal{B}_+)$, Q_1 bzw. Q_2 seien W-Maße auf $\mathcal{A}_1$ bzw. $\mathcal{A}_2$. $Q = Q_1 \otimes Q_2$ sei das Produkt W-Maß von Q_1 und Q_2.

   Dann gilt:

$$\begin{aligned}\int\limits_{\Omega_1\times\Omega_2} f\,dQ &= \int\limits_{\Omega_1} \underbrace{[\int\limits_{\Omega_2} f(\omega_1,\omega_2)dQ_2(\omega_2)]}_{\text{als Funktion von } \omega_1 \text{ messbar (!)}} dQ_1(\omega_1) \\ &= \int\limits_{\Omega_2} \underbrace{[\int\limits_{\Omega_1} f(\omega_1,\omega_2)dQ_1(\omega_1)]}_{\text{als Funktion von } \omega_2 \text{ messbar (!)}} dQ_2(\omega_2)\end{aligned}$$

b) Für

$$f : (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \to (\mathbb{R}, B)$$

   mit

$$\int\limits_{\Omega_1\times\Omega_2} |f|\,dQ \overset{\text{gemäß a)}}{=} \int\limits_{\Omega_1}\int\limits_{\Omega_2} |f(\omega_1,\omega_2)| dQ_2(\omega_2) dQ_1(\omega_1) < \infty$$

   gilt

$$\begin{aligned}\int\limits_{\Omega_1\times\Omega_2} f\,dQ &= \int\limits_{\Omega_1}[\int\limits_{\Omega_2} f(\omega_1,\omega_2)dQ_2(\omega_2)]dQ_1(\omega_1) \\ &= \int\limits_{\Omega_2}[\int\limits_{\Omega_1} f(\omega_1,\omega_2)dQ_1(\omega_1)]dQ_2(\omega_2),\end{aligned}$$

   wobei jeweils $[\ldots]$ als Funktion von ω_1 bzw. ω_2 integrierbar ist.

Damit Nachweis von $E(X_1 \cdot X_2) = EX_1 \cdot EX_2$:
Gemäß a) gilt für die messbare Funktion

$$f : \mathbb{R}^2 \to \mathbb{R}, \quad f(x_1, x_2) = x_1 \cdot x_2 :$$

$$\begin{aligned}
\int |f| d(P_{X_1} \otimes P_{X_2}) &= \int\limits_{\mathbb{R}} \int\limits_{\mathbb{R}} |x_1| \cdot |x_2| dP_{X_1} dP_{X_2} \\
&= \int\limits_{\mathbb{R}} |x_2| \cdot \underbrace{\int\limits_{\mathbb{R}} |x_1| \cdot dP_{X_1}}_{=E|X_1|<\infty} dP_{X_2} \\
&= E|X_1| \cdot \int |x_2| dP_{X_2} \\
&= E|X_1| \cdot E|X_2| \quad < \infty
\end{aligned}$$

nach Voraussetzung.
Damit erhalten wir

$$\begin{aligned}
E(X_1 \cdot X_2) &= \int\limits_{\mathbb{R}\times\mathbb{R}} x_1 \cdot x_2 dP_{(X_1,X_2)}(x_1, x_2) \\
&= \int\limits_{\mathbb{R}\times\mathbb{R}} x_1 \cdot x_2 d(P_{X_1} \otimes P_{X_2})(x_1, x_2) \\
& \text{(da } X_1, X_2 \text{ unabhängig)} \\
&= \int\limits_{\mathbb{R}} [\int\limits_{\mathbb{R}} x_1 \cdot x_2 dP_{X_1}(x_1)] dP_{X_2}(x_2) \\
& \text{(nach Teil b) des Satzes von Fubini,} \\
& \text{der wegen } \int |x_1 \cdot x_2| d(P_{X_1} \otimes P_{X_2}) < \infty \\
& \text{anwendbar ist)} \\
&= \int\limits_{\mathbb{R}} x_2 \cdot \underbrace{\int\limits_{\mathbb{R}} x_1 dP_{X_1}(x_1)}_{=EX_1 \in \mathbb{R}} dP_{X_2}(x_2) \\
&= EX_1 \cdot \int\limits_{\mathbb{R}} x_2 dP_{X_2}(x_2) \\
&= EX_1 \cdot EX_2
\end{aligned}$$

$\rightsquigarrow$ Satz 4.10. $\square$

4.11 Satz. Seien $X_1, \ldots, X_n$ unabhängige reelle ZVn mit endlichen Erwartungswerten. Dann gilt

$$V\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} V(X_j) \, .$$

Beweis:

$$\begin{aligned}
V(X_1 + \ldots + X_n) &= E[\left(\sum_{i=1}^{n} (X_i - EX_i)\right)^2] \\
&= \sum_{i=1}^{n} E[|X_i - EX_i|^2] + 2 \sum_{1\leq i<j\leq n} E[(X_i - EX_i) \cdot (X_j - EX_j)]
\end{aligned}$$

Für $i \neq j$ sind $X_i - EX_i$ und $X_j - EX_j$ unabhängig, also gilt nach Satz 4.10:

$$\begin{aligned}&E[(X_i - EX_i) \cdot (X_j - EX_j)]\\&= E[X_i - EX_i] \cdot E[X_j - EX_j]\\&= (EX_i - EX_i) \cdot (EX_j - EX_j)\\&= 0 \cdot 0 = 0.\end{aligned}$$

Damit

$$V(X_1 + \ldots + X_n) = \sum_{i=1}^{n} E[(X_i - EX_i]^2] + 0 = \sum_{i=1}^{n} V(X_i).$$

$\square$

4.12 Bemerkung. In Satz 4.11 genügt statt der Unabhängigkeit von $\{X_1, \ldots, X_n\}$ die sog. paarweise Unkorreliertheit, d.h.

$$\underset{j \neq k}{\forall} E(X_j - EX_j)(X_k - EX_k) = 0\,.$$

Vorbemerkung zu Definition 4.13:
Sind X, Y unabhängige reelle ZVen, so ist

$$P_{(X,Y)} \overset{s.o.}{=} P_X \otimes P_Y$$

bereits durch P_X und P_Y eindeutig festgelegt, und daher ist auch

$$P_{X+Y}$$

bereits eindeutig durch P_X und P_Y festgelegt, da

$$P_{X+Y}(A) = P_{(X,Y)}(\{(x, y) : x + y \in A\})$$

für $A \in \mathcal{B}$.

4.13 Definition. Sind X, Y zwei unabhängige reelle ZVn auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Verteilungen P_X, P_Y, so wird die Verteilung $P_{X+Y} =: P_X * P_Y$ der ZV $X + Y$ als **Faltung** von P_X und P_Y bezeichnet.

4.14 Bemerkung. Die Faltungsoperation ist kommutativ und assoziativ.
(Klar, denn da Addition in $\mathbb{R}$ kommutativ und assoziativ ist, gilt z.B.

$$X(\omega) + Y(\omega) = Y(\omega) + X(\omega) \quad \text{für alle } \omega \in \Omega \qquad \Longrightarrow \qquad P_{X+Y} = P_{Y+X}).$$

4.15 Satz. Seien X, Y zwei unabhängige reelle ZVn.

a) Besitzen X und Y Dichten f bzw. g, so besitzt $X+Y$ eine [als Faltung von f und g bezeichnete] Dichte h mit

$$h(t) = \underbrace{\int_{\mathbb{R}} f(t-y)g(y)\,dy = \int_{\mathbb{R}} g(t-x)f(x)\,dx}_{\text{Lebesgue-Integrale}},\ t \in \mathbb{R}\,.$$

b) Besitzen X und Y Zähldichten $(p_k)_{k\in\mathbb{N}_0}$ bzw. $(q_k)_{k\in\mathbb{N}_0}$, so besitzt $X+Y$ eine [als Faltung von (p_k) und (q_k) bezeichnete] Zähldichte $(r_k)_{k\in\mathbb{N}_0}$ mit

$$r_k = \sum_{j=0}^{k} p_{k-j}q_j = \sum_{i=0}^{k} q_{k-i}p_i\,,\ k \in \mathbb{N}_0\,.$$

Beweis (von a), Teil b) geht analog):

Wir zeigen: $\quad P[X+Y \le z] = \int\limits_{-\infty}^{z} h(t)dt \quad (z \in \mathbb{R})$

(vgl. Bemerkung 3.11 b))

Dazu:

$$\begin{aligned}
P[X+Y \le z] &= P_{(X,Y)}\big(\underbrace{\{(x,y)\in\mathbb{R}^2 : x+y \le z\}}_{=:A}\big)\\
&= \textstyle\int \mathcal{X}_A(x,y)\,dP_{(X,Y)}(x,y)\\
&= \textstyle\int \mathcal{X}_A(x,y)\ (P_X \otimes P_Y)\,(dx,dy)\\
&\ \text{da } X,Y \text{ unabhängig}\\
&\overset{Fubini}{=} \textstyle\int_{\mathbb{R}}\int_{\mathbb{R}} \mathcal{X}_{(-\infty,z-y]}(x)P_X(dx)P_Y(dy)\\
&= \textstyle\int_{\mathbb{R}}\int_{\mathbb{R}} \mathcal{X}_{(-\infty,z-y]}(x)\cdot f(x)\cdot g(y)\,dx\,dy\\
&\ \text{(nach Satz 3.13)}\\
&= \textstyle\int_{\mathbb{R}}\int_{\mathbb{R}} \mathcal{X}_{(-\infty,z-y]}(t-y)\cdot f(t-y)\cdot g(y)\,dt\,dy\\
&\ \text{(mit Substitution } x = t-y)\\
&\overset{Fubini}{=} \int\limits_{-\infty}^{z} [\textstyle\int_{\mathbb{R}} f(t-y)\cdot g(y))\,dy]\,dt
\end{aligned}$$

□

4.16 Bemerkung. Sei $I \neq \emptyset$ eine beliebige Menge. Gegeben seien Messräume $(\Omega_i, \mathcal{A}_i)$ und W-Maße Q_i auf $\mathcal{A}_i$, $i \in I$. Dann existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und $(\Omega_i, \mathcal{A}_i)$ – ZVn X_i auf $(\Omega, \mathcal{A}, P)$, $i \in I$, mit Unabhängigkeit von $\{X_i \mid i \in I\}$ und $P_{X_i} = Q_i$.

Begründung:

Wähle

$$\begin{aligned}
\Omega := \prod_{i\in I}\Omega_i &= \{(\omega_i)_{i\in I} : \omega_i \in \Omega_i\ (i \in I)\}\\
\mathcal{A} := \bigotimes_{i\in I}\mathcal{A}_i &= \text{kleinste } \sigma-\text{Algebra, die alle Mengen der Form}\\
&\ \prod_{i\in I} A_i, \text{mit } A_i \in \mathcal{A}_i\ (i\in I) \text{ und}\\
&\ A_i \neq \Omega_i \text{ nur für endlich viele } i \in I, \text{enthält.}
\end{aligned}$$

$$P := \bigotimes_{i\in I} Q_i = \text{ Produkt-W-Maß der } Q_i.$$

Hierbei ist $\bigotimes\limits_{i\in I} Q_i$ dasjenige W-Maß $P : \mathcal{A} \to \mathbb{R}$ mit

$$P\left(\prod_{i\in I} A_i\right) = \prod_{\substack{i\in I\\ A_i\neq\Omega_i}} Q_i(A_i)$$

für alle $\prod\limits_{i\in I} A_i$ aus der Definition von $\bigotimes\limits_{i\in I} A_i$.
Ein solches W-Maß P existiert eindeutig nach dem Satz von **Andersen und Jessen**.
Wir definieren nun

$$X_i : \Omega \to \Omega_i$$

als Projektion von Ω auf Ω_i.
Dann gilt:

- $P_{X_i} = Q_i \quad (i \in I)$
- $\{X_i : i \in I\}$ ist unabhängig

Anwendung
In Beweisen kann man annehmen, dass unabhängige ZVen auf dem gleichen W-Raum definiert sind, und dann Aussagen über ihre Verteilung herleiten.

4.17 Satz (2. Lemma von Borel und Cantelli). Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Die Familie $\{A_n;\ n \in \mathbb{N}\}$ von Ereignissen ($\in \mathcal{A}$) sei unabhängig. Dann gilt:

$$\sum_{n=1}^{\infty} P(A_n) = \infty \quad (\Leftarrow)\!\Rightarrow \quad \underbrace{P(\overline{\lim}\, A_n) = 1}_{\text{P-f.s. tritt } A_n \text{ unendlich oft auf}}.$$

Bemerkung:
Ohne Unabhängigkeitsvoraussetzung gilt nach dem 1. Lemma von Borel und Cantelli:

$$\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P(\overline{\lim}\, A_n) = 0.$$

Also folgt "$\Leftarrow$" in Satz 4.17 aus dem 1. Lemma von Borel und Cantelli.
Beispiel:
Unabhängige Folge von Versuchen mit Misserfolgswahrscheinlichkeit $\frac{1}{n}$ im n-ten Versuch.
Wegen

$$\sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

gilt dann nach Satz 4.17: P-f.s. treten unendlich viele Misserfolge auf.
Beweis von Satz 4.17

"$\Leftarrow$": Folgt aus dem 1. Lemma von Borel und Cantelli.
"$\Rightarrow$" Wir zeigen: $0 = P((\overline{\lim}\, A_n)^c) = P\left(\bigcup\limits_{n=1}^{\infty} \bigcap\limits_{k=n}^{\infty} A_k^c\right)$.
Dazu:

$$\begin{aligned}
P\left(\bigcap_{k=n}^{\infty} A_k^c\right) &= \lim_{N\to\infty} P\left(\bigcap_{k=n}^{N} A_k^c\right) \\
&\quad \text{(da W-Maß stetig von oben ist)} \\
&= \lim_{N\to\infty} \textstyle\prod_{k=n}^{N} P(A_k^c) \\
&\quad \text{(da } \{A_n : n\} \text{ unabhängig)} \\
&= \lim_{N\to\infty} \prod_{k=n}^{N} (1 - P(A_k)) \\
&\leq \lim_{N\to\infty} \prod_{k=n}^{N} e^{-P(A_k)} \\
&\quad \text{(wegen } 1 + x \leq e^x \text{ für } x \in \mathbb{R}) \\
&= \lim_{N\to\infty} e^{-\sum\limits_{k=n}^{N} P(A_k)} \\
&= 0 \\
&\quad \text{(da } \sum_{k=n}^{\infty} P(A_k) = \infty)
\end{aligned}$$

Daraus folgt:

$$\begin{aligned}
P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c\right) &\leq \sum_{n=1}^{\infty} P\left(\bigcap_{k=n}^{\infty} A_k^c\right) \\
&\quad \text{(da W-Maße } \sigma-\text{subadditiv sind)} \\
&= \sum_{n=1}^{\infty} 0 = 0.
\end{aligned}$$

$\square$

Bemerkung:

a) In Satz 4.17 darf die Unabhängigkeitsvoraussetzung nicht weggelassen werden.
Beispiel dazu:
Wählt man $A_n = A$ ($n \in \mathbb{N}$) für ein A mit $P(A) = \frac{1}{2}$, so gilt

$$\sum_{n=1}^{\infty} P(A_n) = \infty \text{ und } P(\overline{\lim}\, A_n) = P(A) = \frac{1}{2} \neq 1.$$

b) Sind $(A_n)_n$ unabhängig, so gilt nach dem 1. und 2. Lemma von Borell und Cantelli

$$P(\overline{\lim}\, A_n) \in \{0, 1\}.$$

Im Folgenden verallgemeinern wir diese Aussage.
4.18 Definition. Messraum $(\Omega, \mathcal{A})$. σ-Algebren $\mathcal{A}_n \subset \mathcal{A}$ ($n \in \mathbb{N}$).
$\mathcal{T}_n := \mathcal{F}\left(\bigcup\limits_{k=n}^{\infty} \mathcal{A}_k\right) \ldots$ die von $\mathcal{A}_n, \mathcal{A}_{n+1}, \ldots$ erzeugte σ-Algebra.

$\mathcal{T}_\infty := \bigcap\limits_{n=1}^{\infty} \mathcal{T}_n$ heißt die σ-Algebra der **terminalen Ereignisse** (tail events) der Folge $(\mathcal{A}_n)$.

Beispiele:

a) Messraum $(\Omega, \mathcal{A})$, $A_n \in \mathcal{A}$ $(n \in \mathbb{N})$.
Die σ-Algebra $\mathcal{A}_n$ sei definiert durch

$$\mathcal{A}_n = \{\emptyset, \Omega, A_n, A_n^c\}.$$

1. $\overline{\lim} A_n \in \mathcal{T}_\infty$,
denn:
$$\overline{\lim} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=N}^{\infty} \bigcup_{k=n}^{\infty} A_k \in \mathcal{T}_N$$
für jedes $N \in \mathbb{N}$, da
$$\bigcup_{k=n}^{\infty} A_k \in \mathcal{T}_N \quad \text{für } n \geq N$$
und $\mathcal{T}_n$ σ-Algebra ist.

2.
$$\begin{aligned}\underline{\lim} A_n &= \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k^c \right)^c \\ &= (\overline{\lim} A_n^c)^c \in \mathcal{T}_\infty,\end{aligned}$$
wobei der letzte Schritt unter Beachtung von $A_n^c \in \mathcal{A}_n$ und $\mathcal{T}_\infty$ σ-Algebra aus 1. folgt.

b) $X_n : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$
$\mathcal{A}_n = \mathcal{F}(X_n) = X_n^{-1}(\mathcal{B})$
Hierbei ist $X_n^{-1}(\mathcal{B})$ die kleinste σ-Algebra $\widehat{\mathcal{A}}$, für die X_n $\widehat{\mathcal{A}} - \mathcal{B}$-messbar ist.

1. $[(X_n)_n \text{ konvergent}] \in \mathcal{T}_\infty$
(da
$$\begin{aligned}&[(X_n)_n \text{ konvergent}] \\ &= [(X_n)_{n \geq k} \text{ konvergent}] \\ &= \{\omega \in \Omega \,:\, (X_n(\omega))_{n \geq k} \text{ Cauchy-Folge}\} \\ &= \left\{\omega \in \Omega \,:\, \forall l \in \mathbb{N}\, \exists n_0 \geq k\, \forall m, n \geq n_0 : |X_n(\omega) - X_m(\omega)| < \frac{1}{l}\right\} \\ &= \bigcap_{l \in \mathbb{N}} \bigcup_{n_0 \geq k} \bigcap_{m,n \geq n_0} \left\{\omega \in \Omega \,:\, |X_n(\omega) - X_m(\omega)| < \frac{1}{l}\right\} \in \mathcal{T}_k\end{aligned}$$
für alle $k \in \mathbb{N}$).

2. $[\sum_n X_n \text{ konvergent}] \in \mathcal{T}_\infty$
(da mit der Bezeichnung $S_n = \sum_{k=1}^n X_n$ gilt

$$\left[\sum_n X_n \text{ konvergent}\right] = [(S_n)_{n\geq k} \text{ konvergent}] \in \mathcal{T}_k \text{ für alle } k \in \mathbb{N},$$

was analog zu oben aus

$$\left\{\omega \in \Omega \,:\, |S_n(\omega) - S_m(\omega)| < \frac{1}{l}\right\} \in \mathcal{T}_k$$

für $n, m \geq k$ folgt. Letzteres sieht man ein, indem man beachtet, dass für $n > m \geq k$

$$S_n - S_m = X_{m+1} + \cdots + X_n$$

nach Satz 2.12 $\mathcal{T}_k$-$\mathcal{B}$-messbar ist.)

Erweiterung von Definition 4.1 / 4.2:
W-Raum $(\Omega, \mathcal{A}, P)$, Mengensysteme $\mathcal{C}_i \subset \mathcal{A}$ $(i \in I \neq \emptyset)$.
$\{\mathcal{C}_i : i \in I\}$ heißt **unabhängig**, falls für jede Wahl von Ereignissen $C_i \in \mathcal{C}_i$ $(i \in I)$ die Familie $\{C_i : i \in I\}$ unabhängig ist.

Beispiel:

a) Sind $(A_n)_{n\in\mathbb{N}}$ unabhängig, so ist auch

$$(\{\emptyset, \Omega, A_n, A_n^c\} : n \in \mathbb{N})$$

unabhängig (vgl. Begründung von Bemerkung 4.5).

b) Sind die reellen ZVen $(X_n)_{n\in\mathbb{N}}$ unabhängig, so ist auch

$$(X_n^{-1}(\mathcal{B}) : n \in \mathbb{N})$$

unabhängig
(da

$$\begin{aligned}
&P[X_{n_1}^{-1}(B_1) \cap \ldots \cap X_{n_k}^{-1}(B_k)] \\
&= P[X_{n_1} \in B_1, \ldots, X_{n_k} \in B_k] \\
&\overset{X_i \text{ unabhängig}}{=} P[X_{n_1} \in B_1] \cdot \ldots \cdot P[X_{n_k} \in B_k] \\
&= P(X_{n_1}^{-1}(B_1)) \cdot \ldots \cdot P(X_{n_k}^{-1}(B_k))
\end{aligned}$$

ist).

4.19 Satz (Null–Eins–Gesetz von Kolmogorov). Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$.
Sei $(\mathcal{A}_n)$ eine **unabhängige** Folge von σ-Algebren $\mathcal{A}_n \subset \mathcal{A}$.
Dann gilt für jedes terminale Ereignis A von $(\mathcal{A}_n)$ entweder $P(A) = 0$ oder $P(A) = 1$.

Beispiel: Unabhängige relle ZVen $(X_n)_n$ auf W-Raum $(\Omega, \mathcal{A}, P)$. Dann gilt

$$\begin{aligned} P[\sum_n X_n \text{ konvergent}] &\in \{0,1\} \\ P[(X_n)_n \text{ konvergent}] &\in \{0,1\} \end{aligned}$$

Vorbemerkung zum Beweis von Satz 4.19:
4.20 Def. Sei $\Omega \neq \emptyset$ und $\mathcal{C} \subseteq \mathcal{P}(\Omega)$.
$\mathcal{C}$ heißt **Dynkin-System**, falls gilt:

1. $\Omega \in \mathcal{C}$
2. $A, B \in \mathcal{C}$ mit $A \subseteq B \Rightarrow B \backslash A \in \mathcal{C}$
3. $A_n \in \mathcal{C}$ $(n \in \mathbb{N})$ paarweise disjunkt $\Rightarrow \sum\limits_{n=1}^{\infty} A_n \in \mathcal{C}$.

$\mathcal{D}(\mathcal{C})$ sei das kleinste Dynkin-System, das $\mathcal{C}$ enthält.

4.21 Lemma.

a) $\mathcal{D}$ Dynkin-System $\quad\Rightarrow\quad \emptyset \in \mathcal{D}$ und aus $A \in \mathcal{D}$ folgt $A^c \in \mathcal{D}$

b)
$$\begin{aligned} \mathcal{A}\ \sigma-\text{Algebra} \quad\Leftrightarrow\quad & \mathcal{A} \text{ Dynkin-System und } \mathcal{A} \cup -\text{ stabil} \\ & (\text{ d. h. aus } A, B \in \mathcal{A} \text{ folgt } A \cup B \in \mathcal{A}) \end{aligned}$$

c)
$$\begin{aligned} \mathcal{A}\ \sigma-\text{Algebra} \quad\Leftrightarrow\quad & \mathcal{A} \text{ Dynkin-System und } \mathcal{A} \cap -\text{stabil} \\ & (\text{d. h. aus } A, B \in \mathcal{A} \text{ folgt } A \cap B \in \mathcal{A}). \end{aligned}$$

Beweis

a) $\emptyset = \Omega \backslash \Omega \in \mathcal{D}$ nach Def. 1.) + 2.)

$A \in \mathcal{D} \Rightarrow A^c = \Omega \backslash A \in \mathcal{D}$ nach Def. 2.), beachte $A \subseteq \Omega$ und $\Omega \in \mathcal{D}$.

b) "$\Rightarrow$" klar
"$\Leftarrow$" Sei $\mathcal{A}$ Dynkin-System und $\cup$-stabil. Dann gilt:

(1) $\emptyset \in \mathcal{A}$ nach a)

(2) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$ nach a)

(3) $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A} \quad$ (da $\mathcal{A}$ $\cup$-stabil)

(4) $A_n \in \mathcal{A}$ $(n \in \mathbb{N})$ paarweise disjunkt $\Rightarrow \sum\limits_{n=1}^{\infty} A_n \in \mathcal{A}$ nach Definition Dynkin-System.

Mit

$$A_1 \cup A_2 \cup A_3 \cup \ldots = A_1 + (A_2 \backslash A_1) + (A_3 \backslash (A_1 \cup A_2)) + \ldots$$

und

$$A \setminus B = A \cap B^c = (A^c \cup B)^c$$

folgt daraus die Behauptung.

c) "$\Rightarrow$" klar
"$\Leftarrow$" Sei $\mathcal{A}$ $\cap$-stabiles Dynkin-System. Wegen $A \cup B = (A^c \cap B^c)^c$ ist $\mathcal{A}$ $\cup$-stabil, und mit b) folgt Beh.

$\square$

4.22 Lemma.
W-Raum $(\Omega, \mathcal{A}, P), \mathcal{C} \subseteq \mathcal{P}(\Omega), \mathcal{C}_1, \mathcal{C}_2 \subseteq \mathcal{A}$. Dann gilt:

a) Aus $\mathcal{C}$ $\cap$-stabil (d.h. $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$) folgt

$$\mathcal{D}(\mathcal{C}) = \mathcal{F}(\mathcal{C}).$$

b) Sind $\{\mathcal{C}_1, \mathcal{C}_2\}$ unabhängig, so sind auch

$$\{\mathcal{D}(\mathcal{C}_1), \mathcal{C}_2\} \text{ sowie } \{\mathcal{D}(\mathcal{C}_1), \mathcal{D}(\mathcal{C}_2)\} \text{ unabhängig .}$$

Beweis:

a) Es genügt zu zeigen: (*) Mit $\mathcal{C}$ ist auch $\mathcal{D}(\mathcal{C})$ $\cap$-stabil.

Denn einerseits ist $\mathcal{F}(\mathcal{C})$ ein Dynkin-System, dass $\mathcal{C}$ enthält, was

$$\mathcal{D}(\mathcal{C}) \subset \mathcal{F}(\mathcal{C})$$

impliziert. Und andererseits folgt aus (*) und dem Lemma oben, dass $\mathcal{D}(\mathcal{C})$ eine σ-Algebra ist, die $\mathcal{C}$ enthält, woraus

$$\mathcal{F}(\mathcal{C}) \subset \mathcal{D}(\mathcal{C})$$

folgt.

Nachweis von (*): Sei $\mathcal{C}$ $\cap$-stabil.
Zu zeigen: $E, F \in \mathcal{D}(\mathcal{C}) \Rightarrow E \cap F \in \mathcal{D}(\mathcal{C})$
1. Schritt: Sei $E \in \mathcal{C}$. Gezeigt wird:

$$E \cap F \in \mathcal{D}(\mathcal{C}) \text{ für alle } F \in \mathcal{D}(\mathcal{C}).$$

Setze

$$\mathcal{G}_E := \{\widetilde{F} \in \mathcal{P}(\Omega) : E \cap \widetilde{F} \in \mathcal{D}(\mathcal{C})\}.$$

Behauptung folgt aus

$$\mathcal{D}(\mathcal{C}) \subseteq \mathcal{G}_E \quad \text{(typischer Beweisschritt!)}$$

Nach Definition von $\mathcal{D}(\mathcal{C})$ folgt dies wiederum aus:

(1) $\mathcal{C} \subseteq \mathcal{G}_E$

(2) $\mathcal{G}_E$ ist Dynkin-System.

Nachweis von (1): Sei $\widetilde{F} \in \mathcal{C}$. Da $\mathcal{E}$ $\cap$-stabil und $E \in \mathcal{C}$ nach Annahme oben ist, gilt dann $E \cap \widetilde{F} \in \mathcal{C} \subseteq \mathcal{D}(\mathcal{C}) \Rightarrow \widetilde{F} \in \mathcal{G}_E$

Nachweis von (2):

(1) $\Omega \in \mathcal{G}_E$, da $E \cap \Omega = E \in \mathcal{C} \subseteq \mathcal{D}(\mathcal{C})$

(2) Seien $A, B \in \mathcal{G}_E$ mit $A \subseteq B$.
$E \cap (B \backslash A) = (E \cap B) \backslash (E \cap A) \in \mathcal{D}(\mathcal{C})$ da $E \cap B \in \mathcal{D}(\mathcal{C}), E \cap A \in \mathcal{D}(\mathcal{C}), E \cap A \subseteq E \cap B$ und $\mathcal{D}(\mathcal{C})$ Dynkin-System
$\Rightarrow B \backslash A \in \mathcal{G}_E$.

(3) Seien $A_n \in \mathcal{G}_E$ $(n \in \mathbb{N})$ paarweise disjunkt
$\Rightarrow E \cap A_n \in \mathcal{D}(\mathcal{C})$ (nach Def. $\mathcal{G}_E$) und paarweise disjunkt
$E \cap \left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} (E \cap A_n) \in \mathcal{D}(\mathcal{C})$ (nach Definition Dynkin-System).

$\Rightarrow \sum_{n=1}^{\infty} A_n \in \mathcal{G}_E$ $\rightsquigarrow$ (2).

Bemerkung: Beweis zeigt, dass $\mathcal{G}_E$ auch Dynkin-System ist für $E \in \mathcal{D}(\mathcal{C})$.

2. Schritt: Sei $E \in \mathcal{D}(\mathcal{C})$

Wir zeigen wieder:

$$E \cap F \quad \in \mathcal{D}(\mathcal{C}) \qquad \text{für alle } F \in \mathcal{D}(\mathcal{C})$$

.

Dazu genügt es wieder zu zeigen:

$$\mathcal{D}(\mathcal{C}) \subseteq \mathcal{G}_E := \{\widetilde{F} \in \mathcal{P}(\Omega) : E \cap \widetilde{F} \in \mathcal{D}(\mathcal{C})\}$$

Dies wiederum folgt aus:

(1) $\mathcal{C} \subseteq \mathcal{G}_E$

(2) $\mathcal{G}_E$ Dynkin-System.

Nachweis von (2) erfolgt wie im 1. Schritt.
Nachweis von (1):
Sei $\widetilde{F} \in \mathcal{C}$. Nach Schritt 1 gilt dann

$$\widetilde{F} \cap E \in \mathcal{D}(\mathcal{C})$$

$\Rightarrow \widetilde{F} \in \mathcal{G}_E$ $\rightsquigarrow$ (1) $\rightarrow$ a)

b) Seien $\mathcal{C}_1$ und $\mathcal{C}_2$ unabhängig.
Wir zeigen zuerst: $\{\mathcal{D}(\mathcal{C}_1), \mathcal{C}_2\}$ ist unabhängig.
Äquivalent dazu ist: $\{\{A\}, \mathcal{C}_2\}$ ist unabhängig für alle $A \in \mathcal{D}(\mathcal{C}_1)$.
Um dies nachzuweisen, betrachten wir

$$\mathcal{C}^* = \{A \in \mathcal{A} : \{\{A\}, \mathcal{C}_2\} \text{ unabhängig}\}$$

$\mathcal{C}^*$ ist Dynkin-System, denn es gilt:

- $\Omega \in \mathcal{C}^*$ (klar)
- Sind $A, B \in \mathcal{C}^*$ mit $A \subset B$, so ist auch $B \backslash A \in \mathcal{C}^*$.

  Denn für beliebiges $E \in \mathcal{C}_2$ gilt:

$$\begin{aligned}
& P((B \backslash A) \cap E) = P((B \cap E) \backslash (A \cap E)) \\
& = P(B \cap E) - P(A \cap E) \\
& \quad (\text{da } A \cap E \subset B \cap E) \\
& = P(B) \cdot P(E) - P(A) \cdot P(E) \\
& \quad (\text{da } A, B \in \mathcal{C}^*) \\
& = (P(B) - P(A)) \cdot P(E) \\
& = P(B \backslash A) \cdot P(E) \\
& \quad (\text{da } A \subset B)
\end{aligned}$$

- Sind $A_n \in \mathcal{C}^*$ $(n \in \mathbb{N})$ paarweise disjunkt, so ist auch $\sum\limits_{n=1}^{\infty} A_n \in \mathcal{C}^*$.

Denn für beliebiges $E \in \mathcal{C}_2$ gilt:

$$\begin{aligned}
&P\left(\left(\sum_{n\in\mathbb{N}} A_n\right) \cap E\right) \\
&= P\left(\sum_{n\in\mathbb{N}} A_n \cap E\right) \\
&= \sum_{n=1}^{\infty} P(A_n \cap E) \\
&= \sum_{n=1}^{\infty} P(A_n) \cdot P(E) \\
&\quad \text{(da } A_n \in \mathcal{C}^*\text{)} \\
&= \left(\sum_{n=1}^{\infty} P(A_n)\right) \cdot P(E) \\
&= P\left(\sum_{n\in\mathbb{N}} A_n\right) \cdot P(E).
\end{aligned}$$

Nach Voraussetzung gilt auch $\mathcal{C}_1 \subset \mathcal{C}^*$, und daraus folgt (nach Definition des erzeugten Dynkin-Systems):

$$\mathcal{D}(\mathcal{C}_1) \subset \mathcal{D}(\mathcal{C}^*) = \mathcal{C}^*.$$

Also ist

$$\{\mathcal{D}(\mathcal{C}_1), \mathcal{C}_2\}$$

unabhängig, und nochmaliges Anwenden des gleichen Beweisschrittes ergibt

$$\{\mathcal{D}(\mathcal{C}_1), \mathcal{D}(\mathcal{C}_2)\}$$

unabhängig. □

Beweis von Satz 4.19:

Die Behauptung folgt aus

$(*)$ $\quad \{\mathcal{T}_\infty, \mathcal{T}_\infty\}$ $\quad$ unabhängig,

denn dann gilt für jedes $A \in \mathcal{T}_\infty$: $\quad \{A, A\}$ unabhängig, also

$$P(A) = P(A \cap A) \underset{\text{Unabhängigkeit}}{=} P(A) \cdot P(A),$$

woraus folgt

$$0 = P(A) - P(A) \cdot P(A) = P(A) \cdot (1 - P(A)),$$

was

$$P(A) = 0 \text{ oder } P(A) = 1$$

impliziert.

Zum Nachweis von $(*)$ setzen wir

$$\begin{aligned}
\mathcal{C}_n &= \mathcal{F}(\mathcal{A}_1, \ldots, \mathcal{A}_n) = \mathcal{F}\left(\bigcup_{k=1}^{n} \mathcal{A}_k\right) \\
\mathcal{T}_{n+1} &= \mathcal{F}(\mathcal{A}_{n+1}, \mathcal{A}_{n+2,\ldots}) = \mathcal{F}\left(\bigcup_{k=n+1}^{\infty} \mathcal{A}_k)\right)
\end{aligned}$$

und beachten

$$\mathcal{C}_n = \mathcal{F}(\mathcal{C})$$

mit

$$\mathcal{C} = \{\bigcap_{i=1}^{n} A_i : A_i \in \mathcal{A}_i\} \supseteq \cup_{k=1}^{n} \mathcal{A}_k,$$

sowie

$$\mathcal{T}_{n+1} = \mathcal{F}(\mathcal{T})$$

mit

$$\begin{aligned} \mathcal{T} &:= \{\bigcap_{i \in I} A_i : A_i \in \mathcal{A}_i \ (i \in I) \text{ und } I \text{ endliche} \\ &\quad\ \text{nichtleere Teilmenge von } \{n+1, n+2, \ldots\}\} \\ &\supseteq \cup_{k=n+1}^{\infty} \mathcal{A}_k \end{aligned}$$

Da $(\mathcal{A}_n)_{n \in \mathbb{N}}$ unabhängig sind, sind auch $\{\mathcal{C}, \mathcal{T}\}$ unabhängig.
Mit Lemma 4.22 folgt

$$\{\mathcal{D}(\mathcal{C}), \mathcal{D}(\mathcal{T})\}$$

unabhängig, und da $\mathcal{C}$ und $\mathcal{T}$ beide $\cap$-stabil sind auch

$$\{\mathcal{F}(\mathcal{C}), \mathcal{F}(\mathcal{T})\}$$

unabhängig, also

$$\{\mathcal{C}_n, \mathcal{T}_{n+1}\}$$

unabhängig.
Mit $\quad \mathcal{T}_\infty = \bigcap_{k=1}^{\infty} \mathcal{T}_k \subseteq \mathcal{T}_{n+1} \quad$ folgt daraus aber

$$\{\mathcal{C}_n, \mathcal{T}_\infty\}$$

unabhängig für alle $n \in \mathbb{N}$.
Dies impliziert

$$\left\{\bigcup_{n \in \mathbb{N}} \mathcal{C}_n, \mathcal{T}_\infty\right\}$$

unabhängig. Erneute Anwendung von Lemma 4.22 liefert

$$\left\{\mathcal{D}\left(\bigcup_{n \in \mathbb{N}} \mathcal{C}_n\right), \mathcal{T}_\infty\right\}$$

unabhängig.
Wegen $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \mathcal{C}_3 \subseteq \ldots$ und $\mathcal{C}_n$ $\sigma-$Algebra ist $\bigcup_{n \in \mathbb{N}} \mathcal{C}_n$ $\cap-$stabil, also folgt

$$\begin{aligned} \mathcal{D}\left(\bigcup_{n \in \mathbb{N}} \mathcal{C}_n\right) &= \mathcal{F}\left(\bigcup_{n \in \mathbb{N}} \mathcal{C}_n\right) = \mathcal{F}\left(\bigcup_{k=1}^{\infty} \mathcal{A}_k\right) \\ &= \mathcal{T}_1 \supseteq \mathcal{T}_\infty, \end{aligned}$$

was

$$\{\mathcal{T}_\infty, \mathcal{T}_\infty\}$$

unabhängig impliziert. $\square$

5 Erzeugende und charakteristische Funktionen; Faltungen

Die erzeugenden und charakteristischen Funktionen sind ein methodisches Hilfsmittel zur Untersuchung von W-Maßen.

5.1 Erzeugende Funktionen

5.1 Definition. Es sei $\mu : \mathcal{B} \to \mathbb{R}$ ein auf $\mathbb{N}_0$ konzentriertes W-Maß mit Zähldichte $(b_k)_{k\in\mathbb{N}_0}$ bzw. X eine reelle ZV auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, deren Verteilung P_X auf $\mathbb{N}_0$ konzentriert ist, mit Zähldichte $(b_k)_{k\in\mathbb{N}_0}$. Dann heißt die auf $[0,1]$ (oder auch $[-1,1]$) definierte Funktion g mit

$$g(s) := \sum_{k=0}^{\infty} b_k\, s^k = \begin{cases} \sum_{k=0}^{\infty} \mu(\{k\}) s^k \\ \text{bzw.} \\ \sum_{k=0}^{\infty} P[X=k] s^k = \sum_{k=0}^{\infty} P_X(\{k\}) s^k = E s^X \end{cases}$$

die **erzeugende Funktion** von μ bzw. X.

5.2 Bemerkung. In Definition 5.1 gilt $b_k \in [0,1]$ $(k \in \mathbb{N}_0)$ und $\sum_{k=0}^{\infty} b_k = 1$, weshalb die Reihe nach dem Majorantenkriterium konvergiert. Weiter gilt

$$|g(s)| \le \sum_{k=0}^{\infty} b_k |s|^k \le \sum_{k=0}^{\infty} b_k = 1 \quad (|s| \le 1).$$

5.3 Bemerkung.

a) Die Binomialverteilung $b(n,p)$ mit $n \in \mathbb{N}$, $p \in [0,1]$ hat die erzeugende Funktion g mit

$$g(s) = \sum_{k=0}^{n} \binom{n}{k} \cdot p^k (1-p)^{n-k} \cdot s^k = \sum_{k=0}^{n} \binom{n}{k} \cdot (p \cdot s)^k (1-p)^{n-k} = (1-p+p\cdot s)^n.$$

b) Die Poissonverteilung $\pi(\lambda)$ mit $\lambda \in (0,\infty)$ hat die erzeugende Funktion g mit

$$g(s) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \cdot e^{-\lambda} \cdot s^k = e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{(\lambda \cdot s)^k}{k!} = e^{-\lambda+\lambda\cdot s} \quad \text{für } s \in \mathbb{R}$$

c) Ist X geometrisch verteilt mit Parameter $p \in (0,1)$ (d. h. gilt $P[X=k] = p \cdot (1-p)^k \ \ (k \in \mathbb{N}_0)$), so gilt

$$g(s) = \sum_{k=0}^{\infty} p \cdot (1-p)^k \cdot s^k = p \cdot \sum_{k=0}^{\infty} ((1-p)\cdot s)^k = \frac{p}{1-(1-p)\cdot s}$$

für $s \in [0,1]$.

5.4 Satz (Eindeutigkeitssatz für erzeugende Funktionen).
Besitzen zwei auf $\mathcal{B}$ definierte, aber auf $\mathbb{N}_0$ konzentrierte W-Maße bzw. Verteilungen dieselbe erzeugende Funktion, so stimmen sie überein.

Beweis:
Folgt aus Identitätssatz für Potenzreihen:

$$\sum_{k=0}^{\infty} a_k \cdot x^k = \sum_{k=0}^{\infty} b_k \cdot x^k \; (x \in [0,1]) \Rightarrow a_k = b_k \quad (k \in \mathbb{N}_0).$$

□

5.5 Satz. Sei X eine reelle ZV, deren Verteilung auf $\mathbb{N}_0$ konzentriert ist, mit erzeugender Funktion g.

a) g ist in $(-1,1)$ unendlich oft differenzierbar; für $j \in \mathbb{N}$ gilt

$$g^{(j)}(1-) := \lim_{0 \le s \to 1-0} g^{(j)}(s) = E[X(X-1)\dots(X-j+1)] \quad (\le \infty)\,,$$

insbesondere $g'(1-) = EX$.

b) $EX < \infty \Longrightarrow V(X) = g''(1-) + g'(1-) - (g'(1-))^2$.

Beweis:
Die auf $[-1,1]$ definierte Potenzreihe

$$g(s) = \sum_{k=0}^{\infty} P[X = k] \cdot s^k$$

ist im Inneren $(-1,1)$ des Definitionsbereichs unendlich oft differenzierbar, und es gilt:

$$g^{(j)}(s) = \sum_{k=j}^{\infty} P[X = k] \cdot k \cdot (k-1) \cdot \ldots \cdot (k-j+1) \cdot s^{k-j}.$$

Grenzübergang $\quad s \to 1- \quad$ ergibt

$$\begin{aligned}
\lim_{s \to 1-} g^{(j)}(s) &= \lim_{s \to 1-} \sum_{k=j}^{\infty} P[X = k] \cdot k \cdot (k-1) \cdot \ldots \cdot (k-j+1) \cdot s^{k-j} \\
&= \sum_{k=j}^{\infty} P[X = k] \cdot k \cdot (k-1) \cdot \ldots \cdot (k-j+1) \cdot 1^{k-j}
\end{aligned}$$

(Limes und Reihe sind nach dem Satz von der monotonen Konvergenz vertauschbar. Hierbei fasst man die Reihe als Integral bzgl. dem abzählenden Maß auf)

$$= E[X \cdot (X-1) \cdot \ldots \cdot (X-j+1)]$$

(nach Satz 3.13).

b)

$$\begin{aligned} \text{rechte Seite} &\overset{a)}{=} E[X\cdot(X-1)]+EX-(EX)^2 \\ &= E[X^2]-(EX)^2 \\ &= V(X). \end{aligned}$$

□

5.6 Satz. Seien $X_1,\ldots,X_n$ unabhängige reelle ZVn, deren Verteilungen jeweils auf $\mathbb{N}_0$ konzentriert sind, mit erzeugenden Funktionen $g_1,\ldots,g_n$. Für die erzeugende Funktion g der Summe $X_1+\ldots+X_n$ gilt dann $g=\prod_{j=1}^{n} g_j$.

Beweis:

$$\begin{aligned} g(s) &= E\left(s^{X_1+\cdots+X_n}\right) = E\left(\prod_{j=1}^{n} s^{X_j}\right) \\ &= \prod_{j=1}^{n} E\left(s^{X_j}\right) \\ &\quad \text{(da mit } X_1,\ldots,X_n \text{ auch } s^{X_1},\ldots,s^{X_n} \text{ unabhängig sind, und Satz 4.10 gilt)} \\ &= \prod_{j=1}^{n} g_j(s). \end{aligned}$$

□

Bemerkung:
Zufallsvariablen $X_1, X_2\ldots$ heißen identisch verteilt, falls gilt:

$$P_{X_1}=P_{X_2}=\ldots$$

(also falls die Verteilungen der ZVen übereinstimmen).
Beispiel: Die ZVen $X_1,X_2,\ldots$ seien unabhängig und identisch verteilt mit

$$P[X_1=1]=p,\quad P[X_1=0]=1-p$$

für ein $p\in(0,1)$ fest.
Hierbei ist $[X_k=1]$ das Ereignis, dass im k-ten Bernoulli-Versuch ein Erfolg eintritt.
Die ZV Z_1 (mit Werten in $\mathbb{R}$) gebe die Anzahl der Misserfolge bis zum 1. Erfolg an.
Dann gilt

$$\begin{aligned} P[Z_1=k] &= P[X_1=0,X_2=0,\ldots,X_k=0,X_{k+1}=1] \\ &\overset{Unabh.}{=} P[X_1=0]\cdot P[X_2=0]\cdot\ldots\cdot P[X_k=0]\cdot P[X_{k+1}=1] \\ &= (1-p)^k\cdot p. \end{aligned}$$

Wegen

$$P[Z_1 \in \mathbb{N}_0] = \sum_{k=0}^{\infty} (1-p)^k \cdot p = 1$$

gilt dann $\quad Z_1 < \infty \quad$ f.s.

Nach dem 1. Erfolg tritt die stochastische Situation von neuem auf. Die ZV Z_2 gebe die Anzahl der Misserfolge nach dem 1. Erfolg bis zum 2. Erfolg an. Entsprechend seien $Z_3, Z_4, \ldots$ definiert.

Beh. Die ZVen $Z_1, Z_2, \ldots$ sind unabhängig und identisch verteilt.

Begründung:

$$\begin{aligned}
P[Z_2 = k] &= \sum_{j=0}^{\infty} P[Z_1 = j, Z_2 = k] \\
&= \sum_{j=0}^{\infty} (1-p)^j \cdot p(1-p)^k \cdot p \\
&\quad \text{(folgt analog zur Berechnung von } P[Z_1 = k] \text{ oben)} \\
&= (1-p)^k \cdot p \cdot \sum_{j=0}^{\infty} (1-p)^j \cdot p \\
&= (1-p)^k \cdot p \cdot 1
\end{aligned}$$

Also sind Z_1, Z_2 identisch verteilt.

Weiter gilt

$$P[Z_1 = j, Z_2 = k] = (1-p)^j \cdot p \cdot (1-p)^k \cdot p = P[Z_1 = j] \cdot P[Z_2 = k]$$

für alle $j, k \in \mathbb{N}_0$, woraus die Unabhängigkeit von Z_1 und Z_2 folgt.

(denn

$$\begin{aligned}
P[Z_1 \in A, Z_2 \in B] &= \sum_{\substack{j \in A \cap \mathbb{N}_0 \\ k \in B \cap \mathbb{N}_0}} P[Z_1 = j, Z_2 = k] \\
&\overset{s.o.}{=} \sum_{\substack{j \in A \cap \mathbb{N}_0 \\ k \in B \cap \mathbb{N}_0}} P[Z_1 = j] \cdot P[Z_2 = k] \\
&= \sum_{j \in A \cap \mathbb{N}_0} P[Z_1 = j] \cdot \sum_{k \in B \cap \mathbb{N}_0} P[Z_2 = k] \\
&= P[Z_1 \in A] \cdot P[Z_2 \in B])
\end{aligned}$$

Analog sieht man, dass $Z_1, \ldots, Z_n$ unabhängig und identisch verteilt sind. $\square$

Z_1 hat die erzeugende Funktion

$$h(s) = \sum_{k=0}^{\infty} (1-p)^k \cdot p \cdot s^k = \frac{p}{1-(1-p) \cdot s}$$

Es gilt

$$h'(s) = \frac{p(1-p)}{(1-(1-p) \cdot s)^2} \quad , \quad h''(s) = \frac{2p(1-p)^2}{(1-(1-p) \cdot s)^3} \quad ,$$

woraus mit Satz 5.5 folgt

$$EZ_1 = h'(1) = \frac{p(1-p)}{(1-(1-p))^2} = \frac{1-p}{p}$$

und

$$\begin{aligned} V(Z_1) &= h''(1) + h'(1) - (h'(1))^2 \\ &= \tfrac{2p(1-p)^2}{p^3} + \tfrac{1-p}{p} - \tfrac{(1-p)^2}{p^2} \\ &= \tfrac{(1-p)^2}{p^2} + \tfrac{1-p}{p} = \tfrac{1-p}{p}\left(\tfrac{1-p}{p} + 1\right) \\ &= \tfrac{1-p}{p} \cdot \tfrac{1}{p} = \tfrac{1-p}{p^2}. \end{aligned}$$

P_{Z_1} ist die sogenannte **geometrische Verteilung** mit Parameter $p \in (0,1)$.
Für $r \in \mathbb{N}$ betrachten wir die **Anzahl der Misserfolge bis zum r-ten Erfolg**, gegeben durch

$$X = Z_1 + \ldots + Z_r.$$

Da $Z_1, \ldots, Z_r$ unabhängig und identisch verteilt sind, hat X nach Satz 5.6 die erzeugende Funktion

$$g(s) = (h(s))^r = \frac{p^r}{(1-(1-p)\cdot s)^r}$$

Die Funktion

$$f(x) = \frac{1}{(1-x)^r} = (1-x)^{-r}$$

hat wegen

$$\begin{aligned} f^{(k)}(x) &= (-r)\cdot(-r-1)\cdot\ldots\cdot(-r-k+1)\cdot(1-x)^{-r-k}\cdot(-1)^k \\ &= (r+k-1)\cdot(r+k-2)\cdot\ldots\cdot r\cdot(1-x)^{-r-k} \end{aligned}$$

die Potenzreihendarstellung

$$\frac{1}{(1-x)^r} = \sum_{k=0}^{\infty} \binom{r+k-1}{k} \cdot x^k.$$

Mit dieser erhält man für g:

$$g(s) = p^r \sum_{k=0}^{\infty} \binom{r+k-1}{k} (1-p)^k \cdot s^k,$$

also gilt (nach dem Eindeutigkeitssatz für erzeugende Funktionen, da

$$g(s) = \sum_{k=0}^{\infty} \mathbf{P}[X=k]\cdot s^k$$

unter Beachtung von $X < \infty$ $f.s.$)

$$\underline{\underline{P[X=k] = \binom{r+k-1}{k} \cdot p^r \cdot (1-p)^k}} \qquad (k \in \mathbb{N}_0).$$

P_X ist eine sogenannte **negative Binomialverteilung** (oder auch Pascalverteilung) mit Bezeichnung $Nb(r,p)$.

Die Zähldichte von $Nb(r,p)$ ist

$$k \mapsto \binom{n+k-1}{k} \cdot p^r(1-p)^k$$

und wegen $X = Z_1 + \ldots + Z_r$ mit $Z_1, \ldots, Z_r$ unabhängig identisch verteilt (nach obiger Herleitung) gilt

$$\begin{aligned} EX &= r \cdot EZ_1 = r \cdot \tfrac{1-p}{p} \\ V(X) &= r \cdot V(Z_1) = r \cdot \tfrac{1-p}{p^2}. \end{aligned}$$

5.2 Charakteristische Funktionen

5.7 Definition. Es sei μ ein W-Maß auf $\mathcal{B}$ bzw. X eine reelle ZV mit VF F. Dann heißt die auf $-\infty < u < \infty$ definierte (i.a. komplexwertige) Funktion φ mit

$$\varphi(u) := \int_{\mathbb{R}} e^{iux}\,\mu(dx) = \int_{\mathbb{R}} \cos(ux)\,\mu(dx) + i \cdot \int_{\mathbb{R}} \sin(ux)\,\mu(dx)$$

bzw.

$$\varphi(u) := Ee^{iuX} = \int_{\mathbb{R}} e^{iux}\,P_X(dx) = \int_{\mathbb{R}} \cos(ux)\,P_X(dx) + i \cdot \int_{\mathbb{R}} \sin(ux)\,P_X(dx)$$

die **charakteristische** Funktion von μ bzw. X bzw. F.

5.8 Bemerkung. In Definition 5.7 gilt

a) $\varphi(0) = 1$,

b) $|\,\varphi(u)\,| \leq 1$, $u \in \mathbb{R}$,

c) φ gleichmäßig stetig in $\mathbb{R}$,

d) $\varphi(-u) = \overline{\varphi(u)}$, $u \in \mathbb{R}$.

Begründung: a), b), d) ... trivial.
c)

$$\begin{aligned} |\varphi(u+h) - \varphi(u)| &= |\textstyle\int_{\mathbb{R}} \left(e^{i(u+h)\cdot x} - e^{iux}\right) P_X(dx)| \\ &= |\textstyle\int_{\mathbb{R}} e^{iux} \cdot \left(e^{ihx} - 1\right) P_X(dx)| \\ &\leq \textstyle\int_{\mathbb{R}} |e^{ihx} - 1| P_X(dx) =: f(h) \end{aligned}$$

da

$$|e^{iux}| = 1.$$

Da $f(h)$ nicht von u abhängt, folgt die Behauptung aus

$$f(h) \to 0 \quad \text{für } h \to 0.$$

Dies wiederum gilt nach dem Satz von der majorisierten Konvergenz, da

- $|e^{ihx} - 1| \to |e^0 - 1| = 0$ $(h \to 0)$ für alle $x \in \mathbb{R}$

- $|e^{ihx} - 1| \leq |e^{ihx}| + 1 = 2$ und 2 ist integrierbar.

Also gilt

$$f(h) \to \int_{\mathbb{R}} |e^{i \cdot 0 \cdot x} - 1| P_X(dx) = \int_{\mathbb{R}} 0 P_X(dx) = 0$$

für $h \to 0$, w.z.z.w. □

5.9 Bemerkung. Die Normalverteilung $N(a, \sigma^2)$ mit $a \in \mathbb{R}$, $\sigma^2 \in (0, \infty)$ hat die charakteristische Funktion φ mit

$$\varphi(u) = e^{iau} e^{-\frac{\sigma^2 u^2}{2}}.$$

Insbesondere hat $N(0,1)$ die charakteristische Funktion φ mit $\varphi(u) = e^{-\frac{u^2}{2}}$.

Beweis: Ist X $\mathcal{N}(0,1)$-verteilt, so ist (vgl. Übungen) $Y = \sigma \cdot X + a$ $\mathcal{N}(a, \sigma^2)$-verteilt.
Es gilt:

$$\begin{aligned} \varphi_Y(u) &= E\, e^{iu(\sigma \cdot X + a)} = e^{iua} \cdot E\, e^{i(\sigma \cdot u) \cdot X} \\ &= e^{i \cdot u \cdot a} \cdot \varphi_X(\sigma \cdot u), \end{aligned}$$

also genügt es, die charakteristische Funktion zu $\mathcal{N}(0,1)$ zu bestimmen.
Hierzu:

$$\begin{aligned} \varphi_X(u) &= \int_{\mathbb{R}} e^{iux} \cdot \tfrac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} dx \\ &= \tfrac{1}{\sqrt{2\pi}} \cdot \int_{\mathbb{R}} e^{-\frac{(x-iu)^2}{2}} dx \cdot e^{-u^2/2} \end{aligned}$$

Zu zeigen:

$$\int_{\mathbb{R}} e^{-\frac{(x-iu)^2}{2}} dx = \sqrt{2\pi}.$$

Dazu betrachten wir für $u \in \mathbb{R}$ und $t > 0$:

$$\int_{-t}^{t} e^{-\frac{(x - i \cdot u)^2}{2}} dx = \int_{\mathcal{C}_1} e^{-z^2/2} dz$$

wobei das komplexe Kurvenintegral rechts über den Streckenzug $\mathcal{C}_1$ von $-t - i \cdot u$ bis $t - i \cdot u$ berechnet wird.
Nach dem Cauchyschen Integralsatz gilt

$$\sum_{i=1}^{4} \int_{\mathcal{C}_i} e^{-z^2/2} dz = 0,$$

wobei $\mathcal{C}_2, \mathcal{C}_3$ bzw. $\mathcal{C}_4$ die Streckenzüge von $t - i \cdot u$ bis t, bzw. t bis $-t$, bzw. $-t$ bis $-t - i \cdot u$ sind.

Auf $\mathcal{C}_2$ bzw $\mathcal{C}_4$ ist der Integrand betragsmäßig beschränkt durch

$$\begin{aligned}
&\sup_{0\le|v|\le|u|}\left|e^{-\frac{(\pm t-i\cdot v)^2}{2}}\right| \\
&= \sup_{0\le|v|\le|u|} e^{-t^2/2}\cdot\underbrace{|e^{\pm i\cdot t\cdot v|}}_{=1}\cdot e^{v^2/2} \\
&\le e^{-t^2/2}\cdot e^{u^2/2}.
\end{aligned}$$

Damit gilt für $i\in\{2,4\}$:

$$\left|\int_{\mathcal{C}_i} e^{-z^2/2}dz\right| \le e^{-t^2/2}\cdot e^{u^2/2}\cdot|u|\to 0 \text{ für } t\to\infty.$$

Also gilt:

$$\lim_{t\to\infty}\int_{\mathcal{C}_1} e^{-z^2/2}dz = \lim_{t\to\infty} -\int_{\mathcal{C}_3} e^{-z^2/2}dz = \int_{-\infty}^{\infty} e^{-\frac{s^2}{2}}ds = \sqrt{2\pi}$$

□

5.10 Bemerkung Sei X $\exp(\lambda)$-verteilt, d.h. X habe die Dichte

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

mit $\lambda>0$.
X hat die charakteristische Funktion

$$\varphi(u)=\int_0^\infty e^{iux}\cdot\lambda\cdot e^{-\lambda x}dx = \frac{\lambda}{iu-\lambda}\cdot e^{(iu-\lambda)\cdot x}\Big|_{x=0}^{\infty} = \frac{\lambda}{\lambda-i\cdot u}.$$

Die Exponentialverteilung ist **gedächtnislos** in dem Sinne, dass gilt:

$$(*)\quad P[X>t+s|X>s]=P[X>t]\qquad \text{für alle } s,t>0,$$

denn

$$\begin{aligned}
\text{linke Seite } &= \frac{P[X>t+s]}{P[X>s]} = \frac{\int_{t+s}^{\infty}\lambda\cdot e^{-\lambda x}dx}{\int_s^\infty \lambda\cdot e^{-\lambda x}dx} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda\cdot s}} \\
&= e^{-\lambda t} = \text{ rechte Seite }.
\end{aligned}$$

Ist umgekehrt X eine $\mathbb{R}$-wertige ZV mit $P[X>0]>0$, die die Eigenschaft $(*)$ hat, so ist X $\exp(\lambda)$-verteilt.
Begründung:

Setze
$$S(x) = 1 - F(x) = 1 - P[X \leq x] = P[X > x].$$
Aus $(*)$ folgt dann
$$S(t+s) = S(t) \cdot S(s) \quad \text{für alle } s, t > 0,$$
woraus wiederum folgt

- $S(k \cdot t) = (S(t))^k$ für alle $k \in \mathbb{N}_0, t \geq 0$
- $S\left(\frac{1}{m}\right) = S(1)^{1/m}$ für alle $m \in \mathbb{N}$ (folgt aus oben mit $t = 1/m, k = m$)
- $S\left(\frac{k}{m}\right) = (S(1))^{\frac{k}{m}}$ für alle $k \in \mathbb{N}_0, m \in \mathbb{N}$ (folgt aus den obigen zwei Beziehungen)

Der Fall $S(1) = 1$ kann nicht eintreten, da dies $S(k) = 1$ für alle $k \in \mathbb{N}$ implizieren würde, im Widerspruch zu
$$S(k) = P[X > k] \to 0 \quad (k \to \infty).$$
Ebenso kann der Fall $S(1) = 0$ nicht eintreten, da daraus $S\left(\frac{1}{m}\right) = 0$ für alle $m \in \mathbb{N}$ folgen würde, im Widerspruch zu
$$0 < P[X > 0] = \lim_{m \to \infty} P[X > \frac{1}{m}] = \lim_{m \to \infty} S\left(\frac{1}{m}\right).$$
Also gilt $\quad S(1) \in (0,1)$.
Mit $\lambda = -\ln(S(1)) \in (0, \infty)$ folgt
$$P[X > \frac{k}{m}] = S\left(\frac{k}{m}\right) = (S(1))^{\frac{k}{m}} = e^{-\lambda \cdot \frac{k}{m}}$$
für alle $k \in \mathbb{N}_0, m \in \mathbb{N}$.
Damit gilt für $x \in \mathbb{Q}_+$
$$P[X > x] = e^{-\lambda x}$$
bzw.
$$F(x) = P[X \leq x] = 1 - e^{-\lambda x}.$$
Da F als Verteilungsfunktion rechtsseitig stetig ist, folgt daraus
$$F(x) = 1 - e^{-\lambda x} \text{ für alle } x \in \mathbb{R}_+ \qquad \rightsquigarrow \text{ Beh.}$$

$\square$

Aufgrund der in $(*)$ formulierten **Gedächtnislosigkeit** wird die Exponentialverteilung häufig bei der Modellierung von Wartesystemen eingesetzt.
Es gilt:
$$EX = \int_0^\infty x \cdot \lambda \cdot e^{-\lambda x} dx = \ldots = \tfrac{1}{\lambda}$$
$$V(X) = E(X^2) - (EX)^2 = \ldots = \tfrac{1}{\lambda^2}.$$

5.11 Satz (Eindeutigkeitssatz für charakteristische Funktionen).
a) Ist μ ein W-Maß auf $\mathcal{B}$ bzw. P_X die Verteilung einer reellen ZV X mit zugehöriger VF F und zugehöriger charakteristischer Funktion φ und sind a, b mit $a < b$ Stetigkeitsstellen von F, so gilt die Umkehrformel

$$\underbrace{F(b) - F(a)}_{= \mu((a,b]) \text{ bzw. } P_X((a,b])} = \lim_{U\to\infty} \frac{1}{2\pi} \int_{-U}^{U} \frac{e^{-iua} - e^{-iub}}{iu} \varphi(u)\, du\,.$$

b) Besitzen zwei auf $\mathcal{B}$ definierte W-Maße bzw. Verteilungen dieselbe charakteristische Funktion, so stimmen sie überein.

Beweis: Sind X, Y reelle ZVen mit Vf. F_X, F_Y und gilt für die charakteristischen Funktionen φ_X und φ_X

$$\varphi_X = \varphi_Y,$$

so folgt aus der Umkehrformel

$$F_X(b) - F_X(a) = F_Y(b) - F_Y(a)$$

für alle $a, b \in \mathbb{R}$, die sowohl Stetigkeitsstellen von F_X als auch von F_Y sind. Da Verteilungsfunktionen überall bis auf höchstens abzählbar viele Stellen stetig sind, gilt diese Beziehung also für $L. - f.a.\ a, b \in \mathbb{R}$. Mit $a \to -\infty$ und $F(a) \to 0$ $(a \to -\infty)$ für jede Verteilungsfunktion F folgt daraus

$$F_X = F_Y \quad \text{L.-f.ü.},$$

und da eine Verteilunsgfunktion als rechtsseitig stetige Funktion durch seine Werte auf einer dichten Teilmenge von $\mathbb{R}$ eindeutig bestimmt ist, impliziert dies $F_X = F_Y$. Also genügt es, die Umkehrformel zu zeigen.
Dazu beachten wir für Stetigkeitsstellen a und b von F:

$$\begin{aligned}
&\tfrac{1}{2\pi} \int\limits_{-U}^{U} \tfrac{e^{-iua}-e^{-iub}}{iu} \cdot \varphi(u)du \\
&= \tfrac{1}{2\pi} \int\limits_{-U}^{U} \tfrac{e^{-iua}-e^{-iub}}{i\cdot u} \int_{\mathbb{R}} e^{iux} P_X(dx)du \\
&= \int\limits_{\mathbb{R}} \tfrac{1}{2\pi} \int\limits_{-U}^{U} \tfrac{e^{iu(x-a)}-e^{iu(x-b)}}{iu} du\; P_X(dx) \\
&\qquad \text{(nach dem Satz von Fubini, s.u.)}
\end{aligned}$$

Bei der Anwendung des Satzes von Fubini haben wir dabei benutzt, dass

$$\frac{e^{iu(x-a)} - e^{iu(x-b)}}{iu} = \frac{\cos(u(x-a)) - \cos(u\cdot(x-b))}{iu} + \frac{\sin(u\cdot(x-a)) - \sin(u\cdot(x-b))}{u}$$

nach dem Mittelwertsatz betragsmäßig beschränkt ist und daher bzgl. dem beschränkten (!) Maß $\lambda\big|_{[-U,U]} \otimes \mathcal{P}_X$ integrierbar ist.

Daraus folgt:

$$
\begin{aligned}
&\tfrac{1}{2\pi}\int\limits_{-U}^{U}\tfrac{e^{-iua}-e^{-iub}}{iu}\cdot\varphi(u)du\\
&=\int\limits_{\mathbb{R}}\tfrac{1}{2\pi}\int\limits_{-U}^{U}\Big(\underbrace{\frac{\cos(u(x-a))-\cos(u\cdot(x-b))}{iu}}_{\substack{\text{ungerade als}\\ \text{Funktion von u}}}+\underbrace{\frac{\sin(u\cdot(x-a))-\sin(u\cdot(x-b))}{u}}_{\substack{\text{gerade als}\\ \text{Funktion von u}}}\Big)duP_X\\
&=\int\limits_{\mathbb{R}}\tfrac{1}{\pi}\Big[\underbrace{\int\limits_{0}^{U}\frac{\sin(u\cdot(x-a))}{u}du}_{\overset{t=u\cdot(x-a)}{=}\int\limits_{0}^{U\cdot(x-a)}\frac{\sin t}{t}dt}-\underbrace{\int\limits_{0}^{U}\frac{\sin(u\cdot(x-b))}{u}du}_{\overset{t=u\cdot(x-b)}{=}\int\limits_{0}^{U\cdot(x-b)}\frac{\sin t}{t}dt}\Big]P_X(dx)\\
&=\int\limits_{\mathbb{R}}\tfrac{1}{\pi}\int\limits_{U\cdot(x-b)}^{U\cdot(x-a)}\tfrac{\sin t}{t}dt\ P_X(dx).
\end{aligned}
$$

Nach der Dirichletschen Formel gilt

$$\frac{1}{\pi}\int\limits_{-\infty}^{\infty}\frac{\sin v}{v}dv=2\cdot\frac{1}{\pi}\int\limits_{0}^{\infty}\frac{\sin v}{v}dv=1.$$

Daraus folgt

$$\lim_{U\to\infty}\frac{1}{\pi}\int\limits_{U\cdot(x-b)}^{U\cdot(x-a)}\frac{\sin t}{t}dt=\begin{cases}1 & ,\ \text{falls}\quad a<x<b\\ 0 & ,\ \text{falls}\quad x<a \text{ oder } x>b\\ \frac{1}{2} & ,\ \text{falls}\quad x=a \text{ oder } x=b,\end{cases}$$

wobei die Integrale links wegen der Oszillation des Sinus und

$$\lim_{t\to 0}\frac{\sin t}{t}=\lim_{t\to 0}\frac{\cos t}{1}=1$$

beschränkt sind.
Mit dem Satz von der majorisierten Konvergenz und $P_X(\{a\})=P_X(\{b\})=0$ (da a und b Stetigkeitsstellen von F sind) folgt:

$$
\begin{aligned}
&\lim_{U\to\infty}\tfrac{1}{2\pi}\int\limits_{-U}^{U}\tfrac{e^{-iua}-e^{-iub}}{iu}\cdot\varphi(u)du=\int\limits_{(a,b)}1\ P_X(dx)\\
&=P_X((a,b))=P_X((a,b])=F(b)-F(a),
\end{aligned}
$$

wobei bei der 3. Gleichheit erneut $P_X(\{b\})=0$ verwendet wurde. □

5.12 Satz. Gegeben sei eine VF F mit charakteristischer Funktion φ; es sei $\int_{-\infty}^{\infty}\mid\varphi(u)\mid du<\infty$. Dann besitzt F eine Dichte f, die stetig und beschränkt ist, und es gilt die Umkehrformel

$$f(x)=\frac{1}{2\pi}\int_{\mathbb{R}}e^{-iux}\varphi(u)\,du,\ \ x\in\mathbb{R}$$

[vgl. die Formel $\varphi(u) = \int_{\mathbb{R}} e^{iux} f(x)\, dx,\ \ u \in \mathbb{R}$, bei existierender Dichte f].
Beweis: Siehe Übungen $\square$

5.13 Satz. Sei X eine reelle ZV mit Verteilung P_X und charakteristischer Funktion φ. Für $j \in \mathbb{N}$ gilt:

$$E \mid X \mid^j < \infty$$
$$\Longrightarrow \varphi \text{ hat eine stetige } j\text{-te Ableitung } \varphi^{(j)} \text{ mit } \varphi^{(j)}(u) = i^j \int_{\mathbb{R}} x^j e^{iux} P_X(dx),\ \ u \in \mathbb{R},$$
$$\text{insbesondere } \varphi^{(j)}(0) = i^j EX^j.$$

Beweis:
Fall j=1

$$\frac{\varphi(u+h) - \varphi(u)}{h} = \int_{\mathbb{R}} \underbrace{e^{iux} \cdot \left(\frac{e^{ixh} - 1}{h}\right)}_{\overset{(h\to 0)}{\to} e^{iux} \cdot i \cdot x} P_X(dx).$$

Dabei ist der Integrand betragsmäßig beschränkt durch

$$\begin{aligned}
&\left|\frac{e^{ihx}-1}{h}\right| = \left|\frac{\cos(hx)-1}{h} + i \cdot \frac{\sin(hx)-0}{h}\right| \\
&\leq \left|\frac{\cos(hx)-1}{h}\right| + \left|\frac{\sin(hx)-0}{h}\right| \leq |x| \cdot |-\sin\xi| + |x| \cdot |\cos\overline{\xi}| \\
&\leq 2 \cdot |x|,
\end{aligned}$$

wobei die zweite Ungleichung aus dem Mittelwertsatz folgt. Die Funktion $x \mapsto 2 \cdot |x|$ ist wegen $E|X| < \infty$ bzgl P_X integrierbar. Mit dem Satz von der majorisierten Konvergenz folgt daraus:

$$\varphi'(u) = \lim_{h\to 0} \frac{\varphi(u+h) - \varphi(u)}{h} = \int_{\mathbb{R}} e^{iux} \cdot ix\ P_X(dx).$$

φ' ist gleichmäßig stetig, da

$$\begin{aligned}
|\varphi'(u+h) - \varphi'(u)| &= \left|\int_{\mathbb{R}} ix \cdot \left(e^{i(u+h)\cdot x} - e^{iux}\right)\ P_X(dx)\right| \\
&\leq \int_{\mathbb{R}} |x| \cdot \left|e^{ihx} - 1\right|\ P_X(dx) \\
&\to \int_{\mathbb{R}} 0 P_X(dx) = 0,
\end{aligned}$$

wobei das Integral

$$\int_{\mathbb{R}} |x| \cdot \left|e^{ihx} - 1\right|\ P_X(dx)$$

unabhängig von u ist und für $h \to 0$ wegen

$$|x| \cdot \left|e^{ihx} - 1\right| \to 0 \ \ (h \to 0)$$
$$|x| \cdot \left|e^{ihx} - 1\right| \le 2|x|$$
$$\text{und} \qquad \int 2|x| \ P_X(dx) = 2E|X| < \infty,$$

nach dem Satz von der majorisierten Konvergenz gegen Null konvergiert.
Der allgemeine Fall folgt daraus mit Induktion. □

5.14 Satz. Seien $X_1, \ldots, X_n$ unabhängige reelle ZVn mit charakteristischen Funktionen $\varphi_1, \ldots, \varphi_n$. Für die charakteristische Funktion der Summe $X_1 + \ldots + X_n$ gilt dann

$$\varphi = \prod_{j=1}^{n} \varphi_j.$$

Beweis.

$$\begin{aligned}\varphi(u) &= E\ e^{iu(X_1+\ldots+X_n)} = E\ (e^{iuX_1} \cdot e^{iuX_2} \cdot \ldots \cdot e^{iuX_n}) \\ &= E\ e^{iuX_1} \cdot E\ e^{iuX_2} \cdot \ldots \cdot E\ e^{iuX_n} \\ &\qquad \text{(da mit } X_1, \ldots, X_n \text{ auch } e^{iuX_1}, \ldots, e^{iuX_n} \\ &\qquad \text{unabhängig sind und da Satz 4.10 gilt).} \\ &= \varphi_{X_1}(u) \cdot \varphi_{X_2}(u) \cdot \ldots \cdot \varphi_{X_n}(u).\end{aligned}$$

□

Beispiel: Eine Zufallsvariable mit Dichte

$$f(x) = \begin{cases} \frac{\lambda^\nu}{\Gamma(\nu)} \cdot x^{\nu-1} \cdot e^{-\lambda x} & , x > 0 \\ 0 & x \le 0 \end{cases}$$

(bzgl. des LB-Maßes) heißt **Gamma-verteilt** mit Parameter $\lambda, \nu > 0$ (kurz: $\Gamma_{\lambda,\nu}$-verteilt).
Hierbei ist

$$\Gamma(\nu) = \int_0^\infty x^{\nu-1} \cdot e^{-x} dx$$

die sogenannte Gamma-Funktion, für die gilt:

$$\begin{aligned}\Gamma(\nu+1) &= \nu \cdot \Gamma(\nu), \\ \Gamma(n) &= (n-1)! \qquad (n \in \mathbb{N}) \\ \text{sowie} \quad \Gamma\left(\tfrac{1}{2}\right) &= \sqrt{\pi}.\end{aligned}$$

Spezialfälle sind:

$\Gamma_{\lambda,1} = \exp(\lambda)$

$\Gamma_{\frac{1}{2},\frac{n}{2}} \ldots$ sogenannte Chi-Quadrat-Verteilung χ^2_n mit n Freiheitsgraden.

Für die charakteristische Funktion φ von $\Gamma_{\lambda,\nu}$ gilt:

$$\varphi(u) = \left(\frac{\lambda}{\lambda - i \cdot u}\right)^{\nu} \qquad (u \in \mathbb{R}).$$

Begründung:

$$\begin{aligned}
\varphi(u) &= \int_0^\infty e^{iux} \tfrac{\lambda^\nu}{\Gamma(\nu)} \cdot x^{\nu-1} \cdot e^{-\lambda x} dx \\
&= \tfrac{\lambda^\nu}{(\lambda - i\cdot u)^\nu} \tfrac{1}{\Gamma(\nu)} \cdot \int_0^\infty ((\lambda - i \cdot u) \cdot x)^{\nu-1} \cdot e^{-(\lambda - i\cdot u)\cdot x} \cdot (\lambda - i \cdot u) dx \\
&= \tfrac{\lambda^\nu}{(\lambda - iu)^\nu} \tfrac{1}{\Gamma(\nu)} \cdot \lim_{t\to\infty} \int_{\mathcal{C}_1} z^{\nu-1} \cdot e^{-z} dz
\end{aligned}$$

wobei $\mathcal{C}_1$ der (komplexe) Streckenzug von 0 bis $(\lambda - iu) \cdot t$ ist. Ist $\mathcal{C}_2$ der (komplexe) Streckenzug von $(\lambda - iu) \cdot t$ bis λt und $\mathcal{C}_3$ der von λt bis 0, so gilt:

$$\sum_{i=1}^{3} \int_{\mathcal{C}_i} z^{\nu-1} e^{-z} dz = 0,$$

und mit

$$\lim_{t\to\infty} \int_{\mathcal{C}_2} z^{\nu-1} \cdot e^{-z} dz = 0 \quad (!)$$

folgt

$$\begin{aligned}
\varphi(u) &= \tfrac{\lambda^\nu}{(\lambda - iu)^\nu} \cdot \tfrac{1}{\Gamma(\nu)} \cdot \lim_{t\to\infty} \left(- \int_{\mathcal{C}_3} z^{\nu-1} e^{-z} dz \right) \\
&= \tfrac{\lambda^\nu}{(\lambda - iu)^\nu} \cdot \tfrac{1}{\Gamma(\nu)} \int_0^\infty x^{\nu-1} \cdot e^{-x} dx \\
&= \left(\tfrac{\lambda}{\lambda - iu}\right)^\nu .
\end{aligned}$$

□

Daraus folgt

- $\varphi'(u) = \lambda^\nu \cdot (-\nu) \cdot (\lambda - iu)^{-(\nu+1)} \cdot (-i)$,
  und damit gilt nach Satz 5.13
  $$EX = \frac{\varphi'(0)}{i} = \frac{\lambda^\nu \cdot (-\nu) \cdot \lambda^{-(\nu+1)} \cdot (-i)}{i} = \frac{\nu}{\lambda}.$$
- $$\begin{aligned}
\varphi''(u) &= \lambda^\nu \cdot \nu i \cdot (-(\nu+1)) \cdot (\lambda - iu)^{-(\nu+2)} \cdot (-i) \\
&= -\nu(\nu+1)\lambda^\nu(\lambda - iu)^{-(\nu+2)}
\end{aligned}$$
  und mit Satz 5.13 folgt
  $$E(X^2) = \frac{\varphi''(0)}{i^2} = \frac{\nu(\nu+1)}{\lambda^2},$$
  also gilt
  $$V(X) = E(X^2) - (EX)^2 = \frac{\nu^2 + \nu}{\lambda^2} - \frac{\nu^2}{\lambda^2} = \frac{\nu}{\lambda^2}.$$

Darüber hinaus gilt:
Die Summe von n unabhängigen $\exp(\lambda)$-verteilten Zufallsvariablen ist $\Gamma_{\lambda,n}$ verteilt.
Begründung:
Sind $X_1, \ldots, X_n$ unabhängig und $\exp(\lambda)$-verteilt, so hat nach Satz 5.14 $X_1 + \ldots + X_n$ die charakteristische Funktion

$$\varphi(u) = \frac{\lambda}{\lambda - iu} \cdot \frac{\lambda}{\lambda - iu} \cdot \ldots \cdot \frac{\lambda}{\lambda - iu} = \frac{\lambda^n}{(\lambda - iu)^n},$$

und ist daher nach Satz 5.11 $\Gamma_{\lambda,n}$-verteilt. □

5.15 Definition. Zu zwei W-Maßen P, Q auf $\mathcal{B}$ wird das W-Maß $P * Q$ auf $\mathcal{B}$, die sog. **Faltung** von P und Q, folgendermaßen definiert:

a) sei $T : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ mit $T(x, y) = x + y$;

$$P * Q := (P \otimes Q)_T$$

($P \otimes Q$ W-Maß auf $\mathcal{B}_2$ mit $(P \otimes Q)(B_1 \times B_2) = P(B_1) \cdot Q(B_2)$, $B_1, B_2 \in \mathcal{B}$, sog. Produkt-W-Maß von P und Q)

oder — äquivalent —

b) man wähle ein unabhängiges Paar reeller ZVn X, Y auf $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$ mit $\widetilde{P}_X = P$, $\widetilde{P}_Y = Q$ und setze

$$P * Q := \widetilde{P}_{X+Y}.$$

Beweis der Äquivalenz von a) und b):
Für X, Y wie in b) und $Z = (X, Y)$ gilt:

$$\begin{aligned}
\widetilde{P}_{X+Y} &= \widetilde{P}_{T \circ Z} = \left(\widetilde{P}_Z\right)_T = \left(\widetilde{P}_{(X,Y)}\right)_T \\
&\overset{X,\,Y \text{ unabhängig}}{=} \left(\widetilde{P}_X \otimes \widetilde{P}_Y\right)_T \\
&= (P \otimes Q)_T
\end{aligned}$$

da $\widetilde{P}_X = P, \widetilde{P}_Y = Q$. □

5.16 Satz.

a) Die Faltungsoperation ist kommutativ und assoziativ.

b) Es sei (X, Y) ein unabhängiges Paar reeller ZVn. Für X, Y, $X + Y$ seien die Verteilungsfunktionen mit F, G, H, die Dichten — falls existent — mit f, g, h und die

Zähldichten — falls P_X, P_Y auf $\mathbb{N}_0$ konzentriert sind — mit (p_k), (q_k), (r_k) bezeichnet. Dann gilt

$$H(t) = \int_{\mathbb{R}} F(t-y)\,G(dy) = \int_{\mathbb{R}} G(t-x)\,F(dx), \;\; t \in \mathbb{R}.$$

Falls f existiert, so existiert h mit

$$h(t) = \int_{\mathbb{R}} f(t-y)\,G(dy) \quad \text{für } (L\text{-f.a.}) \; t \in \mathbb{R};$$

falls zusätzlich g existiert, so gilt

$$h(t) = \int_{\mathbb{R}} f(t-y)g(y)\,dy = \int_{\mathbb{R}} g(t-x)f(x)\,dx \;\; \text{für } (L\text{-f.a.}) \; t \in \mathbb{R}.$$

Falls $(p_k)_{k\in\mathbb{N}_0}$, $(q_k)_{k\in\mathbb{N}_0}$ existieren, so existiert $(r_k)_{k\in\mathbb{N}_0}$ mit

$$r_k = \sum_{i=0}^{k} p_{k-i}q_i = \sum_{i=0}^{k} q_{k-i}p_i \;\; (k \in \mathbb{N}_0).$$

Beweis: a) klar nach Definition 5.15 b).

b)

$$H(t) = P[X+Y \le t] = P_{(X,Y)}(B)$$

mit

$$B = \{(x,y) \in \mathbb{R}^2 : x+y \le t\}.$$

Wegen Unabhängigkeit von (X,Y) folgt weiter

$$\begin{aligned}
H(t) &= (P_X \otimes P_Y)(B) \\
&= \int_{\mathbb{R}^2} \chi_B(x,y)d(P_X \otimes P_Y)(x,y) \\
&\overset{\text{Fubini}}{=} \int_{\mathbb{R}}\int_{\mathbb{R}} \chi_B(x,y)dP_X(x)\; dP_Y(y) \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} \chi_{(-\infty,t-y]}(x)\; P_X(dx)\; P_Y(dy) \\
& (\text{da } (x,y) \in B \Leftrightarrow x+y \le t \Leftrightarrow x \le t-y \Leftrightarrow x \in (-\infty, t-y]) \\
&= \int_{\mathbb{R}} F(t-y)\; P_Y(dy)
\end{aligned}$$

Mit $P_{X+Y} = P_{Y+X}$ folgt auch

$$H(t) = \int_{\mathbb{R}} G(t-x)P_X(dx).$$

Existiert Dichte f zu F, so folgt

$$\begin{aligned}
H(t) &= \int\limits_{\mathbb{R}} \int\limits_{-\infty}^{t-y} f(\tau) d\tau \; P_Y(dy) \\
&= \int\limits_{\mathbb{R}} \int\limits_{-\infty}^{t} f(s-y) ds \; P_Y(dy) \\
&\quad \text{(mit Substitution } s = \tau + y) \\
&\overset{\text{Fubini}}{=} \int\limits_{-\infty}^{t} \underbrace{\int\limits_{\mathbb{R}} f(s-y) P_Y(dy)}_{\text{ist als Funktion von } s \text{ Dichte von } H} ds
\end{aligned}$$

Falls zusätzlich Dichte g zu G bzw. P_Y existiert, so gilt für die Dichte h von H:

$$h(t) = \int\limits_{\mathbb{R}} f(s-y) \; P_Y(dy) = \int\limits_{\mathbb{R}} f(t-y) \cdot g(y) \; dy$$

Für Zähldichten geht der Beweis analog, vgl. Übungen □

5.17 Bemerkung.

a) Für $n_1, n_2 \in \mathbb{N},\ p \in [0,1]$ gilt $b(n_1, p) * b(n_2, p) = b(n_1 + n_2, p)$.
Die Summe von n unabhängigen $b(1,p)$ verteilten ZVn ist also $b(n,p)$-verteilt.

b) Für $\lambda_1, \lambda_2 > 0$ gilt $\pi(\lambda_1) * \pi(\lambda_2) = \pi(\lambda_1 + \lambda_2)$.

c) Für $r_1, r_2 \in \mathbb{N},\ p \in (0,1)$ gilt $Nb(r_1, p) * Nb(r_2, p) = Nb(r_1 + r_2, p)$.
Die Summe von r unabhängigen $Nb(1,p)$-verteilten ZVn ist also $Nb(r,p)$-verteilt.

d) Für $a_1, a_2 \in \mathbb{R},\ \sigma_1^2, \sigma_2^2 \in (0,\infty)$ gilt $N(a_1, \sigma_1^2) * N(a_2, \sigma_2^2) = N(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$.

e) Für $\lambda \in (0,\infty),\ \nu_1, \nu_2 \in (0,\infty)$ gilt $\Gamma_{\lambda,\nu_1} * \Gamma_{\lambda,\nu_2} = \Gamma_{\lambda,\nu_1+\nu_2}$.
Die Summe von n unabhängigen $\exp(\lambda)$-verteilten ZVn ist $\Gamma_{\lambda,n}$-verteilt.
Die Summe der Quadrate von n unabhängigen jeweils $N(0,1)$-verteilten ZVn ist χ_n^2-verteilt.

Beweis: Siehe Übungen. □

6 Spezielle Verteilungen

6.1 Diskrete Verteilungen

6.1 Definition. Als **Binomialverteilung** $b(n,p)$ mit Parametern $n \in \mathbb{N},\ p \in (0,1)$ oder auch $[0,1]$ wird ein W-Maß auf $\mathcal{B}$ bezeichnet, das auf $\{0,1,\dots,n\}$ konzentriert ist und die Zähldichte

$$k \mapsto b(n,p;k) := \binom{n}{k} p^k (1-p)^{n-k},\ \ k \in \mathbb{N}_0$$

besitzt.

6.2 Satz. Sei $n \in \mathbb{N},\ p \in [0,1];\ q := 1-p$.

a) Ist $(X_1,\dots,X_n)$ ein n-tupel unabhängiger reeller ZVn mit $P[X_i = 1] = p$, $P[X_i = 0] = q$ — d.h. mit jeweiliger $b(1,p)$-Verteilung — $(i = 1,\dots,n)$, so ist $\sum\limits_{i=1}^{n} X_i \quad b(n,p)$-verteilt.

b) Für eine ZV X mit $P_X = b(n,p)$ gilt $EX = np$, $V(X) = npq$.

c) $b(n,p)$ hat die erzeugende Funktion g mit $g(s) = (q+ps)^n$.

d) Für $n_1, n_2 \in \mathbb{N}$ gilt

$$b(n_1,p) * b(n_2,p) = b(n_1+n_2,p),$$

d.h. für zwei unabhängige ZVn X_1, X_2 mit $P_{X_1} = b(n_1,p)$, $P_{X_2} = b(n_2,p)$ gilt $P_{X_1+X_2} = b(n_1+n_2,p)$.

6.3 Bemerkung. Das n-tupel $(X_1,\dots,X_n)$ von ZVn aus Satz 6.1 beschreibt ein n-tupel von Bernoulli-Versuchen, d.h. Zufallsexperimenten ohne gegenseitige Beeinflussung mit Ausgängen 1 ("Erfolg") oder 0 ("Misserfolg") und jeweiliger Erfolgswahrscheinlichkeit p; $\sum\limits_{k=1}^{n} X_i$ gibt die Anzahl der Erfolge in n Bernoulli-Versuchen an.

6.4 Definition. Als **Poisson-Verteilung** $\pi(\lambda)$ mit Parameter $\lambda > 0$ wird ein W-Maß auf $\mathcal{B}$ bezeichnet, das auf $\mathbb{N}_0$ konzentriert ist und die Zähldichte

$$k \mapsto \pi(\lambda;k) := e^{-\lambda}\frac{\lambda^k}{k!},\ \ k \in \mathbb{N}_0$$

besitzt.

6.5 Satz. Sei $\lambda > 0$.

a) Ist $(b(n,p_n))$ eine Folge von Binomialverteilungen mit $np_n \to \lambda\ (n \to \infty)$, dann konvergiert

$$b(n,p_n;k) \to \pi(\lambda;k) \quad (n \to \infty) \text{ für alle } k \in \mathbb{N}_0\,.$$

b) Für eine ZV X mit $P_X = \pi(\lambda)$ gilt $EX = V(X) = \lambda$.

c) $\pi(\lambda)$ hat die erzeugende Funktion g mit $g(s) = e^{-\lambda+\lambda s}$.

d) Für $\lambda_1, \lambda_2 > 0$ gilt
$$\pi(\lambda_1) * \pi(\lambda_2) = \pi(\lambda_1 + \lambda_2)\,.$$

Beweis von Satz 6.5 a):
Es gelte $n \cdot p_n \to \lambda$ $(n \to \infty)$, woraus $p_n \to 0$ $(n \to \infty)$ folgt. Mit

$$\begin{aligned} b(n, p_n; k) &= \binom{n}{k} \cdot p_n^k \cdot (1-p_n)^{n-k} \\ &= \tfrac{1}{k!} \cdot \underbrace{(n \cdot p_n)}_{\to\lambda} \cdot \underbrace{(n-1) \cdot p_n}_{\to\lambda} \cdot \ldots \cdot \underbrace{(n-k+1) \cdot p_n}_{\to\lambda} \cdot \left((1-p_n)^{\frac{1}{p_n}}\right)^{(n-k)\cdot p_n} \end{aligned}$$

und
$$(n-k) \cdot p_n \to \lambda \quad (n \to \infty)$$
sowie
$$(1-p_n)^{\frac{1}{p_n}} \to e^{-1} \quad (n \to \infty)$$
(wegen $p_n \to 0$ $(n \to \infty)$, woraus
$$\ln\left((1-p_n)^{\frac{1}{p_n}}\right) = \frac{\ln(1-p_n)}{p_n} \to -1 \;\; (n \to \infty)$$
folgt), folgt die Behauptung. □

6.6 Bemerkung. Bei einer rein zufälligen Aufteilung von n unterscheidbaren Teilchen auf m gleichgroße mehrfach besetzbare Zellen in einem Euklidischen Raum wird die Anzahl der Teilchen in einer vorgegebenen Zelle durch eine $b(n, \frac{1}{m})$-verteilte ZV angegeben. Der Grenzübergang $n \to \infty$, $m \to \infty$ mit *Belegungsintensität* $\frac{n}{m} \to \lambda > 0$ führt gemäß Satz 6.5a zu einer $\pi(\lambda)$-verteilten ZV.

6.7 Definition. Als **negative Binomialverteilung** oder **Pascal-Verteilung** $Nb(r,p)$ mit Parametern $r \in \mathbb{N}$, $p \in (0,1)$ wird ein W-Maß auf $\mathcal{B}$ (oder auch $\overline{\mathcal{B}}$) bezeichnet, das auf $\mathbb{N}_0$ konzentriert ist und — mit $q := 1-p$ — die Zähldichte
$$k \mapsto Nb(r,p;k) := \binom{r+k-1}{k} p^r q^k, \quad k \in \mathbb{N}$$
besitzt. Speziell wird $Nb(1,p)$ — mit Zähldichte $k \mapsto p(1-p)^k$, $k \in \mathbb{N}_0$ — als **geometrische Verteilung** mit Parameter $p \in (0,1)$ bezeichnet.

6.8 Satz. Sei $r \in \mathbb{N}$, $p \in (0,1)$, $q := 1-p$.

a) Sei $(X_n)_{n\in\mathbb{N}}$ eine unabhängige Folge von $b(1,p)$-verteilten ZVn. Die erweitert-reelle ZV $X := \inf\{n \in \mathbb{N} \mid \sum\limits_{k=1}^{n} X_k = r\} - r$ mit $\inf \emptyset := \infty$ ist $Nb(r,p)$-verteilt.

b) Für eine ZV X mit $P_X = Nb(r,p)$ gilt $EX = \frac{rq}{p}$, $V(X) = \frac{rq}{p^2}$.

c) $Nb(r,p)$ hat die erzeugende Funktion g mit $g(s) = p^r(1-qs)^{-r}$.

d) Für $r_1, r_2 \in \mathbb{N}$ gilt

$$Nb(r_1,p) * Nb(r_2,p) = Nb(r_1 + r_2, p)\,.$$

Die Summe von r unabhängigen $Nb(1,p)$-verteilten ZVn ist somit $Nb(r,p)$-verteilt.

6.9 Bemerkung. Für die Folge (X_n) in Satz 6.8a wird durch die erweitert-reelle ZV $T := \inf\{n \in \mathbb{N} \mid \sum\limits_{k=1}^{n} X_k = r\}$ mit $\inf \emptyset := \infty$ die *Wartezeit* bis zum r-ten Auftreten der Eins in (X_n) und durch die erweitert-reelle ZV $X = T - r$ die Anzahl der Misserfolge bis zum r-ten Erfolg bei der zu (X_n) gehörigen Folge von Bernoulli-Versuchen angegeben.

6.10 Definition. Als **hypergeometrische Verteilung** mit Parametern $n, r, s \in \mathbb{N}$, wobei $n \leq r + s$, wird ein W-Maß auf $\mathcal{B}$ bezeichnet, das auf $\{0, 1, \ldots, n\}$ — sogar auf $\{\max(0, n-s), \ldots, \min(n, r)\}$ — konzentriert ist und die Zähldichte

$$k \mapsto \begin{cases} \dfrac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}} & , \quad k = 0, 1, \ldots, n \\ 0 & , \quad k = n+1, n+2, \ldots \end{cases}$$

besitzt.

6.11 Bemerkung. Werden aus einer Urne mit r roten und s schwarzen Kugeln rein zufällig n Kugeln ohne Zurücklegen gezogen (n, r, s wie in Definition 6.10), so wird die Anzahl der gezogenen roten Kugeln durch eine ZV angegeben, die eine hypergeometrische Verteilung mit Parametern n, r, s besitzt. Anwendung der hypergeometrischen Verteilung in der Qualitätskontrolle.

6.2 Totalstetige Verteilungen

6.12 Definition. Als **Gleichverteilung auf (a,b)** mit $-\infty < a < b < \infty$ wird ein W-Maß auf $\mathcal{B}$ bezeichnet, das eine Dichte $f = \dfrac{1}{b-a}\chi_{(a,b)}$ besitzt.

6.13 Satz. Für eine ZV X mit Gleichverteilung auf (a,b) gilt $EX = \dfrac{a+b}{2}$, $V(X) = \dfrac{(b-a)^2}{12}$.

6.14 Definition. Als (eindimensionale) **Normalverteilung** oder Gauß-Verteilung $N(a, \sigma^2)$ mit Parametern $a \in \mathbb{R}$, $\sigma > 0$ wird ein W-Maß auf $\mathcal{B}$ bezeichnet, das eine Dichte f mit

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

besitzt. Speziell heißt $N(0,1)$ **standardisierte Normalverteilung**.

6.15 Bemerkung.

a) Sei f wie in Definition 6.14. Der Graph von f heißt Gaußsche Glockenkurve. Im Fall $a = 0$ ist $f(0) = \dfrac{1}{\sqrt{2\pi}\sigma}$ und hat f zwei Wendepunkte $(\pm\sigma; \dfrac{1}{\sqrt{2\pi e}\sigma})$.

b) Die Dichtefunktion und die VF ϕ von $N(0,1)$ sind tabelliert.

c) Ist die ZV X $N(a, \sigma^2)$-verteilt, so ist $\dfrac{X-a}{\sigma}$ $N(0,1)$-verteilt. Es gilt hierbei

$$\begin{aligned}
P[a - \sigma \leq X \leq a + \sigma] &= 2\phi(1) - 1 \approx 0,683 \\
P[a - 2\sigma \leq X \leq a + 2\sigma] &= 2\phi(2) - 1 \approx 0,954 \\
P[a - 3\sigma \leq X \leq a + 3\sigma] &= 2\phi(3) - 1 \approx 0,997 .
\end{aligned}$$

Zu Bemerkung 6.15 c):

$$\begin{aligned}
P[a - k \cdot \sigma \leq X \leq a + k \cdot \sigma] &= P[-k \leq \tfrac{X-a}{\sigma} \leq k] \\
&= \Phi(k) - \Phi(-k) \\
&= 2\Phi(k) - 1
\end{aligned}$$

da aus Symmetriegründen gilt:

$$\Phi(-k) = 1 - \Phi(k).$$

□

6.16 Satz. Sei $a \in \mathbb{R}$, $\sigma > 0$.

a) Für eine ZV X mit Verteilung $N(a, \sigma^2)$ gilt:

$$E(X-a)^{2k-1} = 0, \; E(X-a)^{2k} = \sigma^{2k} \prod_{j=1}^{k} (2j-1), \quad k \in \mathbb{N};$$

insbesondere $EX = a$, $V(X) = \sigma^2$; die ZV $cX + b$ mit $0 \neq c \in \mathbb{R}$, $b \in \mathbb{R}$ hat die Verteilung $N(ca + b, \; c^2\sigma^2)$.

b) Die charakteristische Funktion von $N(a, \sigma^2)$ ist φ mit

$$\varphi(u) = e^{iau} e^{-\frac{\sigma^2 u^2}{2}} .$$

c) Für $a_1, a_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$ gilt

$$N(a_1, \sigma_1^2) * N(a_2, \sigma_2^2) = N(a_1 + a_2, \sigma_1^2 + \sigma_2^2)\,.$$

6.17 Definition. Als **Exponentialverteilung** $\exp(\lambda)$ mit Parameter $\lambda > 0$ wird ein W-Maß auf $\mathcal{B}$ oder $\overline{\mathcal{B}}$ bezeichnet, das eine Dichtefunktion f mit

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

besitzt.

6.18 Satz. Sei $\lambda > 0$.

a) Sei X eine erweitert-reelle ZV, deren Verteilung auf $\overline{\mathbb{R}}_+$ konzentriert sei, mit $P[0 < X < \infty] > 0$. Dann gilt:

$$\underset{s,t\in(0,\infty)}{\forall} \; P[X > t + s \mid X > s] = P[X > t] \iff P_X \text{ ist eine Exponentialverteilung.}$$

"Gedächtnislosigkeit"

b) Für eine ZV X mit $P_X = \exp(\lambda)$ gilt $EX = \dfrac{1}{\lambda}$, $\;V(X) = \dfrac{1}{\lambda^2}$.

c) $\exp(\lambda)$ hat die charakteristische Funktion φ mit

$$\varphi(u) = \frac{\lambda}{\lambda - iu}\,.$$

6.19 Definition. Als **Gamma-Verteilung** $\Gamma_{\lambda,\nu}$ mit Parametern $\lambda, \nu > 0$ wird ein W-Maß auf $\mathcal{B}$ oder $\overline{\mathcal{B}}$ bezeichnet, das eine Dichtefunktion f mit

$$f(x) = \begin{cases} \dfrac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

besitzt. Hierbei $\Gamma_{\lambda,1} = \exp(\lambda)$. $\Gamma_{\frac{1}{2},\frac{n}{2}}$ wird als Chi-Quadrat-Verteilung χ_n^2 mit $n (\in \mathbb{N})$ Freiheitsgraden bezeichnet.

6.20 Satz. Seien $\lambda, \nu > 0$, $\; n \in \mathbb{N}$.

a) Für eine ZV X mit $P_X = \Gamma_{\lambda,\nu}$ gilt $EX = \frac{\nu}{\lambda}$, $V(X) = \frac{\nu}{\lambda^2}$.

b) $\Gamma_{\lambda,\nu}$ hat die charakteristische Funktion φ mit

$$\varphi(u) = \left(\frac{\lambda}{\lambda - iu}\right)^\nu.$$

c) Für $\nu_1, \nu_2 > 0$ gilt

$$\Gamma_{\lambda,\nu_1} * \Gamma_{\lambda,\nu_2} = \Gamma_{\lambda,\nu_1+\nu_2}\,.$$

Insbesondere ist $\Gamma_{\lambda,n}$ die n-fache Faltung von $\exp(\lambda)$.

d) Die Summe der Quadrate von n unabhängigen jeweils $N(0,1)$-verteilten reellen ZVn ist χ_n^2-verteilt.

Zu Satz 6.20 d)
Zu zeigen: $X_1, \ldots, X_n$ unabhängig $\mathcal{N}(0,1)$-verteilt $\Rightarrow$ $X_1^2 + \ldots + X_n^2$ χ_n^2-verteilt, d.h. $\Gamma_{\frac{1}{2},\frac{n}{2}}$-verteilt.
Wegen Faltungseigenschaft von $\Gamma_{\lambda,\nu}$ genügt es zu zeigen:
X_1^2 ist $\Gamma_{1/2,1/2}$-verteilt.

Dazu: Für $x \geq 0$ gilt

$$\begin{aligned}
F(x) &= P[X_1^2 \leq x] = P[-\sqrt{x} \leq X_1 \leq \sqrt{x}] \\
&= \int\limits_{-\sqrt{x}}^{\sqrt{x}} \tfrac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \\
&= \tfrac{2}{\sqrt{2\pi}} \int\limits_{0}^{\sqrt{x}} e^{-t^2/2} dt.
\end{aligned}$$

Ableiten liefert:

$$\begin{aligned}
F'(x) &= \tfrac{2}{\sqrt{2\pi}} e^{-x/2} \cdot \tfrac{1}{2\sqrt{x}} = \tfrac{1}{\sqrt{2\pi}} \cdot x^{-1/2} \cdot e^{-x/2} \\
&= \tfrac{(1/2)^{1/2}}{\Gamma(1/2)} \cdot x^{1/2-1} \cdot e^{-1/2 \cdot x} \ldots \text{Dichte zu } \Gamma_{1/2,1/2} \\
& \quad \text{da } \Gamma(1/2) = \sqrt{\pi})
\end{aligned}$$

$\Rightarrow X_1^2$ ist $\Gamma_{1/2,1/2}$-verteilt.
(denn aus der obigen Beziehung folgt $\int\limits_0^x F'(t)dt = F(x) - F(0+) = F(x)$ für $x > 0$)

6.21 Bemerkung. Sind X, Y zwei unabhängige reelle ZVn mit $P_X = N(0,1)$ und $P_Y = \chi_n^2$, so wird die Verteilung der ZV $\dfrac{X}{\sqrt{Y/n}}$ als ***t*-Verteilung** oder **Student-Verteilung** t_n bzw. St_n mit n Freiheitsgraden bezeichnet ($n \in \mathbb{N}$). t_1 bzw. St_1 ist eine sogenannte **Cauchy-Verteilung**. — Anwendung von χ_n^2 und St_n in der Statistik.

Beispiel aus der Erneuerungstheorie:
Bauelemente werden nach Ausfall sofort ersetzt, wobei:

- die Bauelemente hinsichtlich der Lebensdauer das gleiche stochastische Verhalten haben.

- die Bauelemente sich hinsichtlich der Lebensdauer unabhängig voneinander verhalten.

Die Bauelemente seien in der Reihenfolge des Einsatzes durchnummeriert.
Zur stochastischen Modellierung fassen wir die Lebensdauern als Realisierungen unabhängig, identisch verteilter, nichtnegativer reeller Zufallsvariablen $T_1, T_2, \ldots$ auf, definiert auf W-Raum $(\Omega, \mathcal{A}, P)$.
Es gelte: $P[T_1 = 0] < 1$ (sonst Trivialfall).
Dann beschreibt

$$N_t = \sup\{n \in \mathbb{N} : T_1 + \ldots + T_n \leq t\}$$

(mit $\sup \emptyset = 0$) die Anzahl der Erneuerungen im Zeitintervall $[0, t]$.
Hierbei ist

$$(T_1 + \ldots + T_n)_{n \in \mathbb{N}} \quad \ldots \quad \text{sog. Erneuerungsprozeß}$$

$\{N_t : t \in \mathbb{R}_+\} \ldots$ sog. Zählprozeß zum Erneuerungsprozeß.
N_t ist eine $\bar{\mathbb{N}}_0 = \mathbb{N}_0 \cup \{\infty\}$-wertige ZV, da:

$$[N_t = \infty] = \bigcap_{n \in \mathbb{N}} \underbrace{[T_1 + \ldots + T_n \leq t]}_{\in \mathcal{A}} \in \mathcal{A}$$

$$[N_t = k] = \underbrace{[T_1 + \ldots + T_k \leq t]}_{\in \mathcal{A}} \setminus \underbrace{[T_1 + \ldots + T_{k+1} \leq t]}_{\in \mathcal{A}} \in \mathcal{A}.$$

6.22 Satz. Ist — mit den obigen Bezeichnungen — T_n $\exp(\lambda)$-verteilt ($\lambda > 0$), so ist N_t $\pi(\lambda t)$-verteilt, $t \in \mathbb{R}_+$.

6.23 Bemerkung. Die Familie $\{N_t \mid t \in \mathbb{R}_+\}$ aus Satz 6.22 ist ein sogenannter **Poisson-Prozess**.

Beweis von Satz 6.22:
$S_k = T_1 + \ldots + T_k$ ist $\Gamma_{\lambda,k}$-verteilt nach Satz 6.20 c).
Daher gilt für $k \in \mathbb{N}$

$$\begin{aligned} P[N_t = k] &= P[S_k \leq t] - P[S_{k+1} \leq t] \\ &= \int_0^t \underbrace{\frac{\lambda^k}{(k-1)!} \cdot x^{k-1}}_{u'} \cdot \underbrace{e^{-\lambda x}}_{v} dx - \int_0^t \underbrace{\frac{\lambda^{k+1}}{k!} \cdot x^k \cdot e^{-\lambda x}}_{=-v' \cdot u} dx \end{aligned}$$

Mit partieller Integration des ersten Integrals folgt daraus

$$P[N_t = k] = \left[\lambda^k \cdot \frac{x^k}{k!} \cdot e^{-\lambda x}\right]_{x=0}^t = \frac{(\lambda t)^k}{k!} \cdot e^{-\lambda t},$$

d. h. $P[N_t = k]$ stimmt für $k \in \mathbb{N}$ mit dem Wert der Zähldichte von $\pi(\lambda t)$ an der Stelle k überein.

Für $k = 0$ gilt

$$P[N_t = 0] = P[S_1 > t] = P[T_1 > t] = \int\limits_t^\infty \lambda \cdot e^{-\lambda x} dx = e^{-\lambda t},$$

also gilt die obige Aussage auch für $k = 0$. $\square$

7 Bedingte Erwartungen

Hilfsmittel: **Satz von Radon-Nikodym** (vereinfachte Fassung)
Messraum $(\Omega, \mathcal{A})$, endliches Maß μ auf $\mathcal{A}$, signiertes endliches Maß ν auf $\mathcal{A}$ (d. h. ν ist Differenz zweier endlicher Maße). Es gelte

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

(d. h. ν ist μ stetig).
Dann existiert Funktion

$$f : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$$

die bzgl. μ integrierbar ist, und für die gilt:

$$\forall A \in \mathcal{A} : \nu(A) = \int\limits_A f d\mu.$$

f ist eindeutig bis auf Äquivalenz "$=$ μ-f.ü.".
f ist die sogenannte **Randon-Nikodym-Ableitung** von ν nach μ.
Schreibweise:

$$f = \frac{d\nu}{d\mu}$$

Beweis: Siehe Integrationstheorie bzw. Maßtheorie. $\square$
Es gilt:

$$\frac{d\varphi}{d\mu} = \frac{d\varphi}{d\nu} \cdot \frac{d\nu}{d\mu}$$

mit

μ endliches Maß, φ μ stetig
ν endliches Maß, φ ν stetig
φ endliches signiertes Maß,
ν μ-stetig

(vgl. $\int\limits_{\mathbb{R}} h(x) dP_X(x) = \int\limits_{\mathbb{R}} h(u) \cdot f(u) du$ falls f Dichte von P_X ist).
Aus dem Satz von Radon-Nikodym folgt:

7.1 Satz. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. **Integrierbare** ZV X: $(\Omega, \mathcal{A}, P) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$. σ-Algebra $\mathcal{C} \subset \mathcal{A}$. Dann existiert eine ZV $Z : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B})$ mit folgenden Eigenschaften:

$(*)$ Z ist integrierbar und $\mathcal{C}$-$\mathcal{B}$-messbar,

$(**)$ $\underset{C\in\mathcal{C}}{\forall} \int_C X\,dP = \int_C Z\,dP\,.$

Z ist eindeutig bis auf die Äquivalenz "$=$ Rest$_\mathcal{C}$ P-f.ü.".

Beweis
Definiere $\varphi : \mathcal{C} \to \mathbb{R}$ durch

$$\varphi(C) = \int_C X\ dP = \int_C X^+\ dP - \int_C X^-\ dP.$$

Dann ist φ endliches signiertes Maß auf $\mathcal{C}$, das $P\big|_\mathcal{C}$-stetig ist (denn aus $P(C) = 0$ für ein $C \in \mathcal{C}$ folgt $\varphi(C) = \int_C\ X\ dP = 0$).
Nach dem Satz von Radon-Nikodym (angewendet mir $\nu = \varphi$ wie oben und $\mu = P\big|_\mathcal{C}$) existiert eine Funktion $Z : \Omega \to \mathbb{R}$ mit

- Z ist $\mathcal{C} - \mathcal{B}$-messbar
- Z ist integrierbar bzgl. $P\big|_\mathcal{C}$
- Für alle $C \in \mathcal{C}$ gilt
$$\varphi(C) = \int_C\ Z\ dP\big|_\mathcal{C}$$
d. h.
$$\int_C\ X\ dP = \int_C\ Z\ dP\big|_\mathcal{C} = \int_C\ Z\ dP,$$
wobei die letzte Gleichheit aus der $\mathcal{C} - \mathcal{B}-$Messbarkeit von Z folgt.

Zur Eindeutigkeit:
Sind Z_1, Z_2 zwei $\mathcal{C} - \mathcal{B}$-messbare Zufallsvariablen, für die (**) gilt, so folgt

$$\forall\ C \in \mathcal{C} : \int_C (Z_1 - Z_2) dP = 0.$$

Dies impliziert

$$\int (Z_1 - Z_2) \cdot 1_{\{Z_1 - Z_2 > 0\}}\,dP = \int_{\{Z_1 - Z_2 > 0\}} (Z_1 - Z_2)\,dP = 0,$$

woraus (wegen der Nichtnegativität des Integranden)

$$(Z_1 - Z_2) \cdot 1_{\{Z_1 - Z_2 > 0\}} = 0 \quad P. - f.s.$$

folgt. Analog sieht man

$$(Z_1 - Z_2) \cdot 1_{\{Z_1 - Z_2 < 0\}} = 0 \quad P. - f.s.,$$

womit insgesamt

$$Z_1 = Z_2 \quad P. - f.s.$$

gezeigt ist, woraus $Z_1 = Z_2 \; P\big|_{\mathcal{C}}$-f.ü. folgt. $\square$

7.2 Definition. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Integrierbare ZV $X\colon (\Omega, \mathcal{A}, P) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$. σ-Algebra $\mathcal{C} \subset \mathcal{A}$. Die Äquivalenzklasse (im oberen Sinne) der ZVn $Z\colon (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B})$ mit $(*)$ und $(**)$ — oder auch ein Repräsentant dieser Äquivalenzklasse — heißt **bedingte Erwartung von X bei gegebenem $\mathcal{C}$** $\ldots E(X \mid \mathcal{C})$. Häufig wird ein Repräsentant dieser Äquivalenzklasse als eine Version von $E(X \mid \mathcal{C})$ bezeichnet.

Achtung $E(X|\mathcal{C})$ ist Zufallsvariable / messbare Abbildung!
$E(X|\mathcal{C})$ ist "Vergröberung"von X.
Bemerkung:

$$E(X|\mathcal{C}) = \frac{d\varphi}{dP\big|_{\mathcal{C}}} \text{ mit } \varphi(C) = \int_C X \; dP \; (C \in \mathcal{C})$$

Beispiele

a) $\mathcal{C} = \mathcal{A} \ldots E(X \mid \mathcal{C}) = X$ f.s.
(beachte: $E(X|\mathcal{C})$ ist reellwertig, daher muss X eventuell auf der Nullmenge $[|X| = \infty]$ abgeändert werden).

b) $\mathcal{C} = \{\emptyset, \Omega\} \ldots E(X \mid \mathcal{C}) = EX$ (da $E(X|\mathcal{C})$ $\mathcal{C} - \mathcal{B}$-messbar ist, muss hier $E(X|\mathcal{C})$ konstant sein, und mit $\int_\Omega E(X|\mathcal{C}) \; dP = EX$ folgt die Behauptung).

c) $\mathcal{C} = \{\emptyset, B, B^c, \Omega\}$ mit $0 < P(B) < 1$.

$$(E(X \mid \mathcal{C}))(\omega) = \begin{cases} \dfrac{1}{P(B)} \displaystyle\int_B X \, dP =: E(X \mid B), \;\; \omega \in B \\ \dfrac{1}{P(B^c)} \displaystyle\int_{B^c} X \, dP, \;\; \omega \in B^c \end{cases}$$

$E(X \mid B)$ heißt **bedingter Erwartungswert von X unter der Hypothese B.**

Begründung:

Die rechte Seite ist $\mathcal{C} - \mathcal{B}$-messbar und erfüllt:

- $\int_\emptyset (\text{Re.S.}) \; dP = 0 = \int_\emptyset X \; dP$
- $$\begin{aligned} \int_B (\text{Re.S.}) \; dP &= \int_B \frac{\int_B X \, dP}{P(B)} dP = \frac{\int_B X \, dP}{P(B)} \cdot \int_B 1 \; dP \\ &= \int_B X \; dP. \end{aligned}$$
- $\int_{B^c} (\text{Re.S.}) \; dP \overset{analog}{=} \int_{B^c} X \; dP$

–

$$\begin{aligned}\int_\Omega (\text{Re.S.})\, dP &= \int_B (\text{Re.S.})\, dP + \int_{B^c} (\text{Re.S.})\, dP \\ &\overset{s.o.}{=} \int_B X\, dP + \int_{B^c} X\, dP = \int_\Omega X\, dP.\end{aligned}$$

Aus obiger Rechnung folgt auch, dass die konstanten Funktionswerte von $E(X|\mathcal{C})$ auf B bzw. B^c eindeutig sind. □

7.3 Satz. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. X, X_i integrierbar; σ-Algebra $\mathcal{C} \subset \mathcal{A}$; $c, \alpha_1, \alpha_2 \in \mathbb{R}$.

a) $\underset{C\in\mathcal{C}}{\forall} \displaystyle\int_C E(X \mid \mathcal{C}) dP = \int_C X\, dP$

b) $X = c$ P-f.s. $\Longrightarrow E(X \mid \mathcal{C}) = c$ f.s.

c) $X \geq 0$ P-f.s. $\Longrightarrow E(X \mid \mathcal{C}) \geq 0$ f.s.

d) $E(\alpha_1 X_1 + \alpha_2 X_2 \mid \mathcal{C}) = \alpha_1 E(X_1 \mid \mathcal{C}) + \alpha_2 E(X_2 \mid \mathcal{C})$ f.s.

e) $X_1 \leq X_2$ P-f.s. $\Longrightarrow E(X_1 \mid \mathcal{C}) \leq E(X_2 \mid \mathcal{C})$ f.s.

f) X $\mathcal{C}$-$\mathcal{B}$-messbar $\Longrightarrow X = E(X \mid \mathcal{C})$ f.s.

g) X integrierbar, Y $\mathcal{C}$-$\mathcal{B}$-messbar, XY integrierbar $\Longrightarrow E(XY \mid \mathcal{C}) = Y E(X \mid \mathcal{C})$ f.s.

h) X, X' integrierbar, $XE(X' \mid \mathcal{C})$ integrierbar
$\Longrightarrow E(XE(X' \mid \mathcal{C}) \mid \mathcal{C}) = E(X \mid \mathcal{C}) E(X' \mid \mathcal{C})$ f.s.

i) σ-Algebren $\mathcal{C}_1, \mathcal{C}_2$ mit $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \mathcal{A}$, X integrierbar

$$E(E(X \mid \mathcal{C}_1) \mid \mathcal{C}_2) = E(X \mid \mathcal{C}_1) \text{ f.s.}$$
$$E(E(X \mid \mathcal{C}_2) \mid \mathcal{C}_1) = E(X \mid \mathcal{C}_1) \text{ f.s.}$$

Hier f.s. ... Rest$_\mathcal{C}$ P-f.s. bzw. Rest$_{\mathcal{C}_1}$ P-f.s.

Beweis:

a) Folgt unmittelbar aus der Definition.

b) Klar, da die Konstante die Messbarkeits- und Integralbedingungen erfüllt.

c) $\forall\, C \in \mathcal{C}: \quad \int_C E(X|\mathcal{C})\, dP = \int_C X\, dP \geq 0$
(da $X \geq 0$ P-f.s.)

$\Rightarrow E(X|\mathcal{C}) \geq 0$ f.s.
(da $0 \leq \displaystyle\int_{[E(X|\mathcal{C})<-\frac{1}{n}]} E(X|\mathcal{C})\, dP \leq -\frac{1}{n} \cdot P[E(X|\mathcal{C}) \leq -\frac{1}{n}$,
also $P[E(X|\mathcal{C}) < 0] = \lim\limits_{n\to\infty} P[E(X|\mathcal{C}) < -\frac{1}{n}] = \lim\limits_{n\to\infty} 0 = 0$.)

d) Rechte Seite ist $\mathcal{C}-\mathcal{B}$-messbar, integrierbar und erfüllt für alle $C \in \mathcal{C}$:

$$\begin{aligned}
&\int_C (\alpha_1 E(X_1|\mathcal{C}) + \alpha_2 \cdot E(X_2|\mathcal{C}))\, dP \\
&= \alpha_1 \cdot \int_C E(X_1|\mathcal{C})\, dP + \alpha_2 \cdot \int_C E(X_2|\mathcal{C})\, dP \\
&\overset{Definition}{=} \alpha_1 \cdot \int_C X_1\, dP + \alpha_2 \cdot \int_C X_2\, dP \\
&= \int_C (\alpha_2 \cdot X_1 + \alpha_2 \cdot X_2)\, dP.
\end{aligned}$$

e)

$$\begin{aligned}
&X_1 \leq X_2 \ \text{ P-f.s. } \Rightarrow \ X_2 - X_1 \geq 0 \text{ P-f.s.} \\
&\overset{c)}{\Rightarrow} E(X_2 - X_1|\mathcal{C}) \geq 0 \ \text{ f.s.} \\
&\overset{d)}{\Rightarrow} E(X_2|\mathcal{C}) - E(X_1|\mathcal{C}) \geq 0 \ \text{ f.s.}
\end{aligned}$$

f) Klar, da X $\mathcal{C}-\mathcal{B}$-messbar ist und die Integralbedingung erfüllt.

g) **Fall I:** $X \geq 0$ und $Y \geq 0$
Da $Y \cdot E(X|\mathcal{C})$ $\mathcal{C}-\mathcal{B}$-messbar ist, genügt es zu zeigen:

(1) $Y \cdot E(X|\mathcal{C})$ integrierbar

(2) $\forall\, C \in \mathcal{C}: \quad \int_C Y \cdot E(X|\mathcal{C})\, dP = \int_C X \cdot Y\, dP$

(1) folgt wegen $Y \cdot E(X|\mathcal{C}) \geq 0$ (beachte $X \geq 0$ und c)) aus (2) (setze darin $C = \Omega$ und beachte $X \cdot Y$ integrierbar).

Nachweis von (2):

Fall 1: $X \geq 0$ und $Y = \chi_B$ für ein $B \in \mathcal{C}$.

Dann gilt für $C \in \mathcal{C}$:

$$\begin{aligned}
\int_C Y \cdot E(X|\mathcal{C})\, dP &= \int\limits_{B\cap C} E(X|\mathcal{C})\, dP \\
&= \int\limits_{B\cap C} X\, dP \\
&\quad \text{(nach Definition der bedingten Erwartung und } B \cap C \in \mathcal{C}) \\
&= \int_C X \cdot \chi_B\, dP = \int_C X \cdot Y\, dP
\end{aligned}$$

Fall 2: $X \geq 0$ und Y nichtnegativ einfach.

Beh. folgt mit Linearität des Integrals aus Fall 1.

Fall 3: $X \geq 0$ und $Y \geq 0$

Behauptung folgt aus Fall 2, indem man Y von unten durch nichtnegative Funktionen appoximiert, den Satz von der monotonen Konvergenz und das Resultat aus Fall 2 anwendet.

Fall II: Allgemeine X und Y.

$$\begin{aligned}
E(X \cdot Y|\mathcal{C}) &= E(X^+ \cdot Y^+ - X^- \cdot Y^+ - X^+ \cdot Y^- + X^- \cdot Y^-|\mathcal{C}) \\
&\overset{d)}{=} E(X^+ \cdot Y^+|\mathcal{C}) - E(X^- \cdot Y^+|\mathcal{C}) - E(X^+ \cdot Y^-|\mathcal{C}) + E(X^- \cdot Y^-|\mathcal{C}) \;\; f.s. \\
&= Y^+ \cdot E(X^+|\mathcal{C}) - Y^+ \cdot E(X^-|\mathcal{C}) - Y^- \cdot E(X^+|\mathcal{C}) + Y^- \cdot E(X^-|\mathcal{C}) \quad f.s. \\
& \text{(nach Fall I, wobei alle auftretenden Produkte integrierbar sind,} \\
& \text{ da } X \cdot Y \text{ integrierbar ist)} \\
&= (Y^+ - Y^-) \cdot (E(X^+|\mathcal{C}) - E(X^-|\mathcal{C})) \\
&\overset{d)}{=} Y \cdot E(X|\mathcal{C}) \quad \text{f.s.}
\end{aligned}$$

h) Folgt aus g) mit $Y = E(X'|\mathcal{C})$.

i) i_1) $E(X|\mathcal{C}_1)$ ist $\mathcal{C}_1 - \mathcal{B}$-messbar und damit auch $\mathcal{C}_2 - \mathcal{B}$-messbar. Mit f) folgt die Behauptung.

i_2) $E(X|\mathcal{C}_1)$ ist $\mathcal{C}_1 - \mathcal{B}$-messbar, integrierbar, und es gilt für $C \in \mathcal{C}_1$:

$$\int_C E(X \mid \mathcal{C}_1)\, dP \overset{Def.}{=} \int_C X\, dP = \int_C E(X|\mathcal{C}_2)\, dP$$

wobei die letzte Gleichheit aus $C \in \mathcal{C}_2$ und der Definition von $E(X|\mathcal{C}_2)$ folgt.

$\square$

7.4 Definition. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. σ-Algebra $\mathcal{C} \subset \mathcal{A}$. $A \in \mathcal{A}$.
$P(A \mid \mathcal{C}) := E(\chi_A \mid \mathcal{C})$ heißt **bedingte Wahrscheinlichkeit von A bei gegebenem $\mathcal{C}$.**

7.5 Bemerkung. Zu Definition 7.4.

$$\underset{C \in \mathcal{C}}{\forall} \int_C P(A \mid \mathcal{C})\, dP = P(A \cap C).$$

Denn:

$$\begin{aligned}
\textstyle\int_C P(A|\mathcal{C})\, dP &= \textstyle\int_C E(\chi_A|\mathcal{C})\, dP \\
&= \textstyle\int_C \chi_A\, dP \\
& \text{(nach Definition von } E(\chi_A|\mathcal{C})) \\
&= P(A \cap C).
\end{aligned}$$

Beispiel. $\mathcal{C} = \{\emptyset, B, B^c, \Omega\}$ mit $0 < P(B) < 1$.

$$(P(A \mid \mathcal{C}))(\omega) = \begin{cases} \dfrac{P(A \cap B)}{P(B)} =: P(A \mid B), & \omega \in B \\[2ex] \dfrac{P(A \cap B^c)}{P(B^c)} =: P(A \mid B^c), & \omega \in B^c\,. \end{cases}$$

denn nach dem Beispiel nach Definition 7.1 gilt:

$$(E(\chi_A|\mathcal{C}))(\omega) = \begin{cases} \frac{\int_B \chi_A \, dP}{P(B)} & \text{für} \quad \omega \in B, \\ \frac{\int_{B^c} \chi_A \, dP}{P(B^c)} & \text{für} \quad \omega \in B^c. \end{cases}$$

Definition 7.6. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$.

a) Integrierbare ZV $X\colon (\Omega, \mathcal{A}, P) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$. ZV $Y\colon (\Omega, \mathcal{A}, P) \to (\Omega', \mathcal{A}')$.
$E(X \mid Y) := E(X \mid \underbrace{Y^{-1}(\mathcal{A}')}) \ldots$ **bedingte Erwartung von X bei gegebebenem Y.**
[kleinste σ-Algebra in Ω, bzgl. der Y messbar ist $\ldots \mathcal{F}(Y)(\subset \mathcal{A})$]

b) Integrierbare ZV $X\colon (\Omega, \mathcal{A}, P) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$. ZVn $Y_i\colon (\Omega, \mathcal{A}, P) \to (\Omega'_i, \mathcal{A}'_i)$ $(i \in I)$
$\mathcal{C}(\subset \mathcal{A})$ sei die kleinste σ-Algebra in Ω, bzgl. der alle Y_i messbar sind
$[\mathcal{C} = \mathcal{F}(\bigcup_{i \in I} Y_i^{-1}(\mathcal{A}_i)) \ldots \mathcal{F}(Y_i, \ i \in I)]$
$E(X \mid (Y_i)_{i \in I}) := E(X \mid \mathcal{C}) \ldots$ **bedingte Erwartung von X bei gegebenem $Y_i, \ i \in I$.**

c) $A \in \mathcal{A}$; ZV $Y\colon (\Omega, \mathcal{A}, P) \to (\Omega', \mathcal{A}')$.
$P(A \mid Y) := E(\chi_A \mid Y) \ldots$ **bedingte Wahrscheinlichkeit von A bei gegebenem Y.**

7.7 Bemerkung. Integrierbare ZV $X\colon (\Omega, \mathcal{A}, P) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$.

a) σ-Algebra $\mathcal{C}$ in $\mathcal{A}$
$(X^{-1}(\overline{\mathcal{B}}), \mathcal{C})$ unabhängig $\Longrightarrow E(X \mid \mathcal{C}) = EX$ f.s.

b) ZV $Y\colon (\Omega, \mathcal{A}, P) \to (\Omega', \mathcal{A}')$
(X, Y) unabhängig $\Rightarrow E(X \mid Y) = EX$ f.s.

Beweis:

a) EX ist als konstante Abbildung $\mathcal{C} - \mathcal{B}$-messbar, integrierbar und für $C \in \mathcal{C}$ gilt:

$$\begin{aligned}
&\textstyle\int_C X \, dP = \int X \cdot \chi_C \, dP = E(X \cdot \chi_C) \\
&= E(X) \cdot E(\chi_C) \\
&\quad \text{(da } X, \chi_C \text{ unabhängig wegen } X^{-1}(\overline{\mathcal{B}}),\ \mathcal{C} \text{ unabhängig)} \\
&= E(X) \cdot P(C) = \textstyle\int_C E(X) \, dP.
\end{aligned}$$

b) folgt aus a) mit $\mathcal{C} = \mathcal{F}(Y) = Y^{-1}(\mathcal{A}')$

$\square$

7.8 Satz. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Integrierbare ZV $X\colon (\Omega, \mathcal{A}, P) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$. ZV $Y\colon (\Omega, \mathcal{A}, P) \to (\Omega', \mathcal{A}')$. Dann ex. Abb. $g\colon (\Omega', \mathcal{A}') \to (\mathbb{R}, \mathcal{B})$ mit $E(X \mid Y) = g \circ Y$.
g ist die sog. **Faktorisierung der bedingten Erwartung**.
g ist eindeutig bis auf die Äquivalenz "$= P_Y$-f.ü. ".

Beweis:

a) **Existenz**

Setze $Z = E(X|Y)$. Dann ist Z eine Abbildung $Z : (\Omega, \mathcal{F}(Y)) \to (\mathbb{R}, \mathcal{B})$, d. h. Z ist $\mathcal{F}(Y) - \mathcal{B}$-messbar.

Fall 1: $Z = \chi_A$

Wegen Z $\mathcal{F}(Y) - \mathcal{B}$-messbar gilt $A = \chi_A^{-1}(\{1\}) \in \mathcal{F}(Y) = Y^{-1}(\mathcal{A}')$.
Also existiert $A' \in \mathcal{A}'$ mit $A = Y^{-1}(A')$ und es gilt für $\omega \in \Omega$:

$$Z(\omega) = \chi_{Y^{-1}(A')}(\omega) = \chi_{A'}(Y(\omega)) = (g \circ Y)(\omega)$$

mit $g = \chi_{A'}$.

Fall 2: Z nichtnegativ einfach.

Dann gilt

$$Z = \sum_{i=1}^{n} \alpha_i \cdot \chi_{Y^{-1}(A_i')}$$

und – analog zu Fall 1 – folgt die Behauptung mit

$$g = \sum_{i=1}^{n} \alpha_i \cdot \chi_{A_i'}$$

.

Fall 3: Z nichtnegativ messbar.

Dann existieren nichtnegative einfache $\mathcal{F}(Y) - \mathcal{B}$-messbare (!) Zufallsvariablen Z_n mit $Z_n \uparrow Z$. Nach Fall 2 gilt $Z_n = g_n \circ Y$ mit g_n nichtnegativ einfach. Setze

$$g^* = \sup_n g_n \text{ und } g = g^* \cdot \chi_{[g^* < \infty]}.$$

Dann gilt wegen $Z < \infty$:

$$Z = \sup_n Z_n = \sup_n (g_n \circ Y) = (\sup_n g_n) \circ Y = g \circ Y.$$

Fall 4: Z messbar

Dann gilt $Z = Z^+ - Z^-$ für nichtnegative $\mathcal{F}(Y) - \mathcal{B}$-messbare ZVen Z^+ und Z^-. Nach Fall 3 gilt

$$Z^+ = g^+ \circ Y \text{ und } Z^- = g^- \circ Y$$

und daher

$$Z = (g^+ - g^-) \circ Y$$

$\rightsquigarrow$ Existenz.

b) **Eindeutigkeit**

Aus

$$E(X|Y) = g_1 \circ Y = g_2 \circ Y \quad f.s.$$

folgt für alle $\quad C \in \mathcal{F}(Y) = Y^{-1}(\mathcal{A}')$:

$$\int_C g_1 \circ Y \, dP = \int_C X \, dP = \int_C g_2 \circ Y \, dP.$$

Also gilt für alle $A' \in \mathcal{A}'$:

$$\begin{aligned} 0 &= \int\limits_{Y^{-1}(A')} g_1 \circ Y \, dP - \int\limits_{Y^{-1}(A')} g_2 \circ Y \, dP \\ &= \int\limits_{Y^{-1}(A')} (g_1 - g_2) \circ Y \, dP \\ &= \int\limits_{A'} (g_1 - g_2) \, dP_Y \\ &\quad \text{(nach dem Transformationssatz)}, \end{aligned}$$

woraus $g_1 - g_2 = 0$ P_Y- f.ü., also $g_1 = g_2$ P_Y-f.ü., folgt. $\square$

7.9 Definition. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Integrierbare ZV X: $(\Omega, \mathcal{A}, P) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ bzw. $A \in \mathcal{A}$. ZV Y: $(\Omega, \mathcal{A}, P) \to (\Omega', \mathcal{A}')$. Sei g bzw. g_A eine — bis auf Äquivalenz "$= P_Y$ - f.ü." eindeutig bestimmte — Faktorisierung von $E(X|Y)$ bzw. von $P(A|Y)$.
$E(X \mid Y = y) := g(y) \ldots$ **bedingte Erwartung von X unter der Hypothese** $Y = y$
$P(A \mid Y = y) := g_A(y) \ldots$ **bed. Wahrscheinlichkeit von A unter der Hypoth.** $Y = y$
$E(X \mid Y = \cdot) = g$
$P(A \mid Y = \cdot) = g_A$

Beispiel:
W-Raum $(\Omega, \mathcal{A}, P)$, integrierbare ZV $X : (\Omega, \mathcal{A}, P) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ bzw. $A \in \mathcal{A}$,
ZV $Y : (\Omega, \mathcal{A}, P) \to (\Omega', \mathcal{A}')$. Für $y \in \Omega'$ mit $\{y\} \in \mathcal{A}'$ und $P_Y(\{y\}) > 0$ gilt:

$$\underbrace{E(X|Y = y)}_{\text{im Sinne von Definition 7.9}} \quad = \quad \underbrace{E(X|[Y = y]) = \frac{\int_{[Y=y]} X \, dP}{P[Y = y]}}_{\text{wie im Beispiel nach Definition 7.2}}$$

bzw.

$$\underbrace{P(A|Y=y)}_{\substack{\text{im Sinne von}\\ \text{Definition 7.9}}} = \underbrace{P(A|[Y=y]) = \frac{P(A\cap[Y=y])}{P[Y=y]}}_{\text{wie im Beispiel nach Definition 7.4}}$$

Begründung:
Für $g = E(X|Y=\circ)$ und $C = Y^{-1}(\{y\}) \in Y^{-1}(\mathcal{A}')$ (gilt wegen $\{y\} \in \mathcal{A}'$) gilt nach Definition der bedingten Erwartung:

$$\int_C g\circ Y\,dP = \int_C X\,dP.$$

$$\begin{aligned}
\text{linke Seite} &= \int\limits_{Y^{-1}(\{y\})} g\circ Y\,dP \;=\; g(y)\cdot \int\limits_{Y^{-1}(\{y\})} 1\,dP\\
&= E(X|Y=y)\cdot P[Y=y]
\end{aligned}$$

$$\text{rechte Seite} = \int\limits_{Y^{-1}(\{y\})} X\,dP \;=\; \int\limits_{[Y=y]} X\,dP.$$

Wegen $P[Y=y]>0$ folgt daraus die Behauptung:

$$E(X|Y=y) = \frac{\int_{[Y=y]} X\,dP}{P[Y=y]}$$

2. Teil folgt daraus mit $X=\chi_A$. □

7.10 Satz. Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},P)$. Integrierbare ZV $X\colon (\Omega,\mathcal{A},P)\to(\overline{\mathbb{R}},\overline{\mathcal{B}})$ bzw. $A\in\mathcal{A}$. ZV $Y\colon(\Omega,\mathcal{A},P)\to(\Omega',\mathcal{A}')$.

a) $\underset{A'\in\mathcal{A}'}{\forall}\ \int_{A'} E(X\mid Y=y)\ P_Y(dy) = \int_{Y^{-1}(A')} X\,dP\,,$
insbesondere $\int_{\Omega'} E(X\mid Y=y)\ P_Y(dy) = EX\,.$

b) $\underset{A'\in\mathcal{A}'}{\forall}\ \int_{A'} P(A\mid Y=y)\ P_Y(dy) = P(Y^{-1}(A')\cap A)\,,$
insbesondere $\int_{\Omega'} P(A\mid Y=y)\ P_Y(dy) = P(A)\,.$

Beweis:

a) Sei $g = E(X|Y=y)$, d. h. $g\circ Y = E(X|Y)$.

Dann gilt:

$$\begin{aligned}
\int\limits_{Y^{-1}(A')} X\,dP &\overset{Def. 7.2}{=} \int\limits_{Y^{-1}(A')} g\circ Y\,dP\\
&= \int_{A'} g\,dP_Y\\
&\ \text{(nach dem Transformationssatz)}\\
&= \int_{A'} E(X|Y=y)\,dP_Y(y).
\end{aligned}$$

b) Folgt mit $X = \chi_A$ aus a). $\square$

Beispiel:
In einem zweistufigen Experiment wird rein zufällig ein Punkt $x \in [0,1]$ ausgewählt und dann ein n-Tupel von Bernoulli-Experimenten mit jeweiliger Erfolgswahrscheinlichkeit x durchgeführt. Wie viele Erfolge treten im Mittel auf?
Mit
$Y =$ Anzahl Erfolge,
$X =$ zufällige Erfolgswahrscheinlichkeit, gleichverteilt auf $[0,1]$
gilt:

$$EY \overset{Satz 7.10\,a)}{=} \int E[Y|X=x]\, dP_X(x)$$

wobei

$$E[Y|X=x] = \text{ Erwartungswert zu } b(n,x) = n \cdot x$$

ist, also ist

$$EY = \int_0^1 n \cdot x \, dx = \frac{n}{2}.$$

7.11 Satz. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, integrierbare ZVen $X, X_1, X_2 \colon (\Omega, \mathcal{A}, P) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$, ZV $Y \colon (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$, $\alpha, \beta \in \mathbb{R}$.

a) $X = c$ f.s. $\Longrightarrow E(X \mid Y = \cdot) = c \;\; P_Y$-f.ü.

b) $X \geq 0$ f.s. $\Longrightarrow E(X \mid Y = \cdot) \geq 0 \;\; P_Y$-f.ü.

c) $E(\alpha X_1 + \beta X_2 \mid Y = \cdot) = \alpha E(X_1 \mid Y = \cdot) + \beta E(X_2 \mid Y = \cdot) \;\; P_Y$-f.ü.

d) $X_1 \leq X_2$ f.s. $\Longrightarrow E(X_1 \mid Y = \cdot) \leq E(X_2 \mid Y = \cdot) \;\; P_Y$-f.ü.

Beweis: Folgt aus Satz 7.3 und Definition 7.9, z.B. in a):

$$X = c \; P - f.s. \;\Longrightarrow\; E(X \mid Y = \cdot) \circ Y = c \; P - f.s. \;\Longrightarrow\; E(X \mid Y = \cdot) = c \; P_Y - \text{f.ü.}$$

$\square$

7.12 Satz (Jensensche Ungleichung für bedingte Erwartungen).
Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $\mathcal{C} \subset \mathcal{A}$ eine σ-Algebra, $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B})$ eine reelle Zufallsvariable und sei $f : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ konvex, d. h.

$$f(\alpha x + (1-\alpha) y) \leq \alpha \cdot f(x) + (1-\alpha) \cdot f(y)$$

für alle $x, y \in \mathbb{R}$ und alle $\alpha \in [0,1]$, mit X und $f(X)$ integrierbar. Dann gilt:

$$f(E(X|\mathcal{C})) \leq E(f(X)|\mathcal{C}) \quad f.s..$$

Beweis im Spezialfall f zweimal differenzierbar.
Da f konvex ist, gilt dann

$$f''(x) \geq 0 \quad (x \in \mathbb{R}),$$

woraus für $x, y \in \mathbb{R}$ nach dem Mittelwertsatz und der Monotonie von f' folgt:

$$f(x) - f(y) = f'(\xi_{x,y}) \cdot (x - y) \geq f'(y) \cdot (x - y)$$

(denn die Beziehung ist klar für $x \geq y$, während für $x < y$ $f'(\xi_{x,y}) \leq f'(y)$ und $x - y < 0$, also wieder

$$f'(\xi_{x,y}) \cdot (x - y) \geq f'(y)(x - y)$$

gilt). Also gilt

$$f(x) \geq f(y) + f'(y) \cdot (x - y) \quad (x, y \in \mathbb{R}).$$

Mit der Stetigkeit von f und f' folgt daraus

$$(*) \qquad f(x) = \sup_{y \in \mathbb{Q}} \left(f(y) + f'(y) \cdot (x - y) \right)$$

und

$$f(X) \geq f(y) + f'(y) \cdot (X - y),$$

woraus folgt

$$E(f(X)|\mathcal{C}) \geq f(y) + f'(y) \cdot (E(X|\mathcal{C}) - y) \;\; f.s.$$

für alle $y \in \mathbb{R}$.
Also gilt P-f.s.:

$$\begin{aligned} E(f(X)|\mathcal{C}) &\geq \sup_{y \in \mathbb{Q}} \left(f(y) + f'(y) \cdot (E(X|\mathcal{C}) - y) \right) \\ &\overset{(*)}{=} f(E(X|\mathcal{C})). \end{aligned}$$

□

Beispiel: Mit X ist auch $E(X|\mathcal{C})$ quadratisch integriertbar, da

$$E((E(X|\mathcal{C}))^2) \overset{Satz\ 7.12}{\leq} E(E(X^2|\mathcal{C})) \overset{Def.7.1}{=} E(X^2).$$

8 Martingale

Im diesem Abschnitt leiten wir hinreichende Bedingungen für die Gültigkeit von

$$\frac{1}{n} \sum_{k=1}^{n} (X_k - \mathbf{E}X_k) \to 0 \quad f.s.$$

her. Dabei verwenden wir die folgende deterministische Aussage (Lemma von Kronecker, s.u.):

$$\left(\sum_{k=1}^{n} \frac{x_k}{k}\right)_{n\in\mathbb{N}} \text{ konvergent } \quad \Longrightarrow \quad \frac{1}{n}\sum_{k=1}^{n} x_k \to 0 \quad (n \to \infty),$$

nach der die obige Aussage aus der fast sicheren Konvergenz der Folge

$$\left(\sum_{k=1}^{n} \frac{X_k - \mathbf{E}X_k}{k}\right)_{n\in\mathbb{N}}$$

folgt. Letzteres weisen wir mit Hilfe der sogenannten Martingaltheorie her.

Zur Motivation des Martingalbegriffes dient das folgende Beispiel:

Beispiel: Wir betrachten ein Spiel mit zufälligen (integrierbaren) Gewinnständen X_n in $\overline{\mathbb{R}}$ zur Zeit $n \in \mathbb{N}$. Damit wir dieses als *fair* ansehen, sollte gelten:

1. $EX_1 = 0$.
2. $E(X_{n+1}|X_1 = x_1, \dots, X_n = x_n) = x_n$ für $P_{(X_1,\dots,X_n)}$-f.a. $(x_1, \dots, x_n)$ und alle $n \in \mathbb{N}$.

Letzteres ist äquivalent zu

$$E(X_{n+1}|X_1, \dots, X_n) = X_n \quad f.s.$$

8.1 Definition. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$.
Eine Folge $(X_n)_{n\in\mathbb{N}}$ von integrierbaren ZVn X_n: $(\Omega, \mathcal{A}, P) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ heißt bei gegebener nonoton wachsender Folge $(\mathcal{A}_n)_{n\in\mathbb{N}}$ von σ-Algebren $\mathcal{A}_n \subset \mathcal{A}$ mit $\mathcal{A}_n$-$\overline{\mathcal{B}}$-Messbarkeit von X_n [wichtiger Fall $\mathcal{A}_n = \mathcal{F}(X_1, \dots, X_n)$ $(n \in \mathbb{N})$]

a) ein **Martingal** bzgl. $(\mathcal{A}_n)$, wenn

$$\underset{n\in\mathbb{N}}{\forall} \; E(X_{n+1} \mid \mathcal{A}_n) = X_n \text{ f.s.}$$

$$[\text{d.h.} \quad \underset{n\in\mathbb{N}}{\forall} \; \underset{C\in\mathcal{A}_n}{\forall} \; \int_C X_{n+1}\, dP = \int_C X_n\, dP],$$

b) ein **Submartingal** bzgl. $(\mathcal{A}_n)$, wenn

$$\underset{n\in\mathbb{N}}{\forall} \; E(X_{n+1} \mid \mathcal{A}_n) \geq X_n \text{ f.s.}$$

$$[\text{d.h.} \quad \underset{n\in\mathbb{N}}{\forall} \; \underset{C\in\mathcal{A}_n}{\forall} \; \int_C X_{n+1}\, dP \geq \int_C X_n\, dP],$$

[denn letzteres ist äquivalent zu $\forall n \in \mathbb{N}\ \forall C \in \mathcal{A}_n : \int\limits_C (E(X_{n+1}|\mathcal{A}_n) - X_n)dP \geq 0$, was für $\mathcal{A}_n$-messbare Integranden wiederum äquivalent ist zu

$$E(X_{n+1}|\mathcal{A}_n) - X_n \geq 0 \; f.s.]$$

c) ein **Supermartingal** bzgl. $(\mathcal{A}_n)$, wenn $(-X_n)$ ein Submartingal bzgl. $(\mathcal{A}_n)$ ist.

8.2 Bemerkung. Ein Martingal (X_n) bzgl. $(\mathcal{A}_n)$ ist auch ein Martingal bzgl. $(\mathcal{F}(X_1, \ldots, X_n))$. Entsprechend für Sub-, Supermartingal.

Begründung (für Martingale, für Sub- bzw. Supermartingale analog): Meßbarkeit ist klar. Aus $\mathcal{F}(X_1) \subset \mathcal{A}_1, \mathcal{F}(X_2) \subset \mathcal{A}_2, \ldots$ und $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \ldots$ (was wegen der Messbarkeit der X_i gilt) folgt $\mathcal{F}(X_1, \ldots, X_n) \subset \mathcal{A}_n$, und daher gilt für alle $C \in \mathcal{F}(X_1, \ldots X_n)$:

$$\int\limits_C X_{n+1} dP = \int\limits_C X_n dP.$$

□

8.3 Satz. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, Folge $(V_n)_{n\in\mathbb{N}}$ von Zufallsvariablen V_n: $(\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B})$. Die Partialsummenfolge $(\sum\limits_{j=1}^{n} V_j)_{n\in\mathbb{N}}$ ist genau dann ein Martingal bzw. Submartingal bzgl. $(\mathcal{F}(V_1, V_1 + V_2, \ldots, V_1 + \ldots + V_n)) = (\mathcal{F}(V_1, \ldots, V_n))$, wenn

$$\underset{n\in\mathbb{N}}{\forall} \; V_n \text{ integrierbar und } \underset{n\in\mathbb{N}}{\forall} \; E(V_{n+1} \mid V_1, \ldots, V_n) = 0 \text{ bzw. } \geq 0 \text{ f.s.}$$

Beweis für Martingale:
Es gilt $\mathcal{F}(V_1, \ldots, V_n) = \mathcal{F}(V_1, V_1 + V_2, \ldots, V_1 + \ldots + V_n)$ da $V_1, V_1 + V_2, \ldots, V_1 + \ldots + V_n$ messbar bzgl. $\mathcal{F}(V_1, \ldots, V_n)$ und da $V_1, V_2 = (V_1 + V_2) - V_1, \ldots, V_n = (V_1 + \ldots + V_n) - (V_1 + \ldots + V_{n-1})$ messbar bzgl. $\mathcal{F}(V_1, V_1 + V_2, \ldots, V_1 + \ldots + V_n)$.
Also ist die Messbarkeitsbeziehung klar. Ebenfalls klar ist, dass $V_1, \ldots, V_n$ integrierbar äquivalent ist zu $V_1, V_1 + V_2, \ldots, V_1 + \ldots + V_n$ integrierbar ist.
Also genügt es zu zeigen:

$$E\left(\sum_{j=1}^{n+1} V_j | \mathcal{F}_n\right) = \sum_{j=1}^{n} V_j \text{ f. s. } \Leftrightarrow E(V_{n+1}|\mathcal{F}_n) = 0 \text{ f.s.},$$

wobei $\mathcal{F}_n = \mathcal{F}(V_1, \ldots, V_n)$. Dies folgt aber aus

$$\begin{aligned} E\left(\sum_{j=1}^{n+1} V_j | \mathcal{F}_n\right) &= E(V_{n+1}|\mathcal{F}_n) + E\left(\sum_{j=1}^{n} V_j | \mathcal{F}_n\right) \text{ f.s.} \\ &\quad \text{(nach Satz 7.3 d))} \\ &= E(V_{n+1}|\mathcal{F}_n) + \sum_{j=1}^{n} V_j \text{ f. s.} \\ &\quad \text{(da } \sum_{j=1}^{n} V_j \; \mathcal{F}_n - \text{messbar ist} \\ &\quad \text{folgt dies aus Satz 7.3 f)).} \end{aligned}$$

$\square$

Beispiel für ein Martingal:
Sind $V_1, V_2, \ldots$ **unabhängige** integrierbare reelle Zufallsvariablen mit $EV_n = 0$ $(n \in \mathbb{N})$, so ist

$$\left(\sum_{j=1}^{n} V_j\right)_{n\in\mathbb{N}} \quad \text{ein Martingal bzgl. } (\mathcal{F}(V_1, \ldots, V_n))_{n\in\mathbb{N}}.$$

Begründung: Wegen $V_{n+1}, (V_1, \ldots, V_n)$ unabhängig gilt nach Bemerkung 7.7:

$$E(V_{n+1}|V_1, \ldots, V_n) = E(V_{n+1}) \overset{Voraussetzung}{=} 0.$$

Die Behauptung folgt mit Satz 8.3. $\square$

8.4 Satz. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, Folge (V_n) von **quadratisch integrierbaren** ZVn V_n: $(\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B})$. Die Partialsummenfolge $(\sum_{j=1}^{n} V_j)$ sei ein Martingal.

Dann gilt für alle $i \neq j$: $E(V_i \cdot V_j) = 0$ ("Orthogonalitätseigenschaft")

Beweis: *oBdA* $j > i$.
$V_i \cdot V_j$ integrierbar, da

$$E|V_i \cdot V_j| \overset{Cauchy-Schwarz}{\leq} \sqrt{EV_i^2} \cdot \sqrt{EV_j^2} < \infty.$$

Mit Satz 7.3 folgt

$$\begin{aligned} E(V_i \cdot V_j) &= E(E(V_i \cdot V_j|V_1, \ldots, V_{j-1})) \\ &\quad \text{(nach Satz 7.3. a))} \\ &= E(V_i \cdot E(V_j|V_1, \ldots, V_{j-1})) \\ &\quad \text{(nach Satz 7.3 g), anwendbar} \\ &\quad \text{da } V_i \cdot V_j \text{ und } V_i \text{ integrierbar)} \\ &= 0 \end{aligned}$$

da nach Satz 8.3 gilt $E(V_j|V_1, \ldots, V_{j-1}) = 0$. $\square$

8.5 Satz ((Sub)Martingalkonvergenztheorem (Doob)). Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. (X_n) sei ein Submartingal mit $\overline{\lim}\, E|X_n| < \infty$.
Dann existiert eine integrierbare reelle ZV X mit $X_n \to X$ P-f.s.

Beweis: Siehe Kapitel 11. $\square$

8.6 Korollar. Ist (U_n) eine Folge integrierbarer nichtnegativ-reeller Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ und $(\mathcal{A}_n)$ eine monoton wachsende Folge von Unter-σ-Algebren von $\mathcal{A}$ mit $\mathcal{A}_n$-$\mathcal{B}_+$-Messbarkeit von U_n $(n \in \mathbb{N})$ und gilt

$$E(U_{n+1} \mid \mathcal{A}_n) \leq (1 + \alpha_n)U_n + \beta_n \quad (n \in \mathbb{N})$$

wobei $\alpha_n,\ \beta_n \in \mathbb{R}_+$ mit $\sum_{n=1}^\infty \alpha_n < \infty$, $\sum_{n=1}^\infty \beta_n < \infty$, dann konvergiert (U_n) f.s. — (U_n) ist "fast ein nichtnegatives Supermartingal".
Auch (EU_n) konvergiert.

Beweis: siehe Übungen. □

8.7 Satz. Folge (V_n) von quadratisch integrierbaren reellen ZVn mit $\sum\limits_{n=1}^{\infty} V(V_n) < \infty$. Dann ist

$$\left(\sum_{n=1}^{N}(V_n - E(V_n \mid V_1, \dots, V_{n-1}))\right)_{N\in\mathbb{N}} \quad \text{f.s. konvergent.}$$

[Falls zusätzlich (V_n) unabhängig, dann ist $\left(\sum\limits_{n=1}^{N}(V_n - EV_n)\right)_{N\in\mathbb{N}}$ f.s. konvergent]

Beweis:
oBdA $EV_n = 0,$ sonst V_n durch $V_n - EV_n$ ersetzen.
Setze

$$W_n = V_n - E(V_n|V_1, \dots, V_{n-1}) \quad (n > 1),$$
$$W_1 = V_1 - EV_1 = V_1$$

und

$$S_n = \sum_{j=1}^{n} W_j$$

Zu zeigen ist: $(S_n)_n$ ist P-f.s. konvergent.
Behauptung folgt aus Submartingalkonvergenztheorem, sofern wir zeigen:

1. $(S_n)_n$ ist Martingal bzgl. $(\mathcal{F}(V_1, \dots, V_n))_n$.
2. $\limsup\limits_{n\to\infty} E\{|S_n|\} < \infty$.

Nachweis von (1):
Mit V_n ist auch W_n integrierbar und es gilt

$$\begin{aligned}
E(W_{n+1}|V_1, \dots, V_n) &= E(V_{n+1}|V_1, \dots, V_n) - E(E(V_{n+1}|V_1, \dots, V_n)|V_1, \dots, V_n) \\
&\quad \text{(nach Satz 7.3 d))} \\
&= E(V_{n+1}|V_1, \dots, V_n) - E(V_{n+1}|V_1, \dots, V_n) \\
&\quad \text{(nach Satz 7.3 f))} \\
&= 0.
\end{aligned}$$

Wegen $\mathcal{F}(V_1, \dots, V_n) = \mathcal{F}(W_1, \dots, W_n)$ folgt daraus auch $E(W_{n+1}|W_1, \dots, W_n) = 0$, und (1) ergibt sich mit Satz 8.3.

Nachweis von (2):

Es gilt

$$E\{|S_n|\} \leq \sqrt{1} \cdot \sqrt{ES_n^2}$$

und

$$E(S_n^2) = E\left(\left(\sum_{j=1}^n W_j\right)^2\right) = \sum_{j=1}^n EW_j^2 + \sum_{\substack{i \neq j, \\ i,j \in \{1,\ldots,n\}}} E(W_i \cdot W_j).$$

Für $j > 1$ gilt

$$\begin{aligned} E(W_j^2) &\leq 2E(V_j^2) + 2 \cdot E(|E(V_j|V_1, \ldots, V_{j-1})|^2) \\ &\leq 2E(V_j^2) + 2 \cdot E(E(V_j^2|V_1, \ldots, V_{j-1})) \\ &\quad \text{(nach der Jensenschen Ungleichung für bedingte Erwartungen)} \\ &= 4 \cdot E(V_j^2) < \infty, \end{aligned}$$

also ist W_j quadratisch integrierbar und mit Satz 8.4 folgt

$$\begin{aligned} E(S_n^2) &= \sum_{j=1}^n E(W_j^2) + 0 \\ &\leq EV_1^2 + \sum_{j=2}^n 4 \cdot E(V_j^2) \\ &\leq 4 \cdot \sum_{j=1}^n V(V_j) \\ &\quad \text{(da } EV_j = 0 \text{ nach Annahme zu Beginn des Beweises)} \\ &\leq 4 \cdot \sum_{j=1}^\infty V(V_j) < \infty. \end{aligned}$$

Insgesamt folgt

$$\limsup_{n \to \infty} E|S_n| \leq 2 \cdot \sqrt{\sum_{j=1}^\infty V(V_j)} < \infty, w.z.z.w.$$

□

8.8 Satz. Folge (V_n) von quadratisch integrierbaren reellen ZVn mit

$$\sum_{n=1}^\infty n^{-2} V(V_n) < \infty.$$

Dann

$$\frac{1}{n} \sum_{j=1}^n (V_j - E(V_j \mid V_1, \ldots, V_{j-1})) \to 0 \quad \text{f.s.}$$

Falls zusätzlich (V_n) unabhängig, dann $\frac{1}{n} \sum_{j=1}^n (V_j - EV_j) \to 0$ f.s.

(Kriterium von Kolmogorov zum starken Gesetz der großen Zahlen)

Beweis von Satz 8.8 mit Satz 8.7 und dem folgenden — etwas spezialisierten —
Lemma von Kronecker (Spezialfall)

Ist $(c_n)_n$ eine reelle Folge mit $\sum\limits_{n=1}^{\infty} \frac{c_n}{n}$ konvergent, so gilt

$$\frac{c_1 + \ldots + c_n}{n} \to 0 \quad (n \to \infty).$$

Beweis:
Setze

$$s_n = \sum_{j=1}^{n} \frac{c_j}{j} \text{ und } s = \lim_{n\to\infty} s_n.$$

Dann gilt

$$\frac{1}{n+1} \sum_{j=1}^{n+1} c_j = s_{n+1} - \frac{s_1 + \ldots + s_n}{n+1},$$

was man leicht mit vollständiger Induktion zeigt:
Für $n = 1$ ist nämlich

$$s_2 - \frac{s_1}{2} = c_1 + \frac{c_2}{2} - \frac{c_1}{2} = \frac{1}{2} \cdot (c_1 + c_2),$$

und der Schluss von n auf $n+1$ folgt aus

$$\begin{aligned}
&s_{n+2} - \tfrac{s_1+\ldots+s_{n+1}}{n+2} = s_{n+2} - \tfrac{1}{n+2} \cdot (s_1 + \ldots + s_n + s_{n+1}) \\
&= s_{n+2} - \tfrac{1}{n+2} \cdot \Big((n+1) \cdot \big(\tfrac{s_1+\ldots+s_n}{n+1} - s_{n+1}\big) + (n+1) \cdot s_{n+1} + s_{n+1}\Big) \\
&= s_{n+2} - \tfrac{1}{n+2} \cdot \left((n+1) \cdot \left(-\tfrac{1}{n+1} \sum_{j=1}^{n+1} c_j\right) + (n+2) \cdot s_{n+1}\right) \\
&\quad \text{(nach Induktionsvoraussetzung)} \\
&= s_{n+2} - s_{n+1} + \tfrac{1}{n+2} \sum_{j=1}^{n+1} c_j \\
&= \sum_{j=1}^{n+2} \tfrac{c_j}{j} - \sum_{j=1}^{n+1} \tfrac{c_j}{j} + \tfrac{1}{n+2} \cdot \sum_{j=1}^{n+1} c_j \\
&= \tfrac{1}{n+2} \sum_{j=1}^{n+2} c_j.
\end{aligned}$$

Da $s_n \to s$ $(n \to \infty)$ auch $\frac{s_1+\ldots+s_n}{n+1} \to s$ $(n \to \infty)$ impliziert, folgt daraus aber

$$\lim_{n+\infty} \frac{c_1 + \ldots + c_n}{n} = \lim_{n\to\infty} \left(s_{n+1} - \frac{s_1 + \ldots + s_n}{n+1}\right) = s - s = 0.$$

$\square$

Beweis von Satz 8.8:
Aus Satz 8.7 (angewendet mit $\frac{V_n}{n}$ statt V_n) folgt

$$\left(\sum_{k=1}^{n} \left(\frac{V_k - E(V_k|V_1, \ldots, V_{k-1})}{k}\right)\right)_{n\in\mathbb{N}}$$

ist P-f.s. konvergent. Mit dem Lemma von Kronecker folgt daraus

$$\frac{1}{n}\sum_{k=1}^{n}(V_k - E(V_k|V_1 \ldots, V_{k-1})) \to 0 \ (n \to \infty) \text{ P-f.s.}$$

□

8.9 Bemerkung. Aus Satz 8.7 bzw. 8.8 ergibt sich unmittelbar für eine Folge (V_n) quadratisch integrierbarer reeller ZVn eine hinreichende Bedingung für die f.s. Konvergenz der Reihe $\sum V_n$ bzw. für $\frac{1}{n}\sum\limits_{j=1}^{n} V_j \to 0$ f.s.

8.10 Definition. Eine Folge $(X_n)_{n\in\mathbb{N}}$ von integrierbaren reellen ZVn auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ genügt dem **schwachen** bzw. **starken Gesetz der großen Zahlen**, wenn

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - EX_k) \to 0 \quad (n \to \infty) \text{ nach Wahrscheinlichkeit bzw. P-f.s.}$$

8.11 Satz (Kriterium von Kolmogorov für das starke Gesetz der großen Zahlen).
Eine unabhängige Folge $(X_n)_{n\in\mathbb{N}}$ quadratisch integrierbarer reeller ZVn mit

$$\sum_{n=1}^{\infty} n^{-2} V(X_n) < \infty$$

genügt dem starken Gesetz der großen Zahlen.

Beweis: Wegen

$$E(X_j|X_1, \ldots, X_{j-1}) = EX_j \text{ f.s}$$

(gilt nach der Unabhängigkeit von X_j und $(X_1, \ldots, X_{j-1})$) folgt die Behauptung aus Satz 8.8. □

8.12 Satz (Kolmogorovsches starkes Gesetz der großen Zahlen).
Für eine unabhängige Folge $(X_n)_{n\in\mathbb{N}}$ identisch verteilter integrierbarer reeller ZVn gilt

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to EX_1 \quad (n \to \infty) \text{ f.s.}$$

Beweis: Setze

$$Y_n = X_n \cdot 1_{\{|X_n| \le n\}} \quad (n \in \mathbb{N}).$$

Wegen

$$\frac{1}{n}\sum_{k=1}^{n} X_k = \frac{1}{n}\sum_{k=1}^{n} X_k \cdot 1_{\{|X_k| > k\}} + \frac{1}{n}\sum_{k=1}^{n}(Y_k - \mathbf{E}Y_k) + \frac{1}{n}\sum_{k=1}^{n}\mathbf{E}Y_k$$

genügt es zu zeigen:

$$\frac{1}{n}\sum_{k=1}^{n} X_k \cdot 1_{\{|X_k|>k\}} \to 0 \quad f.s., \tag{1}$$

$$\frac{1}{n}\sum_{k=1}^{n}(Y_k - \mathbf{E}Y_k) \to 0 \quad f.s., \tag{2}$$

und

$$\frac{1}{n}\sum_{k=1}^{n} \mathbf{E}Y_k \to \mathbf{E}X_1 \quad (n \to \infty). \tag{3}$$

Nachweis von (1): Wegen

$$\begin{aligned}
\sum_{k=1}^{\infty} \mathbf{P}\{|X_k| > k\} &= \sum_{k=1}^{\infty} \mathbf{P}\{|X_1| > k\} \leq \sum_{k=1}^{\infty} \int_{k-1}^{k} \mathbf{P}\{|X_1| > t\}\, dt \\
&= \int_{0}^{\infty} \mathbf{P}\{|X_1| > t\}\, dt = \mathbf{E}(|X_1|) < \infty
\end{aligned}$$

gilt nach dem Lemma von Borel-Cantelli, dass mit Wahrscheinlichkeit Eins nur endlich viele der X_k betragsmäßig größer als k sind, woraus (1) folgt.

Nachweis von (2): Wegen dem Lemma von Kronecker genügt es dazu zu zeigen, dass

$$\left(\sum_{k=1}^{n} \frac{Y_k - \mathbf{E}Y_k}{k}\right)_{n\in\mathbb{N}}$$

mit Wahrscheinlichkeit Eins konvergent ist. Gemäß Satz 8.7 folgt dies wiederum aus

$$\begin{aligned}
\sum_{k=1}^{\infty} V\left(\frac{Y_k - \mathbf{E}Y_k}{k}\right) &= \sum_{k=1}^{\infty} \frac{1}{k^2} \cdot V\left(Y_k - \mathbf{E}Y_k\right) = \sum_{k=1}^{\infty} \frac{1}{k^2} \cdot V\left(Y_k\right) \leq \sum_{k=1}^{\infty} \frac{1}{k^2} \cdot \mathbf{E}\left(Y_k^2\right) \\
&= \sum_{k=1}^{\infty} \frac{1}{k^2} \cdot \mathbf{E}\left(X_k^2 \cdot 1_{\{|X_k|\leq k\}}\right) \\
&= \sum_{k=1}^{\infty} \frac{1}{k^2} \cdot \mathbf{E}\left(X_1^2 \cdot 1_{\{|X_1|\leq k\}}\right) \\
&= \sum_{k=1}^{\infty} \frac{1}{k^2} \cdot \left(0 + \sum_{j=1}^{k} \mathbf{E}\left(X_1^2 \cdot 1_{\{j-1<|X_1|\leq j\}}\right)\right) \\
&= \sum_{k=1}^{\infty} \sum_{j=1}^{k} \frac{1}{k^2} \cdot \mathbf{E}\left(X_1^2 \cdot 1_{\{j-1<|X_1|\leq j\}}\right) \\
&= \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} \frac{1}{k^2} \cdot \mathbf{E}\left(X_1^2 \cdot 1_{\{j-1<|X_1|\leq j\}}\right) \\
&= \sum_{j=1}^{\infty} \mathbf{E}\left(X_1^2 \cdot 1_{\{j-1<|X_1|\leq j\}}\right) \cdot \sum_{k=j}^{\infty} \frac{1}{k^2}
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{j=1}^{\infty} \mathbf{E}\left(X_1^2 \cdot 1_{\{j-1<|X_1|\leq j\}}\right) \cdot \left(\frac{1}{j^2} + \sum_{k=j+1}^{\infty}\left(\frac{1}{k-1} - \frac{1}{k}\right)\right) \\
&= \sum_{j=1}^{\infty} \mathbf{E}\left(X_1^2 \cdot 1_{\{j-1<|X_1|\leq j\}}\right) \cdot \frac{2}{j} \\
&\leq 2 \cdot \sum_{j=1}^{\infty} \mathbf{E}\left(|X_1| \cdot 1_{\{j-1<|X_1|\leq j\}}\right) = 2\mathbf{E}(|X_1|) < \infty.
\end{aligned}
$$

Nachweis von (3): $\mathbf{E}|X_1| < \infty$ impliziert $|X_1| < \infty$ $f.s.$, woraus

$$X_1 \cdot 1_{\{|X_1|\leq k\}} \to X_1 \quad f.s.$$

folgt. Mit dem Satz von der majorisierten Konvergenz (wobei $|X_1|$ die integrierbare Majorante ist) folgt daraus

$$\mathbf{E}Y_k = \mathbf{E}\left\{X_1 \cdot 1_{\{|X_1|\leq k\}}\right\} \to \mathbf{E}X_1 \quad (n \to \infty),$$

was (3) impliziert. □

8.13 Bemerkung.

a) In Satz 8.12 darf die Integrierbarkeitsvoraussetzung nicht weggelassen werden.

b) Die Voraussetzung der Unabhängigkeit in Satz 8.12 kann zur Voraussetzung der paarweisen Unabhängigkeit abgeschwächt werden (N. Etemadi, Z. Wahrscheinlichkeitstheorie verw. Gebiete 55 (1981), 119-122).

8.14 Satz (Tschebyschev). Eine Folge $(X_n)_{n\in\mathbb{N}}$ quadratisch integrierbarer paarweise unkorrelierter [d.h. es gilt $\underset{j\neq k}{\forall}$ $E(X_j - EX_j)(X_k - EX_k) = 0$] reeller ZVn mit

$$n^{-2}\sum_{k=1}^{n} V(X_k) \to 0 \ \ (n \to \infty)$$

genügt dem schwachen Gesetz der großen Zahlen.

Beweis: Siehe Übungen

8.15 Bemerkung (Folgerung aus obigen Sätzen). Für eine unabhängige Folge (X_n) identisch verteilter reeller ZVn mit $P[X_1 = 1] = p$, $P[X_1 = 0] = 1 - p$ (festes $p \in [0, 1]$) gilt

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to p \quad (n \to \infty) \text{ P-f.s. und nach Wahrscheinlichkeit}$$

(Borelsches starkes Gesetz der großen Zahlen bzw. Bernoullisches schwaches Gesetz der großen Zahlen).

9 Verteilungskonvergenz in $\mathbb{R}$

9.1 Definition:

a) W-Raum $(\Omega, \mathcal{A}, P), X_n, X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B})$ $(n \in \mathbb{N}), p \geq 1$. Man sagt: X_n **konvergiert gegen** X

a_1) **fast sicher** (kurz: $X_n \to X$ f.s.)
$:\Leftrightarrow$ $P(\{\omega \in \Omega : \lim\limits_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$.

a_2) **im p-ten Mittel** (kurz: $X_n \to X$ im p-ten Mittel)
$:\Leftrightarrow$ $E\{|X_n - X|^p\} \to 0$ $(n \to \infty)$

a_3) **nach Wahrscheinlichkeit** (kurz: $X_n \to^P X$)
$:\Leftrightarrow$ $\forall \varepsilon > 0 :$ $P[|X_n - X| > \varepsilon] \to 0$ $(n \to \infty)$

a_4) **nach Verteilung** (kurz: $X_n \to^{\mathcal{D}} X$)
$:\Leftrightarrow$ Für jede beschränkte und stetige Funktion $g : \mathbb{R} \to \mathbb{R}$ gilt:

$$Eg(X_n) \to Eg(X) \ (n \to \infty)$$

b) W-Maße Q_n, Q $(n \in \mathbb{N})$ auf $\mathcal{B}$. Man sagt: Q_n **konvergiert gegen** Q **schwach** (kurz: $Q_n \to Q$ schwach)
$:\Leftrightarrow$ Für jede beschränkte und stetige Funktion $g : \mathbb{R} \to \mathbb{R}$ gilt:
$\int\limits_{\mathbb{R}} g \, dQ_n \to \int\limits_{\mathbb{R}} g \, dQ$ $(n \to \infty)$.

9.2 Bemerkung.

a) Zwischen den Konvergenzbegriffen bestehen die folgenden Zusammenhänge:

- $X_n \to X$ $f.s.$ $\Rightarrow$ $X_n \to^P X$, denn nach dem Satz von der majorisierten Konvergenz folgt aus $X_n \to X$ $f.s.$:

$$P[|X_n - X| > \epsilon] = \int \chi_{\{|X_n - X| > \epsilon\}} \, dP \to 0 \quad (n \to \infty).$$

- $X_n \to X$ im p-ten Mittel $\Rightarrow$ $X_n \to^P X$ (nach der Ungleichung von Markov).
- $X_n \to^P X$ $\Rightarrow$ $X_n \to^{\mathcal{D}} X$ (vgl. Satz 9.3).
- $X_n \to^P X$ $\Rightarrow$ Es existiert Teilfolge $(X_{n_k})_k$ mit $X_{n_k} \to X$ $f.s.$ (vgl. Übungen).

b) Für die Konvergenz nach Verteilung (äquivalent: $P_{X_n} \to P_X$ schwach) müssen die Zufallsvariablen nicht auf dem gleichen W-Raum definiert sein.

c) $X_n \to^{\mathcal{D}} X$ ist nach Satz 9.4 äquivalent zu:
Für jeden Stetigkeitspunkt $x \in \mathbb{R}$ der Verteilungsfunktion F von X gilt

$$F_n(x) \to F(x)\ (n \to \infty),$$

wobei F_n die Verteilungsfunktion von X_n ist.

Die allgemeinere Definition oben lässt sich aber auf Zufallsvariablen mit Werten in (geeigneten) metrischen Räumen verallgemeinern.

d) Die Grenzverteilung bei Konvergenz nach Verteilung ist eindeutig (und damit auch bei allen anderen Konvergenzarten in Definition 9.1):

$$\left.\begin{array}{l} X_n \to^{\mathcal{D}} X^* \\ X_n \to^{\mathcal{D}} X^{**} \end{array}\right\} \Rightarrow^{c)} \left\{\begin{array}{l} F_n(x) \to F^*(x) \text{ in allen Stetigkeitsstellen von } F^* \\ F_n(x) \to F^{**}(x) \text{ in allen Stetigkeitsstellen von } F^{**} \end{array}\right.$$

wobei F^* bzw. F^{**} die Verteilungsfunktion von X^* bzw. X^{**} ist.

Da $\{x \in \mathbb{R} \mid x \text{ ist Unstetigkeitsstelle von } F^* \text{ bzw. } F^{**}\}$ abzählbar ist, folgt mit der rechtsseitigen Stetigkeit der Verteilungsfunktion: $F^*(x) = F^{**}(x)$ für alle $x \in \mathbb{R}$, woraus $P_{X^*} = P_{X^{**}}$ folgt.

e) In der Definition der Konvergenz nach Verteilung lassen sich X_n und X ersetzen durch X_n' und X', sofern $P_{X_n} = P_{X_n'}$ und $P_{X'} = P_X$.

f) Es gilt:

$$X_n \to^P X \quad \Leftrightarrow \quad \begin{array}{l} \text{Für jede Teilfolge } (n_k)_k \text{ von } (n)_n \\ \text{existiert Teilteilfolge } (n_{k_l})_l \text{ mit } X_{n_{k_l}} \to X \ f.s. \end{array}$$

(Vgl. Übungen.)

9.3 Satz. Reelle ZVn X_n $(n \in \mathbb{N})$, X auf $(\Omega, \mathcal{A}, P)$.
Aus $X_n \to X$ nach Wahrscheinlichkeit (Schreibweise $X_n \xrightarrow{P} X$) folgt $X_n \xrightarrow{\mathcal{D}} X$.

Beweis: Beliebiges stetiges und beschränktes $g : \mathbb{R} \to \mathbb{R}$. Zu zeigen:

$$Eg(X_n) \to Eg(X) \ \ (n \to \infty).$$

Äquivalent (bei Konvergenz reeller Folgen):
Für jede Teilfolge $(n_k)_k$ von $(n)_n$ existiert Teilteilfolge $(n_{k_l})_l$ mit

$$Eg(X_{n_{k_l}}) \to Eg(X) \ \ (l \to \infty).$$

Dazu: Sei $(n_k)_k$ beliebige Teilfolge von $(n)_n$.
Nach Bemerkung 9.2 f) existiert dann Teilteilfolge $(n_{k_l})_l$ mit

$$X_{n_{k_l}} \to X \quad f.s.$$

Da g stetig ist, folgt daraus

$$g(X_{n_{k_l}}) \to g(X) \;\; f.s.,$$

und der Satz von der majorisierten Konvergenz impliziert

$$Eg(X_{n_{k_l}}) \to Eg(X) \;\; (l \to \infty).$$

$\square$

9.4 Satz. W-Maße $Q_n \;\; (n \in \mathbb{N})$, Q auf $\mathcal{B}$ bzw. reelle ZVn $X_n \;\; (n \in \mathbb{N})$, X mit VFn F_n, F. $Q_n \to Q$ schwach bzw. $X_n \overset{\mathcal{D}}{\to} X \;\; (n \to \infty)$ gilt genau dann, wenn $F_n(x) \to F(x) \;\; (n \to \infty)$ für alle Stetigkeitspunkte x von F.

Beweisidee: Approximiere in

$$F_n(x) = \int \chi_{(-\infty,x]}(t)\, dP_{X_n}(t)$$

die Indikatorfunktion durch stetige und beschränkte Funktionen bzw. approximiere stetige und beschränkte Funktionen durch Linearkombinationen von Indikatorfunktionen. $\square$

9.5 Satz. Reelle ZVn $X_n \;\; (n \in \mathbb{N})$, X auf $(\Omega, \mathcal{A}, P)$. Ist X P-f.s. konstant, so gilt:

$$X_n \overset{\mathcal{D}}{\to} X \iff X_n \overset{P}{\to} X\,.$$

Beweis:
"$\Leftarrow$" Klar nach Satz 9.3.
"$\Rightarrow$" oBdA $\;\; X_n \to^{\mathcal{D}} c$, da Verteilung von X mit der von c übereinstimmt.
Sei F die Verteilungsfunktion von c, also

$$F(x) = \begin{cases} 0 & , \;\; x < c \\ 1 & , \;\; x \geq c \end{cases}$$

Nach Satz 9.4 gilt für die Verteilungsfunktion F_n von X_n:

$$F_n(x) \to F(x) \;\; (n \to \infty) \text{ für alle } x \neq c.$$

Daraus folgt für $\varepsilon > 0$:

$$\begin{aligned}
&P[|X_n - c| > \varepsilon] \\
&= P[X_n > c + \varepsilon] + P[X_n < c - \varepsilon] \\
&\leq P[X_n > c + \varepsilon] + P[X_n \leq c - \varepsilon] \\
&= 1 - F_n(c + \varepsilon) + F_n(c - \varepsilon) \\
&\overset{(n\to\infty)}{\longrightarrow} 1 - F(c + \varepsilon) - F(c - \varepsilon) = 1 - 1 - 0 = 0.
\end{aligned}$$

□

9.6 Satz. Reelle ZVn X_n, Y_n $(n \in \mathbb{N})$, X auf $(\Omega, \mathcal{A}, P)$

$$\left.\begin{array}{l} X_n \stackrel{\mathcal{D}}{\to} X \\ |X_n - Y_n| \stackrel{P}{\to} 0 \end{array}\right\} \Longrightarrow Y_n \stackrel{\mathcal{D}}{\to} X\,.$$

Beweis: Beliebiges beschränktes und stetiges $g : \mathbb{R} \to \mathbb{R}$.
Es genügt zu zeigen: $Eg(X_n) - Eg(Y_n) \to 0$ $(n \to \infty)$
oBdA $X_n - Y_n \to 0$ f.s. (sonst mit Teilteilfolgen arbeiten).
Sei $|g(x)| \leq \beta$ für $x \in \mathbb{R}$. Wähle $M > 0$ mit $-M$ und M sind Stetigkeitspunkte der Verteilungsfunktion F von X.
Dann gilt:

$$\begin{aligned} &|Eg(X_n) - Eg(Y_n)| \\ &\leq E\left[|g(X_n) - g(Y_n)| \cdot \chi_{[-M<X_n\leq M, |X_n-Y_n|\leq 1]}\right] \\ &\quad +2\beta \cdot P[X_n > M] + 2\beta \cdot P[X_n \leq -M] + 2\beta \cdot P[|X_n - Y_n| > 1] \end{aligned}$$

Da g auf $[-M-1, M+1]$ gleichmäßig stetig ist, folgt aus $X_n - Y_n \to 0$ f.s.

$$|g(X_n) - g(Y_n)| \cdot \chi_{[-M<X_n\leq M, |X_n-Y_n|\leq 1]} \to 0 \;\; f.s.$$

Mit der Beschränktheit von g folgt daraus mit dem Satz von der majorisierten Konvergenz

$$E\left[|g(X_n) - g(Y_n)| \cdot \chi_{[-M<X_n\leq M, |X_n-Y_n|\leq 1]}\right] \to 0 \; (n \to \infty).$$

Da $X_n \to^{\mathcal{D}} X$ gilt und $-M$ und M Stetigkeitspunkte der Verteilungsfunktion von X sind, gilt

$$P[X_n > M] = 1 - P[X_n \leq M] \stackrel{(n\to\infty)}{\longrightarrow} 1 - P[X \leq M] = P[X > M]$$

und

$$P[X_n \leq -M] \stackrel{(n\to\infty)}{\longrightarrow} P[X \leq -M].$$

Aus $X_n - Y_n \to^{P} 0$ folgt weiter

$$P[|X_n - Y_n| > 1] \to 0 \; (n \to \infty).$$

Insgesamt erhält man

$$\limsup_{n\to\infty} |Eg(X_n) - Eg(Y_n)| \leq 0 + 2\beta P[X > M] + 2\beta \cdot P[X \leq -M] + 0.$$

Mit $M \to \infty$ folgt die Behauptung. □

9.7 Satz (Spezialfall des **Darstellungssatzes von Skorokhod).**

Reelle ZVn X_n $(n \in \mathbb{N})$, X mit $X_n \xrightarrow{\mathcal{D}} X$. Dann existiert ein Wahrscheinlichkeitsraum $(\Omega^*, \mathcal{A}^*, P^*)$ — wobei man $\Omega^* = (0,1)$, $\mathcal{A}^* = (0,1) \cap \mathcal{B}$, $P^* =$ L-B-Maß auf $\mathcal{A}^*$ wählen kann — und reelle ZVn X_n^* $(n \in \mathbb{N})$, X^* auf $(\Omega^*, \mathcal{A}^*, P^*)$ derart, dass

$$\underset{n}{\forall}\ P_{X_n} = P_{X_n^*},\ \ P_X = P_{X^*},\ \ X_n^* \to X^* \quad (n \to \infty)\ P^*\text{-f.s.}$$

Beweis:
Sei $\Omega^* = (0,1), \mathcal{A}^* = (0,1) \cap \mathcal{B}$ und $P^* = LB\text{-Maß}\ |_{\mathcal{A}^*}$.
Sei F_n bzw. F die Verteilungsunktion zu X_n bzw. X.
Setze

$$X_n^*(\omega) = \min\{t \in \mathbb{R} : F_n(t) \geq \omega\}$$
$$X^*(\omega) = \min\{t \in \mathbb{R} : F(t) \geq \omega\}$$

für $\omega \in (0,1)$.
X_n^* bzw. X^* sind wohldefiniert, messbar (da monoton) und es gilt

$$P^*_{X_n^*} = P_{X_n} \text{ und } P^*_{X^*} = P_X$$

(vgl. Übungen)
X^* hat höchstens abzählbar viele Unstetigkeitsstellen (da X^* monotone Abbildung ist, ist die Anzahl der Spünge mit Sprunghöhe $> \varepsilon$ auf jedem kompakten Intervall endlich), diese bilden eine Nullmenge bzgl. des LB-Maßes.
Also genügt es zu zeigen:

$$X_n^*(\omega) \to X^*(\omega) = \ (n \to \infty) \text{ für jede Stetigkeitsstelle } \omega \text{ von } X^*.$$

Dazu: Sei ω Stetigkeitsstelle von X^*
Sei $\varepsilon > 0$ beliebig.
Dann existieren Stetigkeitspunkte x_1, x_2 von F (da diese dicht in $\mathbb{R}$ liegen) mit

$$(*) \qquad x_1 < X^*(\omega) < x_2 \quad \text{und} \quad x_2 - x_1 < \varepsilon.$$

Aus $x_1 < X^*(\omega) = \min\{t \in \mathbb{R} : F(t) \geq \omega\}$ folgt

$$F(x_1) < \omega.$$

Da ω Stetigkeitsstelle von X^* ist, ist F rechts von der Stelle $X^*(\omega)$ nicht konstant gleich ω. Daher folgt aus $X^*(\omega) < x_2 :\quad \omega < F(x_2)$.
Insgesamt gilt also:

$$F(x_1) < \omega < F(x_2).$$

Aus $X_n \to^{\mathcal{D}} X$ folgt

$$F_n(x_1) \to F(x_1) \ \ (n \to \infty),$$
$$F_n(x_2) \to F(x_2) \ \ (n \to \infty).$$

Also existiert n_0 so, dass für alle $n \geq n_0$ gilt:

$$F_n(x_1) < \omega < F_n(x_2).$$

Nach Definition von $X_n^*(\omega) = \min\{t \in \mathbb{R} : F_n(t) \geq \omega\}$ folgt daraus aber:

$$(**) \quad x_1 \leq X_n^*(\omega) \leq x_2.$$

Mit $(*)$ und $(**)$ ergibt sich für $n \geq n_0$:

$$|X_n^*(\omega) - X^*(\omega)| \leq x_2 - x_1 < \varepsilon,$$

woraus die Behauptung folgt. □

9.8 Satz (Satz von der stetigen Abbildung).
Reelle ZVn X_n $(n \in \mathbb{N})$, X mit $X_n \xrightarrow{\mathcal{D}} X$; Abbildung $h : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ sei P_X-f.ü. stetig. Dann gilt $h(X_n) \xrightarrow{\mathcal{D}} h(X)$.

9.9 Bemerkung. Für h stetig ist Satz 9.8 trivial.
Beweis von Satz 9.8:
Für

$$D = \{x \in \mathbb{R} : h \text{ unstetig in } x\} \in \mathcal{B}(!)$$

gilt

$$P_X(D) = 0.$$

Nach Satz 9.7 existiert W-Raum $(\Omega^*, \mathcal{A}^*, P^*)$ und Zven X_n^* und X^* auf $(\Omega^*, \mathcal{A}^*, P^*)$ mit

$$P^*_{X_n^*} = P_{X_n}, P^*_{X^*} = P_X \text{ und } X_n^*(\omega) \to X^*(\omega) \text{ für } P^*\text{-fast alle } \omega \in \Omega^*.$$

Daraus folgt

$$h(X_n^*(\omega)) \to h(X^*(\omega)) \; (n \to \infty)$$

für P^*-fast alle $\omega \in \Omega^* \backslash \{\varpi : X^*(\varpi) \in D\}$.
Wegen

$$P^*(\{\varpi \in \Omega^* : X^*(\varpi) \in D\}) = P^*_{X^*}(D) = P_X(D) = 0$$

gilt also

$$h(X_n^*) \to h(X^*) \;\; P^* - f.s.,$$

was

$$h(X_n^*) \to^{\mathcal{D}} h(X^*)$$

bzw.

$$P^*_{h(X_n^*)} \to P^*_{h(X^*)} \text{ schwach}$$

impliziert.

Wegen

$$P^*_{h(X_n^*)} = \left(P^*_{X_n^*}\right)_h \overset{s.o.}{=} (P_{X_n})_h = P_{h(X_n)}$$

und

$$P^*_{h(X^*)} = (P^*_{X^*})_h \overset{s.o.}{=} (P_X)_h = P_{h(X)}$$

(vergleiche Bemerkung 2.16) ist dies wiederum äquivalent zu

$$h(X_n) \to^{\mathcal{D}} h(X) \quad w.z.z.w.$$

□

9.10 Satz (Satz von Slutsky).
Reelle ZVn X_n, Y_n, X auf $(\Omega, \mathcal{A}, P)$; $c \in \mathbb{R}$.

$$\left.\begin{array}{l} X_n \overset{\mathcal{D}}{\to} X \\ Y_n \overset{P}{\to} c \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} X_n + Y_n \overset{\mathcal{D}}{\to} X + c \\ Y_n X_n \overset{\mathcal{D}}{\to} cX\,. \end{array}\right.$$

Beweis: Wir zeigen zunächst den ersten Teil: Aus $Y_n \to^P c$ folgt

$$|X_n - (X_n + Y_n - c)| = |c - Y_n| \to^P 0.$$

Mit Satz 9.6 folgt daraus

$$X_n + Y_n - c \to^{\mathcal{D}} X,$$

woraus mit Satz 9.8 (angewendet mit $(h(z) = z + c)$

$$X_n + Y_n \to^{\mathcal{D}} X + c$$

folgt.
Für den zweiten Teil wählen wir ein beliebiges stetiges und beschränktes $f : \mathbb{R} \to \mathbb{R}$ und zeigen

$$\mathbf{E}f(Y_n X_n) \to \mathbf{E}f(c \cdot X) \quad (n \to \infty).$$

Dazu beachten wir

$$\mathbf{E}f(Y_n X_n) - \mathbf{E}f(c \cdot X) = \left(\mathbf{E}f(Y_n X_n) - \mathbf{E}f(c \cdot X_n)\right) + \left(\mathbf{E}f(c \cdot X_n) - \mathbf{E}f(c \cdot X)\right),$$

wobei wegen $X_n \to^{\mathcal{D}} X$ und der Stetigkeit und Beschränktheit von $h(z) = f(c \cdot z)$ der Termn in der zweiten Klammer gegen Null konvergiert. Um auch noch zu zeigen, dass der Term in der ersten Klammer gegen Null konvergiert, betrachten wir ein beliebiges $\epsilon > 0$, wählen Stetigkeitspunkte K_1, K_2 von der Verteilungsfunktion von X mit $\mathbf{P}\{X \leq K_1\} < \epsilon$ und $\mathbf{P}\{X > K_2\} < \epsilon$ und setzen $M = \max\{|K_1|, |K_2|\} \cdot |c| + \epsilon$. Da f auf $[-M, M]$ gleichmäßig stetig ist, existieren $C \geq 1$ und $\delta > 0$ mit

$$|f(x) - f(z)| \leq C \cdot |x - z|$$

für alle $x, z \in [-M, M]$. Damit folgt

$$\begin{aligned}
&|\mathbf{E}f(Y_nX_n) - \mathbf{E}f(c \cdot X_n)| \\
&\leq \mathbf{E}|f(Y_nX_n) - f(c \cdot X_n)| \cdot I_{\{X_n \notin (K_1,K_2]\}} \\
&\quad + \mathbf{E}|f(Y_nX_n) - f(c \cdot X_n)| \cdot I_{\{|Y_n - c| > \epsilon/(C \cdot M)\}} \\
&\quad + \mathbf{E}|f(Y_nX_n) - f(c \cdot X_n)| \cdot I_{\{X_n \in (K_1,K_2], |Y_n - c| \leq \epsilon/(C \cdot M)\}} \\
&\leq 2 \cdot \|f\|_\infty \cdot \mathbf{P}\{X_n \notin (K_1, K_2]\} + 2 \cdot \|f\|_\infty \cdot \mathbf{P}\{|Y_n - c| > \epsilon/(C \cdot M)\} + \epsilon \\
&\to 2 \cdot \|f\|_\infty \cdot \mathbf{P}\{X \notin (K_1, K_2]\} + 0 + \epsilon \leq 2 \cdot \|f\|_\infty \cdot 2\epsilon + \epsilon.
\end{aligned}$$

Mit $\epsilon \to 0$ folgt die Behauptung. □

9.11 Definition. Eine Menge $\mathcal{Q}$ von W-Maßen auf $\mathcal{B}$ heißt

a) **relativ (folgen)kompakt**, wenn jede Folge von Elementen aus $\mathcal{Q}$ eine Teilfolge besitzt, die schwach gegen ein W-Maß auf $\mathcal{B}$ konvergiert,

b) **gleichmäßig straff**, wenn

$$\underset{\varepsilon > 0}{\forall} \quad \underset{\text{komp. } K \subset \mathbb{R}}{\exists} \quad \underset{Q \in \mathcal{Q}}{\forall} \quad Q(K^c) \leq \varepsilon\,.$$

9.12 Satz (Spezialfall des **Satzes von Prokhorov**).
Eine Menge $\mathcal{Q}$ von W-Maßen auf $\mathcal{B}$ ist genau dann relativ kompakt, wenn sie gleichmäßig straff ist.

Bemerkung: Eine Funktion $F : \mathbb{R} \to \mathbb{R}$ heißt maßdefinierende Funktion, wenn sie monoton wachsend und rechtsseitig stetig ist.
Man kann zeigen: Zu jeder maßdefinierenden Funktion F existiert genau ein Maß Q auf $\mathcal{B}$ mit

$$Q((a, b]) = F(b) - F(a) \quad \text{für alle } a, b \in \mathbb{R}, a < b$$

(vgl. Maßtheorie).
Weiter gilt:

9.13 Satz (Auswahlsatz von Helly).
Ist (F_n) eine Folge von Verteilungsfunktionen, dann existiert eine Indexfolge (n_i) und eine maßdefinierender Funktion $F : \mathbb{R} \to \mathbb{R}$ (nicht notwendigerweise eine Verteilungsfunktion) derart, dass $F_{n_i}(x) \to F(x)$ $(i \to \infty)$ für alle Stetigkeitspunkte x von F.

Bemerkung:
Hierbei ist $F(x) = 0$ $(x \in \mathbb{R})$ möglich, z. B. gilt für

$$F_n(x) = \begin{cases} 0, & x < n \\ 1, & x \geq n \end{cases}$$

(d.h. zugehöriges Maß ist auf den Punkt $\{n\}$ konzentriert):

$$F_n(x) \to 0 \ (n \to \infty) \quad \text{für alle } x \in \mathbb{R}$$

("Wegrutschen von Masse ins Unendliche").

Beweisskizze: Sei $Q = \{x_1, x_2, \ldots\}$.
Da jede beschränkte reelle Folge eine konvergente Teilfolge enthält, kann man rekursiv Teilfolgen $(n_{k+1,l})_l$ von $(n_{k,l})_l$ (mit $n_{0,l} = l \ (l \in \mathbb{N})$) konstruieren, für die

$$\left(F_{n_{k+1},l}(x_i)\right)_{l\in\mathbb{N}}$$

konvergent ist für $i = 1, 2, \ldots, k+1$.
Setzt man dann

$$n_k = n_{k,k},$$

so sieht man leicht, dass

$$(F_{n_k}(x_i))_{k\in\mathbb{N}} \text{ für alle } i \in \mathbb{N} \text{ konvergiert.}$$

Nun definiert man

$$F^* : \mathbb{Q} \to \mathbb{R}$$

durch

$$F^*(x) = \lim_{n\to\infty} F_{n_k}(x).$$

Man macht sich leicht klar, dass dann gilt:

1. $F^*(x) \in [0,1] \ \ (x \in \mathbb{Q})$,

2. F^* ist monoton wachsend.

Die Funktion $\quad F : \mathbb{R} \to \mathbb{R} \quad$ definiert durch

$$F(x) = \inf_{x<z, z\in\mathbb{Q}} F^*(z)$$

erfüllt dann die Behauptung, denn:

i) $F(x) \in [0,1] \quad (x \in \mathbb{R})$

II) F ist monoton wachsend

III) F ist rechtsseitig stetig (folgt aus Definition als Infimum!)

IV) Für jeden Stetigkeitspunkt $x \in \mathbb{R}$ von F gilt:

$$F_{n_k}(x) \to F(x) \ \ (k \to \infty).$$

Nachweis von IV): Für beliebiges $\varepsilon > 0$ wähle $x_1, x_2 \in \mathbb{Q}$ und $\delta > 0$ mit

$$x_1 \leq x \leq x_2 \text{ und } |F(x_2 + \delta) - F(x_1 - \delta)| < \varepsilon.$$

Dann gilt:

$$\begin{aligned}
& F_{n_k}(x) - F(x) \overset{\substack{Monotonie \\ von\ F_{n_k}}}{\leq} F_{n_k}(x_2) - F(x) \overset{(k\to\infty)}{\to} F^*(x_2) - F(x) \\
& \leq F(x_2 + \delta) - F(x) \\
& \overset{\substack{Monotonie \\ vonF}}{\leq} F(x_2 + \delta) - F(x_1 - \delta) < \varepsilon
\end{aligned}$$

und analog

$$F_{n_k}(x) - F(x) \geq F_{n_k}(x_1) - F(x) \overset{(k\to\infty)}{\to} F^*(x_1) - F(x) \geq F(x_1 - \delta) - F(x_2 + \delta) > -\varepsilon$$

Daraus folgt:

$$\limsup_{n\to\infty} |F_{n_k}(x) - F(x)| < \varepsilon.$$

□

Beweis von Satz 9.12

a) Wir zeigen:$\mathcal{Q}$ relativ kompakt $\Rightarrow$ $\mathcal{Q}$ gleichmäßig straff.

Sei $\mathcal{Q}$ relativ kompakt. Angenommen, $\mathcal{Q}$ ist nicht gleichmäßig straff. Dann gilt:

$$\exists \varepsilon > 0 \ \ \forall \text{ kompakten } K \subset \mathbb{R} \ \exists\, Q \in \mathcal{Q} : Q(K) < 1 - \varepsilon.$$

Ist $\varepsilon > 0$ entsprechend gewählt, so gilt also:

$$\forall n \in \mathbb{N} \ \ \exists\, Q_n \in \mathcal{Q} \ : \ Q_n([-n, n]) < 1 - \varepsilon.$$

Da $\mathcal{Q}$ relativ kompakt ist, existiert Indexteilfolge $(n_i)_i$ und W-Maß Q^* mit

$$Q_{n_i} \to Q^* \text{ schwach.}$$

Sei F^* die Verteilungsfunktion zu Q^*. Dann existieren Stetigkeitspunkte $a < b$ von F^* mit

$$F^*(b) - F^*(a) \geq 1 - \frac{\varepsilon}{2}.$$

Daraus folgt für die Verteilungsfunktionen F_{n_i} von Q_{n_i}

$$Q_{n_i}((a, b]) = F_{n_i}(b) - F_{n_i}(a) \overset{(i\to\infty)}{\longrightarrow} F^*(b) - F^*(a) \geq 1 - \frac{\varepsilon}{2},$$

was ein Widerspruch zu

$$\begin{aligned}
Q_{n_i}((a, b]) \ &\leq \ Q_{n_i}([-n_i, n_i]) \\
&\quad\ \text{(für } i \text{ genügend groß)} \\
&\overset{s.o.}{<} \ 1 - \varepsilon
\end{aligned}$$

ist.

b) Wir zeigen: $\mathcal{Q}$ gleichmäßig straff $\Rightarrow$ $\mathcal{Q}$ relativ kompakt.

Sie $(Q_n)_n$ beliebige Folge in $\mathcal{Q}$. Sei F_n die Verteilungsfunktion zu Q_n.

Nach dem Auswahlsatz von Helly existiert Indexteilfolge $(n_i)_i$ und maßdefinierende Funktion F mit $F_{n_i}(x) \to F(x)$ $(i \to \infty)$ in allen Stetigkeitspunkten x von F.

Sei Q das zu F gehörende Maß, d. h.

$$Q((a,b]) = F(b) - F(a) \text{ für alle } a, b \in \mathbb{R}, a < b$$

(vgl. Integrationstheorie bzw. Maßtheorie). Wegen $F_{n_i}(x) \in [0,1]$ gilt auch $F(x) \in [0,1]$ für alle Stetigkeitspunkte x von F, und da F monoton ist und diese dicht in $\mathbb{R}$ liegen folgt daraus sogar $F(x) \in [0,1]$ für alle $x \in \mathbb{R}$. Mit der Stetigkeit des Maßes Q von unten folgt

$$Q((-\infty, x]) = F(x) - \lim_{a \to -\infty} F(a)$$

bzw.

$$F(x) = Q((-\infty, x]) + c \quad (x \in \mathbb{R})$$

mit

$$c = \lim_{a \to -\infty} F(a) \quad \in [0,1].$$

Wir zeigen im Folgenden:

$$Q(\mathbb{R}) \geq 1. \qquad (*)$$

Daraus folgt die Behauptung, denn: Gilt $Q(\mathbb{R}) \geq 1$, so ist

$$1 \geq \lim_{x \to \infty} F(x) \overset{s.o.}{=} \lim_{x \to \infty} (Q((-\infty, x]) + c) = Q(\mathbb{R}) + c \geq 1 + c,$$

woraus mit $c \in [0,1]$ $c = 0$ und $Q(\mathbb{R}) = 1$ folgt, was wiederum Q W-Maß,

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

und damit auch

$$Q_n \to Q \quad \text{schwach}$$

impliziert.

Nachweis von $(*)$:

Sei $\varepsilon > 0$ beliebig.

Da $\mathcal{Q}$ gleichmäßig straff ist, existiert kompaktes $K \subseteq \mathbb{R}$ mit

$$Q_{n_i}(K) \geq 1 - \varepsilon \quad \text{ für alle } i \in \mathbb{N}.$$

Wähle Stetigkeitspunkte a, b von F mit $K \subset (a,b]$.

Dann gilt

$$\begin{aligned}
Q(\mathbb{R}) &\geq Q((a,b]) \\
&= F(b) - F(a) \\
&= \lim_{i\to\infty} (F_{n_i}(b) - F_{n_i}(a)) \\
&\quad \text{(nach Wahl der } F_{n_i}\text{)} \\
&= \lim_{i\to\infty} Q_{n_i}((a,b]) \\
&\geq \lim_{i\to\infty} Q_{n_i}(K) \\
&\quad \text{(da } K \subset (a,b]\text{)} \\
&\geq 1-\varepsilon.
\end{aligned}$$

$\rightsquigarrow$ Beh. $\square$.

Der folgende Satz gestattet, Probleme der Verteilungskonvergenz in $\mathbb{R}$ zurückzuführen auf Probleme der punktweisen Konvergenz einer Folge von charakteristischen Funktionen.

9.14 Satz (Stetigkeitssatz von Lévy-Cramér)

a) W-Maße Q_n $(n \in \mathbb{N})$ bzw. Q auf $\mathcal{B}$ mit charakteristischen Funktionen φ_n bzw. φ, d. h.

$$\varphi_n(u) = \int_{\mathbb{R}} e^{iux} dQ_n(x) \ (u \in \mathbb{R}).$$

Dann gilt:

$$Q_n \to Q \text{ schwach } \Leftrightarrow \ \forall u \in \mathbb{R}: \ \varphi_n(u) \to \varphi(u) \ (n \to \infty).$$

b) Genauer gilt:

b_1) $Q_n \to Q$ schwach $\Rightarrow$ $\forall u \in \mathbb{R}: \varphi_n(u) \to \varphi(u)$ $(n \to \infty)$.

b_2) Wenn für jedes $u \in \mathbb{R}$

$$\varphi(u) = \lim_{n\to\infty} \varphi_n(u)$$

existiert und $\varphi : \mathbb{R} \to \mathbb{C}$ im Punkt 0 stetig ist, dann ist φ charakteristische Funktion zu einem W-Maß Q und es gilt:

$$Q_n \to Q \text{ schwach}.$$

Beweis:
Da die charakteristische Funktion stetig ist, folgt a) aus b).
Nachweis von b_1):

Da $x \mapsto \cos(ux)$ und $x \mapsto \sin(ux)$ stetig und beschränkt sind, folgt aus $Q_n \to Q$ schwach

$$\begin{aligned}\varphi_n(u) &= \int_{\mathbb{R}} \cos(ux)\, dQ_n(x) + i \cdot \int_{\mathbb{R}} \sin(ux) dQ_n(x) \\ &\overset{(n\to\infty)}{\longrightarrow} \int_{\mathbb{R}} \cos(ux)\, dQ(x) + i \cdot \int \sin(ux)\, dQ(x) \\ &= \int_{\mathbb{R}} e^{iux}\, dQ(x) = \varphi(u).\end{aligned}$$

Nachweis von b_2):
Es genügt zu zeigen: $(Q_n)_n$ ist gleichmäßig straff.
Denn nach dem Satz von Prokohorov ist $(Q_n)_n$ dann auch relativ kompakt, also existiert zu $(Q_n)_n$ Teilfolge $(Q_{n_i})_i$ und W-Maß Q mit

$$Q_{n_i} \to Q \text{ schwach.}$$

Nach b_1) gilt dann für die charakteristischen Funktionen φ_n von Q_n:

$$\varphi_{n_i}(x) \to \varphi^*(x)\ (i \to \infty) \quad \text{für alle } x \in \mathbb{R},$$

wobei φ^* die charakteristische Funktion zu Q ist.
Nach Voraussetzung impliziert dies aber

$$\varphi^* = \varphi,$$

also existiert W-Maß Q mit charakteristischer Funktion φ.
Ist nun $g : \mathbb{R} \to \mathbb{R}$ stetig und beschränkt, so existiert zu jeder Teilfolge $(n_k)_k$ vor $(n)_n$ Teilteilfolge $(n_{k_l})_l$ mit $Q_{n_{k_l}} \to$ W-Maß schwach, da $(Q_n)_n$ gleichmäßig straff ist. Wie oben sieht man, dass dieses W-Maß gerade Q ist, was

$$\int g(x) dQ_{n_{k_l}}(x) \to \int g(x) dQ(x)\ (l \to \infty) \text{ impliziert.}$$

Nachweis von: $\{Q_n : n \in \mathbb{N}\}$ ist gleichmäßig straff
Aus

$$\varphi(0) = \lim_{n\to\infty} \varphi_n(0) = \lim_{n\to\infty} 1 = 1$$

folgt mit der Stetigkeit von φ im Nullpunkt:

$$\frac{1}{v} \int_{-v}^{v} |1 - \varphi(u)| du \to 0 \quad \text{für } v \to 0.$$

Sei $\varepsilon > 0$ beliebig. Dann existiert $v > 0$ mit

$$\frac{1}{v} \int_{-v}^{v} |1 - \varphi(u)| du < \varepsilon$$

und mit $\varphi_n(u) \to \varphi(u)$ $(n \to \infty)$ und dem Satz von der majorisierten Konvergenz folgt

$$\frac{1}{v}\int_{-v}^{v} |1-\varphi_n(u)|du \overset{(n\to\infty)}{\longrightarrow} \frac{1}{v}\int_{-v}^{v} |1-\varphi(u)|du < \epsilon.$$

Also gilt: $\exists\, n_0\, \forall n \geq n_0 : \frac{1}{v}\int\limits_{-v}^{v} |1-\varphi_n(u)|du < 2\varepsilon.$

Wir zeigen nun:

$$Q_n\left(\left\{x \in \mathbb{R} : |x| \geq \frac{2}{v}\right\}\right) \leq \frac{1}{v}\int_{-v}^{v} |1-\varphi_n(u)|du.$$

Dazu:

$$\begin{aligned}
&\tfrac{1}{v}\int\limits_{-v}^{v}(1-\varphi_n(u))du = \tfrac{1}{v}\int\limits_{-v}^{v}\left(1-\int\limits_{\mathbb{R}} e^{iux}dQ_n(x)\right)du \\
&= \tfrac{1}{v}\int\limits_{\mathbb{R}}\left\{\int\limits_{-v}^{v}(1-e^{iux})du\right\}dQ_n(x) \\
&\qquad \text{(nach Fubini und da } Q_n \text{ W-Maß)} \\
&= \tfrac{1}{v}\int\limits_{\mathbb{R}}\left(2v - 2\cdot\tfrac{\sin\, vx}{x}\right)dQ_n(x) = \int\limits_{\mathbb{R}}\underbrace{\left(2-2\cdot\frac{\sin\, vx}{vx}\right)}_{\substack{\geq 0 \\ (da\ |\sin(u)|\leq u\ für\ u\in\mathbb{R})}}dQ_n(x)
\end{aligned}$$

impliziert

$$\begin{aligned}
\tfrac{1}{v}\int\limits_{-v}^{v}|1-\varphi_n(u)|du &\geq \left|\tfrac{1}{v}\int\limits_{-v}^{v}(1-\varphi_n(u))du\right| \\
&= \int\limits_{\mathbb{R}} 2\cdot\left(1-\tfrac{\sin vx}{vx}\right)dQ_n(x) \\
&\geq \int\limits_{\{x\in\mathbb{R}:|x|\geq\frac{2}{v}\}} 2\cdot\left(1-\underbrace{\frac{\sin vx}{vx}}_{\leq\frac{1}{v\cdot\frac{2}{v}}=\frac{1}{2}}\right)dQ_n(x) \\
&\qquad\qquad \text{im Integrationsbereich} \\
&\geq \int\limits_{\{x\in\mathbb{R}:|x|\geq\frac{2}{v}\}} 2\cdot\left(1-\tfrac{1}{2}\right)dQ_n(x) \\
&= Q_n\left(\left\{x\in\mathbb{R}:|x|\geq\tfrac{2}{v}\right\}\right), w.z.z.w.
\end{aligned}$$

Abschluss des Beweises:

Bisher gezeigt: $\forall\varepsilon > 0\ \exists v > 0\ \exists n_0\ \forall n \geq n_0$:

$$Q_n\left(\left\{x \in \mathbb{R} : |x| \geq \frac{2}{v}\right\}\right) < 2\varepsilon.$$

Sei nun $\varepsilon > 0$ beliebig.

Wir wählen dann $K = [-a, a] \subseteq \mathbb{R}$ so, dass gilt:

1. $K^c \subset \left(-\frac{2}{v}, \frac{2}{v}\right)^c$

2. $Q_n(K^c) \leq 2\varepsilon$ für $n = 1, 2, \ldots, n_0$.

Hierbei gilt 1. für a groß genug immer, und 2. ist möglich, da für $a \to \infty$

$$Q_n([-a, a]^c) \to Q_n(\emptyset) = 0$$

konvergiert (nach der Stetigkeit des W-Maßes Q_n von oben).
Dann gilt aber für $n \leq n_0$: $Q_n(K^c) \leq 2\varepsilon$, und für $n > n_0$:

$$Q_n(K^c) \leq Q_n\left(\left\{x \in \mathbb{R} : |x| \geq \frac{2}{v}\right\}\right) \overset{s.o.}{\leq} 2\varepsilon,$$

womit "$\{Q_n : n \in \mathbb{N}\}$ gleichmäßig straff" gezeigt ist.

$\rightsquigarrow$ Beh. $\square$

9.15 Bemerkung. Die obigen Definitionen und Sätze lassen sich auf W-Maße auf $\mathcal{B}_k$ bzw. k-dim. Zufallsvektoren übertragen.

10 Zentrale Grenzwertsätze

Ausssagen über die Approximation von Verteilungen, insbesondere der Verteilungen von Summen von Zufallsvariablen, durch die Normalverteilung im Sinne der schwachen Konvergenz werden als zentrale Grenzwertsätze bezeichnet.

10.1 Satz (Zentraler Grenzwertsatz von Lindeberg-Lévy im 1-dim. Fall).
Sei $(X_n)_{n\in\mathbb{N}}$ eine unabhängige Folge identisch verteilter quadratisch integrierbarer reeller ZVn mit $EX_1 =: a$, $V(X_1) =: \sigma^2$ mit $\sigma > 0$. Dann gilt:

$$\frac{1}{\sqrt{n}\sigma}\sum_{k=1}^{n}(X_k - a) \xrightarrow{\mathcal{D}} N(0,1)\text{-verteilte reelle ZV.}$$

Sprechweise:
$\sum\limits_{k=1}^{n} X_k$ ist bei großem n annähernd $\mathcal{N}(n \cdot a, n \cdot \sigma^2)$-verteilt.

Beweis:
oBdA $a = 0$ und $\sigma^2 = 1$, ansonsten ersetzt man X_k durch $\frac{X_k - a}{\sigma}$.
Sei $\varphi_{\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k}$ die charakteristische Funktion von $\frac{1}{\sqrt{n}}\sum\limits_{k=1}^{n} X_k$. Nach dem Stetigkeitssatz von Lévy-Cramér genügt es dann zu zeigen:

$$\varphi_{\frac{1}{\sqrt{n}}\sum\limits_{k=1}^{n} X_k}(u) \to e^{-u^2/2} \quad (n \to \infty) \quad (u \in \mathbb{R}).$$

Hierbei ist $u \mapsto e^{-u^2/2}$ die charakteristische Funktion einer $\mathcal{N}(0,1)$-verteilten Zufallsvariablen.

Da $X_1, X_2, \ldots$ unabhängig und identisch verteilt sind, gilt

$$\begin{aligned}
\varphi_{\frac{1}{\sqrt{n}}\sum\limits_{k=1}^{n} X_k}(u) &= E\, e^{iu\left(\frac{1}{\sqrt{n}}\sum\limits_{k=1}^{n} X_k\right)} \\
&= E \prod_{i=1}^{n} e^{i\frac{u}{\sqrt{n}}X_k} \\
&= \prod_{i=1}^{n} E\, e^{i\frac{u}{\sqrt{n}}X_k} \\
&= \left(\varphi_{X_1}\left(\tfrac{u}{\sqrt{n}}\right)\right)^n,
\end{aligned}$$

also ist zu zeigen:

$$\left(\varphi_{X_1}\left(\frac{u}{\sqrt{n}}\right)\right)^n \to e^{-u^2/2} \; (n \to \infty).$$

Wegen $EX_1^2 < \infty$ ist φ_{X_1} zweimal stetig differenzierbar, und damit gilt nach Taylor:

$$\begin{aligned}
\varphi_{X_1}(u) &= \varphi_{X_1}(0) + \varphi'_{X_1}(0) \cdot u + \tfrac{1}{2} \cdot \varphi''_{X_1}(0) \cdot u^2 + \rho(u) \\
&= 1 + i \cdot EX_1 \cdot u + \tfrac{1}{2} \cdot (-1) \cdot EX_1^2 \cdot u^2 + \rho(u) \\
&= 1 - \tfrac{1}{2} \cdot u^2 + \rho(u)
\end{aligned}$$

(wegen $EX_1 = 0$ und $EX_1^2 = 1$ nach Voraussetzung), wobei für das Restglied $\rho(u)$ gilt:

$$\frac{\rho(u)}{u^2} \to 0 \; (u \to 0)$$

Die letzte Beziehung sieht man (unter Verwendung der Lagrangeschen Formel für das Restglied des linearen Taylor-Polynoms) so ein:

$$\frac{\rho(u)}{\frac{1}{2}u^2} = \left(Re\varphi''_{X_1}(\xi_u) - Re\varphi''_{X_1}(0)\right) + i \cdot \left(Im\varphi''_{X_1}(\mu_u) - Im\varphi''_{X_1}(0)\right)$$

mit $|\xi_u| \leq u$ und $|\mu_u| \leq u$, woraus mit der Stetigkeit von φ''_{X_1} die Behauptung folgt.

Damit gilt für $u \in \mathbb{R}$ fest

$$\begin{aligned}
\left(\varphi_{X_1}\left(\tfrac{u}{\sqrt{n}}\right)\right)^n &= \left(1 - \tfrac{1}{2} \cdot \tfrac{u^2}{n} + \rho\left(\tfrac{u}{\sqrt{n}}\right)\right)^n \\
&= \left(1 + \tfrac{-\frac{1}{2}u^2 + n \cdot \rho\left(\frac{u}{\sqrt{n}}\right)}{n}\right)^n.
\end{aligned}$$

Mit

$$n \cdot \rho\left(\frac{u}{\sqrt{n}}\right) = u^2 \cdot \frac{\rho(u/\sqrt{n})}{(u/\sqrt{n})^2} \to u^2 \cdot 0 = 0 \; (n \to \infty)$$

und

$$\left(1 + \frac{x_n}{n}\right)^n \to e^x \; (n \to \infty)$$

für jede Folge $(x_n)_n$ mit $x_n \to x \ (n \to \infty)$
(der letzte Grenzwert folgt z. B. aus

$$\left(1+\frac{x_n}{n}\right)^n = e^{n\cdot\ln(1+\frac{x_n}{n})} = e^{x_n \frac{\ln(1+\frac{x_n}{n})}{\frac{x_n}{n}}}$$

und

$$\lim_{z\to 0}\frac{\ln(1+z)}{z} \overset{de\ l'Hospital}{=} \lim_{z\to 0}\frac{\frac{1}{1+z}}{1} = 1)$$

folgt die Behauptung. □

10.2 Korollar (Zentraler Grenzwertsatz von de Moivre und Laplace).
Für eine unabhängige Folge $(X_n)_{n\in\mathbb{N}}$ identisch verteilter reeller ZVn auf $(\Omega, \mathcal{A}, P)$ mit $P[X_1 = 1] = p$, $P[X_1 = 0] = 1 - p =: q \ (0 < p < 1)$ gilt

$$\underset{\substack{\alpha<\beta\\ \in\mathbb{R}}}{\forall} \ P\left[\alpha \underset{(=)}{<} \frac{\sum_{k=1}^n X_k - np}{\sqrt{npq}} \underset{(=)}{<} \beta\right] \to \frac{1}{\sqrt{2\pi}}\int_\alpha^\beta e^{-t^2/2}\,dt \ \ (n\to\infty).$$

11 Beweis des Submartingalkonvergenztheorems

11.1 Definition. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Monoton wachsende Folge $(\mathcal{A}_n)_{n\in\mathbb{N}}$ von σ-Algebren $\mathcal{A}_n \subset \mathcal{A}$. Eine ZV T: $(\Omega, \mathcal{A}, P) \to (\overline{\mathbb{N}} := \mathbb{N}\cup\{\infty\}, \mathcal{P}(\overline{\mathbb{N}}))$ heißt **Stoppzeit** bzgl. $(\mathcal{A}_n)$, wenn $\underset{k\in\mathbb{N}}{\forall} \ [T = k] \in \mathcal{A}_k$; hierbei heißt T **Stoppzeit im engeren Sinne**, falls $P[T < \infty] = 1$ ["Kein Vorgriff auf die Zukunft"].

Wichtiger Fall: $\mathcal{A}_n = \mathcal{F}(X_1, \ldots, X_n)$ $(n \in \mathbb{N})$ mit ZVn X_n
Deutung (X_n) … Folge der Gewinnstände in einem Spiel.
Ein Spieler ohne prophetische Gaben bricht das Spiel im zufälligen Zeitpunkt T aufgrund des bisherigen Spielverlaufs ab.
Beispiel: $T(\omega) = \inf\{n \in \mathbb{N}: X_n(\omega) \in B\}$, $\omega \in \Omega$ — festes messbares B.

11.2 Definition. Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$
Folge $(X_n)_{n\in\mathbb{N}}$ von ZVn X_n: $(\Omega, \mathcal{A}, P) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$;
Folge $(T_n)_{n\in\mathbb{N}}$ von Stoppzeiten bzgl. (X_n) [d.h. bzgl. $(\mathcal{F}(X_1, \ldots, X_n))$]
mit $T_1 \le T_2 \le \ldots < \infty$. Neue Folge: Folge $(X_{T_n})_{n\in\mathbb{N}}$ von ZVn definiert durch

$$(X_{T_n})(\omega) := X_{T_n(\omega)}(\omega), \ \ \omega \in \Omega\,.$$

Übergang von (X_n) zu (X_{T_n}) heißt **optional sampling** [frei gewählte Stichprobenbildung]

Deutung … Testen des Spielverlaufs zu den Zeitpunkten $T_n(\omega)$.

11.3 Satz (Optional sampling theorem).
Submartingal $(X_n)_{n\in\mathbb{N}}$. Festes $M \in \mathbb{N}$.

Folge $(T_n)_{n\in\mathbb{N}}$ von Stoppzeiten bzgl. (X_n) mit $T_1 \leq T_2 \leq \ldots \leq M$.
Die durch optional sampling erhaltene Folge $(X_{T_n})_{n\in\mathbb{N}}$ ist ebenfalls ein Submartingal. — Entsprechend für Martingal statt Submartingal (Deutung: die Fairness eines Spielverlaufs wird nicht geändert).

Beweis von Satz 11.3 für Submartingale (analog für Martingale):
Sei $n \in \mathbb{N}$ und $C \in \mathcal{F}(X_{T_1}, \ldots, X_{T_n})$.
zu zeigen: $\int_C X_{T_{n+1}} dP \geq \int_C X_{T_n} dP$.
Wegen $C = \sum\limits_{j=1}^{M} D_j$ mit $D_j = C \cap [T_n = j]$ genügt es für $j \in \{1, \ldots, M\}$ zu zeigen:

$$\int_{D_j} X_{T_{n+1}} dP \geq \int_{D_j} X_{T_n} dP \left(= \int_{D_j} X_j dP \right)$$

Dazu:
Für $j = M$ ist die Behauptung trivial, da dann auf D_j gilt

$$T_n = M = T_{n+1}.$$

Wir zeigen zunächst:

$$D_j \in \mathcal{F}(X_1, \ldots, X_j)$$

Wegen

$$C = [(X_{T_1}, \ldots X_{T_n}) \in B] \text{ für ein } B \in \mathcal{B}_n$$

folgt dies aus

$$\begin{aligned} D_j &= [(X_{T_1}, \ldots, X_{T_n}) \in B, T_n = j] \\ &= \sum_{j_1 \leq \cdots \leq j_{n-1} \leq j} ((X_{T_1}, \ldots, X_{T_n}) \in B, T_1 = j_1, \ldots, T_{n-1} = j_{n-1}, T_n = j) \\ &= \sum_{j_1 \leq \cdots \leq j_{n-1} \leq j} \underbrace{\left[(X_{j_1}, \ldots, X_{j_{n+1}}, X_j) \in B, T_1 = j_1, \ldots, T_{n-1} = j_{n-1}, T_n = j\right]}_{\in \mathcal{F}(X_1, \ldots, X_j)} \end{aligned}$$

Nun beachten wir, dass aus $T_{n+1} \geq T_n$ folgt

$$\begin{aligned} \int_{D_j} X_{T_{n+1}} dP &= \sum_{k=j}^{M} \int\limits_{D_j \cap [T_{n+1} = k]} X_{T_{n+1}} dP \\ &= \sum_{k=j}^{M} \int\limits_{D_j \cap [T_{n+1} = k]} X_k \, dP. \end{aligned}$$

Für $\underline{j = M}$ war die Behauptung trivial, s. o.
Für $\underline{j \leq M - 1}$ gilt

$$D_j \cap [T_{n+1} = M] = D_j \cap [T_{n+1} \leq M - 1]^c \in \mathcal{F}(X_1, \ldots, X_{M-1}),$$

also ist in diesem Fall

$$\int\limits_{D_j \cap [T_{n+1} = M]} X_M dP \geq \int\limits_{D_j \cap [T_{n+1} = M]} X_{M-1} dP$$

(da $(X_n)_n$ Submartingal ist), was

$$\int\limits_{D_j\cap[T_{n+1}=M-1]} X_{M-1}dP + \int\limits_{D_j\cap[T_{n+1}=M]} X_M dP$$
$$\geq \int\limits_{D_j\cap[T_{n+1}\geq M-1]} X_{M-1}dP$$

impliziert.
Für $j \leq M-2$ ist

$$D_j \cap [T_{n+1} \geq M-1] = D_j \cap [T_{n+1} \leq M-2]^c \in \mathcal{F}(X_1, \ldots, X_{M-2})$$

und damit

$$\int\limits_{D_j\cap[T_{n+1}\geq M-1]} X_{M-1}dP \geq \int\limits_{D_j\cap[T_{n+1}\geq M-1]} X_{M-2}dP,$$

was wiederum

$$\sum_{k=M-2}^{M} \int\limits_{D_j\cap[T_{n+1}=k]} X_k dP \geq \int\limits_{D_j\cap[T_{n+1}\geq M-2]} X_{M-2}dP$$

impliziert.
Schließt man analog weiter, so erhält man schließlich

$$\sum_{k=j}^{M} \int\limits_{D_j\cap[T_{n+1}=k]} X_k dP$$
$$\geq \int\limits_{D_j\cap[T_{n+1}\geq j]} X_j dP$$
$$\overset{(*)}{=} \int\limits_{D_j} X_j dP \overset{(**)}{=} \int\limits_{D_j} X_{T_n} dP$$

wobei $(*)$ aus $T_n = j$ auf D_j und $T_{n+1} \geq T_n$ folgt, und $(**)$ wegen $T_n = j$ auf D_j gilt.
Insgesamt ist damit gezeigt:

$$\int\limits_{D_j} X_{T_{n+1}} dP = \sum_{k=j}^{M} \int\limits_{D_j\cap[T_{n+1}=k]} X_k dP \geq \int\limits_{D_j} X_{T_n} dP,$$

w.z.z.w. □

11.4 Satz (**"upcrossing inequality"** (von Doob)).
Sei $(X_1, \ldots, X_n)$ ein — beim festen Index $n \in \mathbb{N}$ abbrechendes — Submartingal. Feste reelle Zahlen a, b mit $a < b$. Die ZV $U[a,b]$ gebe die Anzahl der aufsteigenden Überquerungen des Intervalls $[a,b]$ durch $X_1, \ldots, X_n$ an (d.h. die Anzahl der Übergänge der abbrechenden Folge von einem Wert $\leq a$ zu einem Wert $\geq b$). Dann gilt

$$(b-a)EU[a,b] \leq E(X_n - a)^+ - E(X_1 - a)^+ .$$

Beweis von Satz 11.4.
Mit $X_1, \dots, X_n$ ist auch $(X_1 - a)^+, \dots, (X_n - a)^+$ ein Submartingal, da

$$\begin{aligned} E((X_{k+1} - a)^+ | X_1, \dots, X_k) &= \max\{E\{(X_{k+1} - a)^+ | X_1, \dots, X_k), 0\} \\ &\geq \max\{E\{X_{k+1} - a | X_1, \dots, X_k\}, 0\} \\ &= \max\{E(X_{k+1} | X_1, \dots, X_n) - a, 0\} \\ &= \max\{X_k - a, 0\} = (X_k - a)^+ \end{aligned}$$

da $X_1, \dots, X_n$ Martingal ist.

$U[a, b]$ gibt auch die Anzahl der aufsteigenden Überdeckungen des Intervalls $[0, b - a]$ durch $(X_1 - a)^+, \dots, (X_n - a)^+$ an.

Daher oBdA: $a = 0, X_1 \geq 0, \dots, X_n \geq 0$.

zu zeigen: $b \cdot E\, U[a, b] \leq EX_n - EX_1$.

Dazu setzen wir

$$\begin{aligned} T_1(\omega) &:= 1 \\ T_2(\omega) &:= \begin{cases} \min\{i \in \{T_1(\omega), \dots, n\} : X_i(\omega) = 0\} & , \text{ falls } \exists_{i \in \{T_1(\omega), \dots, n\}} X_i(\omega) = 0 \\ \\ n & , \text{ sonst} \end{cases} \\ T_3(\omega) &:= \begin{cases} \min\{i \in \{T_2(\omega), \dots, n\} : X_i(\omega) \geq b\} & , \text{ falls } \exists_{i \in \{T_2(\omega), \dots, k\}} X_i(\omega) \geq b \\ n & , \text{ sonst} \end{cases} \end{aligned}$$

u.s.w.
$T_{n+1}(\omega) = n$.
Nach Konstruktion gilt dann

$$\sum_j \left(X_{T_{2j+1}} - X_{T_{2j}}\right) \geq b \cdot U[0, b].$$

Weiter sieht man leicht, dass $T_1, \dots, T_{n+1}$ Stoppzeiten sind mit

$$T_1 \leq T_2 \leq \ldots \leq T_{n+1} \leq n.$$

Nach Satz 11.3 ist damit $\left(X_{T_1}, \dots, X_{T_{n+1}}\right)$ ein Submartingal, folglich gilt

$$E\left(X_{T_{2j}} - X_{T_{2j-1}}\right) \geq 0.$$

Damit folgt:

$$\begin{aligned} EX_n - EX_1 &= E\left(\sum_{k=1}^{n} \left(X_{T_{k+1}} - X_{T_k}\right)\right) \\ &= E\left(\sum_j \left(X_{T_{2j+1}} - X_{T_{2j}}\right)\right) + E\left(\sum_j \left(X_{T_{2j}} - X_{T_{2j-1}}\right)\right) \\ &\overset{s.o.}{\geq} b \cdot E(U[0, b]) + 0 \end{aligned}$$

$\rightsquigarrow$ Beh. $\square$

Beweis von Satz 8.5:
Für $a, b \in \mathbb{R}$ mit $a < b$ setze

$$A(a,b) = [\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n].$$

Im ersten Schritt des Beweises zeigen wir:

$$P(A(a,b)) = 0.$$

Dazu sei $U_n(a,b)$ die Anzahl der aufsteigenden Überquerungen von $[a,b]$ durch $(X_1, \dots, X_n)$. Nach Satz 11.4 gilt dann

$$E[U_n(a,b)] \leq \frac{E(X_n - a)^+ - E(X_1 - a)^+}{b-a},$$

woraus

$$\limsup_{n\to\infty} E[U_n(a,b)] \leq \limsup_{n\to\infty} \frac{E|X_n| + a}{b-a} < \infty \qquad (*)$$

folgt.
Angenommen, $P(A(a,b)) > 0$. Auf $A(a,b)$ gilt

$$U_n[a,b] \uparrow \infty \;\; (n \to \infty)$$

(da aus $\liminf_{n\to\infty} X_n < a$ und $\limsup_{n\to\infty} X_n > b$, folgt, dass unendlich viele der X_n kleiner als a bzw. größer als b sind). Dies impliziert

$$EU_n(a,b) \geq \int_{A(a,b)} U_n(a,b) dP \to \infty \;\; (n \to \infty)$$

nach dem Satz von der monotonen Konvergenz, was ein Widerspruch zu $(*)$ ist. Also gilt

$$P(A(a,b)) = 0.$$

Im zweiten Schritt des Beweises zeigen wir

$$X_n \to X^* \;\; P - f.s.$$

für eine erweitert rellwertige Zufallsvariable X^*.
Dazu setzen wir

$$X^* = \limsup_{n\to\infty} X_n$$

und folgern die Behauptung aus

$$\begin{aligned}
&P[(X_n)_n \text{ konvergent}] = 1 - P[\liminf_{n\to\infty} X_n < \limsup_{n\to\infty} X_n] \\
&= 1 - \sum_{a,b\in\mathbb{Q}, a<b} P[\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n] \\
&\overset{Schritt\ 1}{=} 1 - \sum_{a,b\in\mathbb{Q}, a<b} 0 = 1.
\end{aligned}$$

Im dritten Schritt zeigen wir die Behauptung des Satzes.
Wegen

$$\begin{aligned} E|X^*| &= E[\liminf_{n\to\infty} |X_n|] \\ &\quad \text{(da } (X_n)_n \text{ mit Wahrscheinlichkeit 1 konvergent ist)} \\ &\leq \liminf_{n\to\infty} E|X_n| \\ &\quad \text{(nach dem Lemma von Fatou)} \\ &\leq \limsup_{n\to\infty} E|X_n| < \infty \end{aligned}$$

ist X^* integrierbar und das Ereignis $[|X^*| = \infty]$ hat die Wahrscheinlichkeit Null.
Setzen wir nun

$$X = X^* \cdot 1_{[|X^*|<\infty]},$$

so ist X reelle Zufallsvariable, für die wegen

$$X = X^* \;\; P - f.s.$$

gilt: $X_n \to X \;\; P - f.s.$ und X ist integrierbar. $\square$

Literatur

Bauer, H.: Wahrscheinlichkeitstheorie; de Gruyter, Berlin (1990), 4. Aufl.

Behnen, K., Neuhaus, G.: Grundkurs Stochastik; Teubner, Stuttgart (1995), 3. Aufl.

Billingsley, P.: Probability and Measure; Wiley, New York (1979).

Borowkow, A. A.: Wahrscheinlichkeitstheorie; Birkhäuser, Basel (1976).

Breiman, L.: Probability; Addison-Wesley, Reading Mass. (1968).

Chow, Y. S., Teicher, H.: Probability Theory - Independence, Interchangeability, Martingales; Springer, Berlin (1978).

Chung, K. L.: A Course in Probability Theory; Academic Press, New York (1974), 2nd ed.

Dinges, H., Rost, H.: Prinzipien der Stochastik; Teubner, Stuttgart (1982).

Dudley, R.M.: Real Analysis and Probability; Chapman & Hall, New York (1989).

Durrett, R..: Probability:Theory and Examples; Wadsworth & Brooks/Cole, Pacific Grove (1991). New York (1989).

Fabian, V., Hannan, J.: Introduction to Probability and Mathematical Statistics; Wiley, New York (1985).

Feller, W.: An Introduction to Probability and Its Applications I,II; Wiley, New York (1970/71), 3rd ed./2nd ed.

Fisz, M.: Wahrscheinlichkeitsrechnung und mathematische Statistik; VEB Deutscher Verlag der Wissenschaften, Berlin (1970), 5. Aufl.

Gänssler, P., Stute, W.: Wahrscheinlichkeitstheorie; Springer, Berlin (1977).

Gnedenko, B W.: Lehrbuch der Wahrscheinlichkeitsrechnung; Akademie-Verlag, Berlin (1968), 5. Aufl.

Hesse, C.: Angewandte Wahrscheinlichkeitstheorie; Vieweg (2003).

Hinderer, K.: Grundbegriffe der Wahrscheinlichkeitstheorie; Springer, Berlin (1985), 3. korr. Nachdruck.

Jacod, J., Protter, P.: Probability Essentials. Springer, Berlin (2000).

Kallenberg, O.: Foundations of Modern Probability. Springer, New York (1997).

Krengel, U.: Einführung in die Wahrscheinlichkeitstheorie und Statistik; Vieweg, Braunschweig (1991), 3. Aufl.

Krickeberg, K.: Wahrscheinlichkeitstheorie; Teubner, Stuttgart (1963).

Krickeberg, K., Ziezold, H.: Stochastische Methoden; Springer, Berlin (1988), 3. Aufl.

Laha, R. G., Rohatgi, V. K.: Probability Theory; New York, Wiley & Sons (1979).

LAMPERTI, H.: Probability; Benjamin, New York (1966).

LOÈVE, M.: Probability Theory I,II; Springer, Berlin (1977/78), 4th ed.

NEVEU, J.: Mathematische Grundlagen der Wahrscheinlichkeitstheorie; Oldenbourg, München (1969).

PFANZAGL, J.: Elementare Wahrscheinlichkeitsrechnung; de Gruyter, Berlin (1988).

PLACHKY, D.; BARINGHAUS, SCHMITZ, N./PLACHKY, D.: Stochastik I/II; Akad. Verlagsges., Frankfurt (1978/1981).

PROHOROV, YU. V., ROZANOV, YU. A.: Probability Theory; Springer, Berlin (1969).

RÉNYI, A.: Wahrscheinlichkeitsrechnung; VEB Deutscher Verlag der Wissenschaften, Berlin (1962).

RÉNYI, A.: Foundations of Probability Theory; Holden-Day, San Francisco (1970).

RESNICK, S.: A Probability Path; Birkhäuser, Boston (1999).

RICHTER, H.: Wahrscheinlichkeitstheorie; Springer, Berlin (1966), 2. Aufl.

SCHMITZ, N.: Vorlesungen über Wahrscheinlichkeitstheorie; Teubner, Stuttgart (1996).

SHIRYAYEV, A. N.: Probability; Springer, Berlin (1984).

SINAI, Y. G.: Probability Theory; Springer, Berlin (1992).

VOGEL, W.: Wahrscheinlichkeitstheorie; Vandenhoeck & Ruprecht, Göttingen (1970).

WHITTLE, P.: Probability via Expectation; Springer, New York (2000), 4. Aufl.

Bücher mit Übungsaufgaben:

CACOULLOS, T. Exercises in Probability; Springer, New York (1989).

CAPINSKI M.,ZASTAWNIAK, T. Probability through Problems; Springer, New York (2001).