Einführung in die

Wahrscheinlichkeitstheorie und Statistik

Robert Denk

Sommersemester 1999

Universität Regensburg

Naturwissenschaftliche Fakultät I

– Mathematik –

Inhaltsverzeichnis

Einleitung

Das vorliegende Skript gibt den Inhalt einer von mir im Sommersemester 1999 gehaltenen (vierstündigen) Vorlesung zur Einführung in die Wahrscheinlichkeitstheorie und Statistik wieder. Es besteht im wesentlichen in einer fast wörtlichen Wiedergabe des vorgetragenen Stoffes.

Die Vorlesung richtete sich an Studierende der Richtungen Diplom-Mathematik und Lehramt Mathematik für Gymnasien ab dem vierten Semester und sollte eine erste Einführung in typische Denkweisen und Aussagen der Stochastik und der Statistik liefern. Da die Zeit im Sommersemester recht knapp bemessen ist, war es unumgänglich, sich auf eine relativ kleine Auswahl des möglichen Stoffes zu beschränken. Dabei legte ich das Hauptgewicht auf die klassischen Aussagen der Stochastik, insbesondere wurden etwa die verschiedenen Konvergenzbegriffen für Folgen von Zufallsvariablen und die dafür geltenden Aussagen behandelt. Unter anderem sind hier die schwachen und starken Gesetze der großen Zahlen und der zentrale Grenzwertsatz zu nennen. Bei einigen Beweisen mußte der Hinweis auf die entsprechende Maßtheorie-Vorlesung genügen, welche im gleichen Semeseter als zweistündige Vorlesung gehalten wurde. Für die Statistik blieb relativ wenig Zeit, und ich versuchte, wenigstens die wichtigsten Ideen aus der Theorie der Parameter-Punktschätzung und der Signifikanztests zu erläutern.

Ein ständiges Problem bei einer Vorlesung über Wahrscheinlichkeitstheorie liegt darin, daß diese nicht der Zeitpunkt und der Rahmen dafür ist, die Grundzüge der Maß- und Integrationstheorie zu entwickeln. Hier war ich auf die (mehr oder weniger) vorhandenen Kenntnisse aus der Analysis-Vorlesung angewiesen. Eine zusätzliche Unterstützung in dieser Richtung konnten die Studenten in einem von Holger Plank abgehaltenen Tutorium zur Maßtheorie erhalten. Unter anderem wurden endliche Produkte von Maßräumen dort behandelt; der zugehörige Text findet sich in Anhang, ebenso wie die ebenfalls von Herrn Plank gestellten Übungsaufgaben.

Der behandelte Stoff wurde von mir durch eine Reihe von Graphiken veranschaulicht, welche in der Vorlesung als Folien gezeigt wurden und von denen sich einige auch in diesem Skript wiederfinden. Diese Zeichnungen wurden mit dem Mathematik-Programmpaket Maple erzeugt, die Interessierten finden die entsprechenden Maple-Befehle im Anhang.

Ich hoffe, daß meinen Studenten diese Vorlesung Spaß gemacht hat (mir schon) und daß dieses Skript für den einen oder anderen nützlich sein wird. Schließlich möchte mich noch bei Stephan Otto, Michaela Lautenschlager und Holger Plank für die Unterstützung bei der Anfertigung dieses Skripts bedanken.

Regensburg, November 1999 Robert Denk

1. Bemerkungen zur Maßtheorie

Mathematisch gesehen ist eine Wahrscheinlichkeit ein normiertes Maß. Daher werden zunächst einige fundamentale Begriffe aus der Maßtheorie zusammengestellt und wiederholt.

Definition 1.1. Sei Ω eine Menge, $\mathscr{P}(\Omega) := \{A : A \subset \Omega\}$ die Potenzmenge von Ω und $\mathscr{A} \subset \mathscr{P}(\Omega)$.

a) $\mathscr{A}$ heißt σ-Algebra, falls gilt:

(i) $\emptyset \in \mathscr{A}$,
(ii) Für jedes $A \in \mathscr{A}$ gilt $A^c := \{\omega \in \Omega : \omega \notin A\} \in \mathscr{A}$.
(iii) Für $A_n \in \mathscr{A}$ $(n \in \mathbb{N})$ gilt $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{A}$.

In diesem Fall heißt $(\Omega, \mathscr{A})$ Meßraum.

b) Falls statt (iii) nur gilt

(iii′) Für $A, B \in \mathscr{A}$ gilt $A \cup B \in \mathscr{A}$,

so heißt $\mathscr{A}$ eine Algebra.

c) Falls statt (iii) nur gilt:

(iii″) Falls $A_n \in \mathscr{A}$ disjunkt sind (d.h. $A_n \cap A_m = \emptyset$ für $n \neq m$), dann ist $\dot{\bigcup\limits_{n\in\mathbb{N}}} A_n \in \mathscr{A}$,

so heißt $\mathscr{A}$ ein Dynkin-System.

Bemerkung 1.2. a) Die größte σ-Algebra ist $\mathscr{P}(\Omega)$, die kleinste ist $\{\emptyset, \Omega\}$. Falls $\mathscr{A}_i$ eine σ-Algebra ist für $i \in I$, wobei I eine nichtleere Indexmenge ist, dann ist $\bigcap_{i\in I} \mathscr{A}_i$ wieder eine σ-Algebra.

Sei $\mathscr{E} \subset \mathscr{P}(\Omega)$ beliebig. Dann ist

$$\sigma(\mathscr{E}) := \bigcap \{\mathscr{A} \supset \mathscr{E} : \mathscr{A} \text{ ist } \sigma\text{−Algebra über } \Omega\}$$

die kleinste σ-Algebra, die $\mathscr{E}$ enthält (von $\mathscr{E}$ erzeugte σ-Algebra). Analog existieren ein kleinstes Dynkin-System $\mathscr{D}(\mathscr{E})$ und eine kleinste Algebra, die $\mathscr{E}$ enthält.

Die von $\mathscr{E}$ erzeugte Algebra kann man explizit angeben:

$$\Big\{ \bigcup_{i=1}^{n} \bigcap_{j=1}^{n} A_{ij} : A_{ij} \in \mathscr{E} \cup \mathscr{E}^c,\ n \in \mathbb{N} \Big\},$$

wobei $\mathscr{E}^c := \{A^c : A \in \mathscr{E}\}$.

Für σ-Algebren gilt dies keineswegs, auch nicht, wenn man n durch ∞ ersetzt und abzählbar oft iteriert!

b) Manchmal betrachtet man statt einer Algebra einen Ring, d.h. ein $\mathscr{A} \subset \mathscr{P}(\Omega)$ mit (i), (iii′) und

(ii′) Für alle $A, B \in \mathscr{A}$ ist $A \backslash B \in \mathscr{A}$.

Lemma 1.3. a) *Ein Dynkin-System $\mathscr{D}$ ist genau dann eine σ-Algebra, falls gilt:*

$$\text{Für alle } A, B \in \mathscr{D} \text{ ist } A \cap B \in \mathscr{D}$$

(*d.h. wenn $\mathscr{D}$ $\cap$-stabil ist*).

b) (Dynkin-Lemma). *Sei $\mathscr{E} \subset \mathscr{P}(\Omega)$ $\cap$-stabil. Dann ist $\sigma(\mathscr{E}) = \mathscr{D}(\mathscr{E})$.*

Beweis. a) Sei $\mathscr{D}$ $\cap$-stabil. Dann gilt für $A, B \in \mathscr{D}$:

$$A \backslash B = A \backslash (A \cap B) = A \cap (A \cap B)^c = \Big[A^c \,\dot{\cup}\, (A \cap B)\Big]^c \in \mathscr{D}\,,$$

also

$$A \cup B = (A \backslash B) \,\dot{\cup}\, B \in \mathscr{D}\,.$$

Seien $A_n \in \mathscr{D}$ für alle $n \in \mathbb{N}$. Setze $\tilde{A}_0 := \emptyset$ und $\tilde{A}_n := A_1 \cup \ldots \cup A_n \in \mathscr{D}$. Dann ist

$$\bigcup_{n \in \mathbb{N}} A_n = \dot{\bigcup_{n \in \mathbb{N}_0}} \tilde{A}_{n+1} \backslash \tilde{A}_n \in \mathscr{D}\,,$$

d.h. $\mathscr{D}$ ist σ-Algebra.

b) Zu zeigen ist nur, daß $\mathscr{D}(\mathscr{E})$ eine σ-Algebra ist. Zu $A \in \mathscr{D}(\mathscr{E})$ definiere

$$\mathscr{D}_A := \{B \in \mathscr{P}(\Omega) : A \cap B \in \mathscr{D}(\mathscr{E})\}\,.$$

Dann ist $\mathscr{D}_A$ ein Dynkin-System. Da $\mathscr{E}$ $\cap$-stabil ist, gilt

$$\mathscr{E} \subset \mathscr{D}_A \text{ für alle } A \in \mathscr{E}$$

und damit $\mathscr{D}(\mathscr{E}) \subset \mathscr{D}_A$ für alle $A \in \mathscr{E}$, d.h.

$$A \cap B \in \mathscr{D}(\mathscr{E}) \text{ für alle } A \in \mathscr{E},\ B \in \mathscr{D}(\mathscr{E})\,.$$

Dies heißt $\mathscr{E} \subset \mathscr{D}_B$ für alle $B \in \mathscr{D}(\mathscr{E})$ und damit

$$\mathscr{D}(\mathscr{E}) \subset \mathscr{D}_B \text{ für alle } B \in \mathscr{D}(\mathscr{E})\,,$$

d.h. $\mathscr{D}(\mathscr{E})$ ist $\cap$-stabil. Mit Teil a) folgt nun die Behauptung. □

Definition 1.4. a) Sei $(\Omega, \mathscr{A})$ ein Meßraum. Dann heißt $\mu : \mathscr{A} \to [0, \infty]$ ein Maß, falls gilt:

(i) $\mu(\emptyset) = 0$,
(ii) σ-Additivität: Für $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ mit $A_n \cap A_m = \emptyset$ $(n \neq m)$ gilt

$$\mu\Big(\dot{\bigcup_{n\in\mathbb{N}}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n)\,.$$

In diesem Fall heißt $(\Omega, \mathscr{A}, \mu)$ ein Maßraum.

b) Ein Maß μ heißt

- σ-finit (oder normal), falls es eine Folge $(A_n)_n \subset \mathscr{A}$ gibt mit $\bigcup_{n\in\mathbb{N}} A_n = \Omega$ und $\mu(A_n) < \infty$ für alle $n \in \mathbb{N}$.
- Wahrscheinlichkeitsmaß (W-Maß), falls $\mu(\Omega) = 1$.

c) Sei $\mathscr{A}$ ein Ring. Dann heißt $\mu : \mathscr{A} \to [0, \infty]$ ein Inhalt, falls $\mu(\emptyset) = 0$ und

$$\mu(A \cup B) = \mu(A) + \mu(B) \text{ für alle } A, B \in \mathscr{A} \text{ mit } A \cap B = \emptyset$$

(endliche Additivität) gilt.

Beispiele 1.5. a) Dirac-Maß: Zu $x \in \Omega$ definiere

$$\delta_x(A) := 1_A(x) := \begin{cases} 1\,, & x \in A\,, \\ 0\,, & x \notin A\,. \end{cases}$$

Dann ist δ_x ein Maß auf $\mathscr{P}(\Omega)$ und damit auf jeder σ-Algebra.

b) Elementargeometrischer Inhalt: Betrachte

$$\mathscr{E} := \{(a, b] : -\infty < a < b < \infty\}$$

und

$$\mathscr{R} := \Big\{ \dot{\bigcup_{j=1}^{n}} A_j : A_j \in \mathscr{E},\ A_i \cap A_j = \emptyset\ (i \neq j)\,,\ n \in \mathbb{N} \Big\}\,.$$

Dann ist $\mathscr{R}$ ein Ring. Setze

$$\lambda\Big(\dot{\bigcup_{j=1}^{n}} A_j\Big) := \sum_{j=1}^{n} (b_j - a_j)$$

für $A_j = (a_j, b_j] \in \mathscr{E}$ disjunkt. Dann ist $\lambda : \mathscr{R} \to [0, \infty)$ ein Inhalt. λ ist σ-finit in dem Sinne, daß 1.4 b) gilt, und σ-additiv, d.h. für $(A_n)_{n\in\mathbb{N}} \subset \mathscr{R}$ disjunkt mit $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{R}$ gilt

$$\lambda\Big(\dot{\bigcup_{n\in\mathbb{N}}} A_n\Big) = \sum_{n\in\mathbb{N}} \lambda(A_n)\,.$$

c) Sei $\Omega = \mathbb{N}$, $\mathscr{A} := \{A \subset \mathbb{N} : A \text{ oder } A^c \text{ endlich }\}$ und

$$\mu(A) := \begin{cases} 0\,, & \text{falls } A \text{ endlich,} \\ 1\,, & \text{falls } A^c \text{ endlich.} \end{cases}$$

Dann ist $\mathscr{A}$ eine Algebra und μ ein Inhalt, aber $\mathscr{A}$ ist keine σ-Algebra und μ ist nicht σ-additiv auf $\mathscr{A}$.

d) Zählmaß: Definiere

$$\mu(A) := \begin{cases} |A|\,, & \text{falls } A \text{ endlich,} \\ \infty\,, & \text{falls } A \text{ unendlich.} \end{cases}$$

Dann ist μ ein Maß auf $\mathscr{P}(\Omega)$, welches genau dann σ-finit ist, falls Ω abzählbar ist.

Bemerkung 1.6. Sei μ ein Inhalt auf einem Ring $\mathscr{A}$. Dann gilt:

(i) μ ist monoton, d.h. für $A, B \in \mathscr{A}$ mit $A \subset B$ gilt $\mu(A) \leq \mu(B)$.

(ii) μ ist subtraktiv, d.h. für $A, B \in \mathscr{A}$ mit $A \subset B$ und $\mu(A) < \infty$ gilt $\mu(B \backslash A) = \mu(B) - \mu(A)$.

(iii) μ ist sub-additiv, d.h. für $A_1, \dots, A_n \in \mathscr{A}$ gilt

$$\mu\Big(\bigcup_{i=1}^{n} A_i\Big) \leq \sum_{i=1}^{n} \mu(A_i)\,.$$

Satz 1.7. *Sei μ ein Inhalt auf einem Ring $\mathscr{A}$. Betrachte die folgenden Aussagen:*

(a) *μ ist σ-additiv.*

(b) *Für alle $A_n \in \mathscr{A}$ mit $A_1 \subset A_2 \subset \dots$ und $\bigcap_{n\in\mathbb{N}} A_n =: A \in \mathscr{A}$ gilt*

$$\lim_{n\to\infty} \mu(A_n) = \mu(A)$$

(*d.h. μ ist stetig von unten*).

(c) *Für alle $A_n \in \mathscr{A}$ mit $A_1 \supset A_2 \supset \dots$, $\bigcup_{n\in\mathbb{N}} A_n = \emptyset$ und $\mu(A_1) < \infty$ gilt*

$$\lim_{n\to\infty} \mu(A_n) = 0$$

(*d.h. μ ist stetig von oben*).

Dann gilt (a) $\Longleftrightarrow$ (b) $\Longrightarrow$ (c). *Falls μ endlich ist, sind alle drei Aussagen äquivalent.*

Beweis. (a) $\Rightarrow$ (b). Mit $A_0 := \emptyset$ und $\tilde{A}_n := A_n \backslash A_{n-1}$ ist $A = \dot{\bigcup\limits_{n\in\mathbb{N}}} \tilde{A}_n$ und $A_n = \bigcup\limits_{k=1}^{n} \tilde{A}_k$. Also ist

$$\mu(A) = \sum_{n\in\mathbb{N}} \mu(\tilde{A}_n) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(\tilde{A}_k) = \lim_{n\to\infty} \mu(A_n)\,.$$

(b) ⇒ (a). Sei $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ paarweise disjunkt, $A := \bigcup_{n\in\mathbb{N}} A_n \in \mathscr{A}$. Setze $\tilde{A}_n := A_1 \dot\cup \ldots \dot\cup A_n$. Dann gilt $\tilde{A}_n \nearrow A$ (d.h. $\tilde{A}_1 \subset \tilde{A}_2 \subset \ldots$ und $\bigcup \tilde{A}_n = A$), und nach (b) gilt $\mu(\tilde{A}_n) \to \mu(A)$. Wegen $\mu(\tilde{A}_n) = \sum_{k=1}^n \mu(A_k)$ gilt also $\sum_{k=1}^\infty \mu(A_k) = \mu(A)$.

(b) ⇒ (c). Wegen $\mu(A_1 \backslash A_n) = \mu(A_1) - \mu(A_n)$ und $A_1 \backslash A_n \nearrow A_1$ gilt nach (b)

$$\mu(A_1) = \lim_{n\to\infty} \mu(A_1\backslash A_n) = \mu(A_1) - \lim_{n\to\infty} \mu(A_n)$$

und damit $\mu(A_n) \to 0$.

Sei nun μ endlich.

(c) ⇒ (d). Falls $A_n \nearrow A$, gilt $A\backslash A_n \searrow \emptyset$ und damit gilt $\mu(A\backslash A_n) \to 0$ nach (c). Somit folgt $\mu(A_n) \to \mu(A)$. □

In vielen Fällen ist nicht ein Maß auf einer σ-Algebra gegeben, sondern ein Inhalt auf einer Algebra oder einem Ring. Daher stellt sich die Frage, ob sich dieser Inhalt eindeutig zu einem Maß fortsetzen läßt. Die folgende Konstruktion liefert die Antwort.

Definition 1.8. Sei μ ein σ-additiver Inhalt auf einem Ring $\mathscr{A}$. Dann heißt $\mu^* : \mathscr{P}(\Omega) \to [0,\infty]$, definiert durch

$$\mu^*(B) := \begin{cases} \inf\left\{ \sum_{n\in\mathbb{N}} \mu(A_n) : A_n \in \mathscr{A},\ B \subset \bigcup_{n\in\mathbb{N}} A_n \right\}, & \text{falls } \{\cdots\} \neq \emptyset\,, \\ \infty\,, & \text{sonst,} \end{cases}$$

das zu μ gehörige äußere Maß.

Eine Menge $A \subset \Omega$ heißt meßbar, falls

$$\mu^*(A) = \mu^*(A\cap B) + \mu^*(A \cap B^c) \quad \text{für alle } B \in \mathscr{P}(\Omega)\,.$$

Satz 1.9. (Caratheodory) *Sei μ ein σ-additiver, σ-finiter Inhalt auf einem Ring $\mathscr{A}$. Dann ist das Mengensystem $\overline{\sigma}(\mathscr{A})$ aller μ^*-meßbaren Mengen eine ($\sigma(\mathscr{A})$ enthaltende) σ-Algebra, und $\mu^*|_{\overline{\sigma}(\mathscr{A})}$ ist ein Maß. Das Maß $\mu^*|_{\sigma(\mathscr{A})}$ ist die einzige Maßfortsetzung von μ.*

Der Beweis findet sich etwa im Buch von Bauer [1] oder im Buch von Halmos [8].

Die σ-Algebra $\overline{\sigma}(\mathscr{A})$ hat selbst eine Bedeutung:

Definition 1.10. Ein Maß μ auf einer σ-Algebra $\mathscr{A}$ heißt vollständig, falls gilt: Aus $A \subset B$, $B \in \mathscr{A}$ und $\mu(B) = 0$ folgt $A \in \mathscr{A}$. Ein vollständiges Maß $\mu : \mathscr{A} \to [0,\infty]$ heißt Vervollständigung des Maßes $\mu_0 : \mathscr{A}_0 \to [0,\infty]$, falls $\mathscr{A}_0 \subset \mathscr{A}$, $\mu|_{\mathscr{A}_0} = \mu_0$ und folgende (universelle) Eigenschaft gilt:

Sei $\mu' : \mathscr{A}' \to [0,\infty]$ vollständige Fortsetzung von μ_0. Dann ist $\mathscr{A}' \subset \mathscr{A}$ und $\mu'|_{\mathscr{A}} = \mu$ (d.h. μ ist minimale vollständige Fortsetzung von μ_0).

Der folgende Satz besagt, daß $\mu^*|_{\overline{\sigma}(\mathscr{A})}$ die Vervollständigung des von μ auf $\sigma(\mathscr{A})$ induzierten Maßes ist, und dies zugleich eine Art Abschluß darstellt. Dazu benutzt man die Halbmetrik (!)

$$d_{\mu^*}(A,B) := \mu^*(A \triangle B) \quad \text{auf } \mathscr{P}(\Omega) \times \mathscr{P}(\Omega)\,.$$

Dabei ist $A \triangle B := (A \backslash B) \cup (B \backslash A)$.

Satz 1.11. *Sei μ endlicher, σ-additiver Inhalt auf einer Algebra $\mathscr{A}$. Dann gilt:*

a) $\mu^*|_{\overline{\sigma}(\mathscr{A})}$ *ist die Vervollständigung des Maßes* $\mu^*|_{\sigma(\mathscr{A})}$.

b) $\overline{\sigma}(\mathscr{A}) = \{B \in \mathscr{P}(\Omega) :$ *Für alle* $\varepsilon > 0$ *existiert ein* $A \in \mathscr{A}$ *mit* $d_{\mu^*}(A,B) < \varepsilon\}$ (*d.h.* $\overline{\sigma}(\mathscr{A})$ *ist der Abschluß von* $\mathscr{A}$ *bzgl.* d_{μ^*} *in* $\mathscr{P}(\Omega)$).

(Beweis siehe etwa Halmos [8].)

Bemerkung 1.12. a) Die obigen Sätze besagen, daß ein σ-finiter und σ-additiver Inhalt auf einer Algebra (oder einem Ring) bereits eindeutig ein Maß auf der erzeugten σ-Algebra definiert. Das zum elementargeometrischen Inhalt (Beispiel 1.5 c) gehörige Maß heißt Lebesgue-Maß λ, die σ-Algebra ist die Borel-σ-Algebra $\mathscr{B}(\mathbb{R})$. Die zur Vervollständigung gehörige σ-Algebra heißt das System aller Lebesgue-meßbaren Mengen.

b) Es gibt viele verschiedene Beschreibungen von $\overline{\sigma}(\mathscr{A})$, etwa

$$\overline{\sigma}(\mathscr{A}) = \{A \triangle N : A \in \sigma(\mathscr{A})\,,\ N \subset \tilde{N} \in \sigma(\mathscr{A}) \text{ mit } \mu(\tilde{N}) = 0\}\,.$$

Es gilt für das Lebesgue-Maß: Die Mächtigkeit (Kardinalität) von $\mathscr{B}(\mathbb{R})$ ist dieselbe wie die von $\mathbb{R}$, aber die Kardinalität der Lebesgue-meßbaren Mengen ist $2^{|\mathbb{R}|}$ und damit größer. Das letzte sieht man, indem man eine Menge C mit $|C| = |\mathbb{R}|$ und $\lambda(C) = 0$ angibt (z.B. die Cantor-Menge). Dann ist jede Teilmenge von C Lebesgue-meßbar. Es gibt also i.a. sehr viel mehr Mengen in der Vervollständigung $\overline{\sigma}(\mathscr{A})$ als in $\sigma(\mathscr{A})$.

Definition 1.13. Seien $(\Omega, \mathscr{A})$ und $(S, \mathscr{S})$ Meßräume. Für $X \colon \Omega \to S$ definiere

$$X^{-1}(B) := \{X \in B\} := \{\omega \in \Omega : X(\omega) \in B\} \quad (B \in \mathscr{P}(S))$$

und

$$X^{-1}(\mathscr{S}) := \{X^{-1}(B) : B \in \mathscr{S}\}\,.$$

Dann heißt X $\mathscr{A}$-$\mathscr{S}$-meßbar, falls für alle $B \in \mathscr{S}$ gilt $X^{-1}(B) \in \mathscr{A}$ (d.h. falls $X^{-1}(\mathscr{S}) \subset \mathscr{A}$ gilt).

Bemerkung 1.14. a) Jede konstante Funktion ist meßbar bezüglich jeder σ-Algebra.

b) Sind $X \colon (\Omega, \mathscr{A}) \to (S_1, \mathscr{S}_1)$ und $Y \colon (S_1, \mathscr{S}_1) \to (S_2, \mathscr{S}_2)$ meßbar, so auch $Y \circ X \colon (\Omega, \mathscr{A}) \to (S_2, \mathscr{S}_2)$. Denn es gilt

$$(Y \circ X)^{-1}(\mathscr{S}_2) = X^{-1}(Y^{-1}(\mathscr{S}_2)) \subset X^{-1}(\mathscr{S}_1) \subset \mathscr{A}\,.$$

Lemma 1.15. *Seien* $(\Omega, \mathscr{A})$ *und* $(S, \mathscr{S})$ *Meßräume und* $\mathscr{S} = \sigma(\mathscr{E})$ *(d.h.* $\mathscr{E}$ *ist ein Erzeugendensystem von* $\mathscr{S}$*). Dann ist* $X : \Omega \to S$ *genau dann* $\mathscr{A}$*-*$\mathscr{S}$*-meßbar, wenn* $X^{-1}(\mathscr{E}) \subset \mathscr{A}$*.*

Beweis. Das Mengensystem $\mathscr{S}' := \{B \in \mathscr{P}(S) : X^{-1}(B) \in \mathscr{A}\}$ ist eine σ-Algebra über S. Nach Definition ist X genau dann $\mathscr{A}$-$\mathscr{S}$-meßbar, wenn $\mathscr{S} = \sigma(\mathscr{E}) \subset \mathscr{S}'$. Dies ist aber äquivalent zu $\mathscr{E} \subset \mathscr{S}'$, d.h. zu $X^{-1}(\mathscr{E}) \subset \mathscr{A}$. □

Falls Ω und S topologische Räume sind (mit Topologien τ_Ω und τ_S), so wählt man gewöhnlich

$$\mathscr{A} := \mathscr{B}(\Omega) := \sigma(\tau_\Omega)$$

und $\mathscr{S} := \mathscr{B}(S)$ als σ-Algebren, die sogenannten Borel-σ-Algebren. In diesem Fall spricht man kurz von Borel-meßbar oder auch nur von meßbar. Aus obigem Lemma folgt sofort (mit $\mathscr{E} = \tau_S$), daß eine stetige Abbildung $X : \Omega \to S$ Borel-meßbar ist. Falls $(S, \mathscr{S}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ oder $(S, \mathscr{S}) = (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ gilt, so folgt aus Lemma 1.15, daß X genau dann $\mathscr{A}$-meßbar ist, falls

$$\{X \leq \alpha\} \in \mathscr{A} \qquad \text{für alle } \alpha \in \mathbb{R}\,.$$

2. Zufallsexperimente und Wahrscheinlichkeit

Was ist eine Wahrscheinlichkeit? Jeder wird antworten, daß bei einem Werfen eines Würfels die Wahrscheinlichkeit für die Zahl 1 den Wert $\frac{1}{6}$ besitzt. Aber wie kann man dies begründen? In der Vergangenheit wurde der Versuch gemacht, die Wahrscheinlichkeit eines Ereignisses A als Limes der relativen Häufigkeit des Auftretens von A zu definieren. Eine solche Definition (zu finden etwa bei Mises (1919)) stößt jedoch auf mathematische Schwierigkeiten. Daher definiert man heute die Wahrscheinlichkeit axiomatisch, nämlich als W-Maß (nach Kolmogorov (1933)). Es wird sich später zeigen, daß sich die Häufigkeitsinterpretation bei dieser Wahl der Axiome präzisieren und beweisen läßt (Gesetze der großen Zahl).

Definition 2.1. Ein Zufallsexperiment ist ein W-Raum $(\Omega, \mathscr{A}, P)$ (d.h. ein Maßraum, wobei das Maß ein W-Maß ist) mit folgender Interpretation:

a) $x \in \Omega$ heißt Ergebnis oder mögliche Realisierung des Experiments.

b) $A \in \mathscr{A}$ heißt Ereignis, d.h. eine Menge von Ergebnissen. (Die Wahl der σ-Algebra $\mathscr{A}$ ist aus dem Experiment oder mathematisch begründet).

c) $P(A)$ heißt die Wahrscheinlichkeit für das Eintreten von $A \in \mathscr{A}$. Das Maß $P\colon \mathscr{A} \to [0,1]$ heißt die zum Experiment gehörige W-Verteilung.

Bemerkung 2.2. a) Die Forderung, daß $\mathscr{A}$ eine Algebra ist, scheint natürlich. Dabei entspricht $\emptyset$ dem unmöglichen Ereignis, $A \cap B$ dem gleichzeitigen Eintreten von A und B und A^c der logischen Negation von A. Analog ist die Forderung, daß P ein Inhalt ist, naheliegend (aus der Häufigkeitsinterpretation). Die Bedingung, daß $\mathscr{A}$ eine σ-Algebra und P ein Maß (d.h. σ-additiv) sind, ist eine mathematische Idealisierung (vgl. dazu auch Satz 1.7, der dies als Stetigkeit beschreibt).

b) Warum betrachtet man überhaupt verschiedene σ-Algebren und nimmt nicht stets $\mathscr{A} = \mathscr{P}(\Omega)$? Ein tiefliegender Satz von Ulam (siehe etwa [11], S. 29, Satz 5.6) besagt unter Annahme der Kontinuumshypothese, daß es kein W-Maß P auf der Potenzmenge von $[0,1]$ gibt mit $P(\{x\}) = 0$ für alle $x \in [0,1]$. Insbesondere ist das Lebesgue-Maß $\lambda : \mathscr{B}([0,1]) \to [0,1]$ nicht auf $\mathscr{P}([0,1])$ fortsetzbar. Falls Ω abzählbar ist (insbesondere falls Ω endlich ist), wird man stets $\mathscr{A} = \mathscr{P}(\Omega)$ wählen.

Beispiele 2.3. a) Laplace-Experiment: Hier ist $|\Omega| < \infty$, $\mathscr{A} = \mathscr{P}(\Omega)$ und

$$P(A) = \frac{|A|}{|\Omega|} \quad \text{für alle } A \in \mathscr{P}(\Omega)\,.$$

b) Gleichverteilung auf dem Intervall $[a,b]$ $(-\infty < a < b < \infty)$: Hier ist $\Omega = [a,b]$, $\mathscr{A} = \mathscr{B}(\Omega)$ und $P(A) = \frac{\lambda(A)}{\lambda(\Omega)}$ für alle $A \in \mathscr{A}$.

c) Mehrstufige Experimente: Hier wird ein Zufallsexperiment $(\Omega_1, \mathscr{A}_1, P_1)$ n-fach wiederholt. Man erhält $\Omega = \Omega_1 \times \ldots \times \Omega_1$, $\mathscr{A} = \mathscr{A}_1 \otimes \ldots \otimes \mathscr{A}_1$, und für das zugehörige

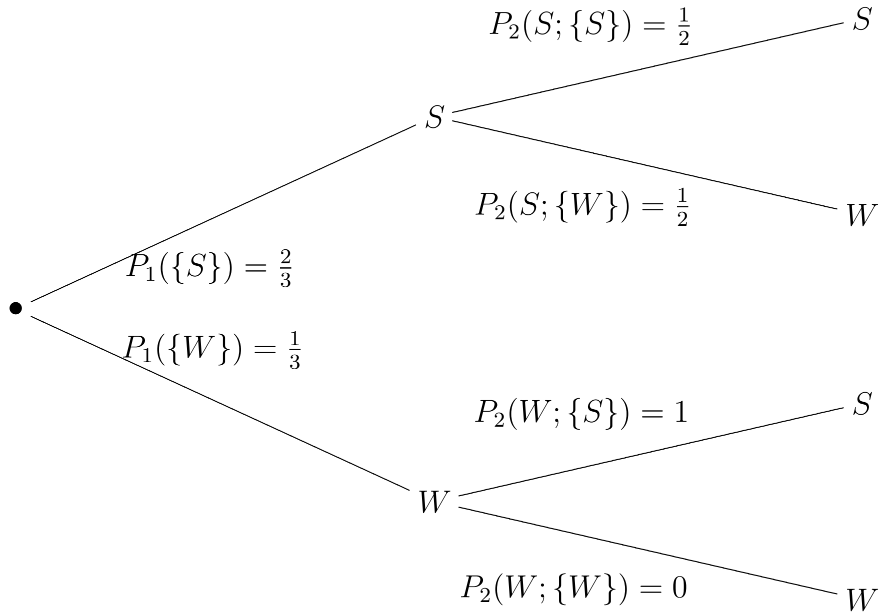


Abbildung 1: Beispiel eines Baumdiagramms.

Maß P gilt, falls $|\Omega| < \infty$, die Darstellung

$$P(\{(x_1, \dots, x_n)\} = P(\{x_1\}) \cdot P_2(x_1; \{x_2\}) \cdot P_3(x_1, x_2; \{x_3\}) \cdot \ldots \\ \ldots \cdot P_n(x_1, \dots, x_{n-1}; \{x_n\}),$$

wobei $P(x_1, \dots, x_{k-1}; \cdot)$ die Übergangswahrscheinlichkeit für die k-te Wiederholung ist (abhängig von der bereits erzielten Realisierung $x_1, \dots, x_{k-1}$).

Beispiel: Ziehen von zwei Kugeln aus einer Urne mit zwei schwarzen und einer weißen Kugel. Hier ist $\Omega_1 = \{S, W\}$ und $\mathscr{A}_1 = \mathscr{P}(\Omega_1)$. Die Übergangswahrscheinlichkeiten werden am besten durch ein Baumdiagramm beschrieben (Abbildung 1).

Definition 2.4. Sei $(\Omega, \mathscr{A}, P)$ ein W-Raum. Dann heißt P

(i) auf $A \in \mathscr{A}$ konzentriert, falls $P(A) = 1$,

(ii) diskret, falls P auf einer abzählbaren Menge konzentriert ist.

Satz 2.5. *Sei $(\Omega, \mathscr{A}, P)$ W-Raum mit $\{x\} \in \mathscr{A}$ für alle $x \in \Omega$. Dann ist die Menge $\Omega_0 := \{x \in \Omega : P(\{x\}) > 0\}$ abzählbar, und folgende Aussagen sind äquivalent:*

(i) *P ist diskret.*

(ii) $P(\Omega_0) = 1$.

(iii) *Es gilt*

$$\int f \mathrm{d}P = \sum_{x\in\Omega_0} f(x)P(\{x\})$$

für alle beschränkten und $\mathscr{A}$*-meßbaren* $f : \Omega \to \mathbb{R}$.

(iv) *Es gilt*

$$P = \sum_{x\in\Omega_0} P(\{x\})\delta_x\,.$$

Dabei ist (iv) als Gleichheit von Abbildungen zu verstehen, d.h. diese Gleichheit gilt punktweise für jede Menge $A \in \mathscr{A}$. Man beachte jedoch, daß δ_x auf $\mathscr{P}(\Omega)$ definiert ist und damit wegen (iv) jedes diskrete Maß auf $\mathscr{P}(\Omega)$ definiert werden kann und üblicherweise definiert wird.

Beweis. Betrachte die Mengen $A_n := \{x \in \Omega : P(\{x\}) \geq \frac{1}{n}\}$. Dann enthält A_n endlich viele, nämlich nicht mehr als n, Elemente, und daher ist $\Omega_0 = \bigcup_{n\in\mathbb{N}} A_n$ als abzählbare Vereinigung endlicher Mengen abzählbar.

(i) $\Rightarrow$ (ii). Sei $A \in \mathscr{A}$ abzählbar mit $P(A) = 1$. Dann ist

$$1 = P(A) = \sum_{x\in A} P(\{x\}) \leq \sum_{x\in\Omega_0} P(\{x\}) = P(\Omega_0) \leq P(\Omega) = 1\,.$$

(ii) $\Rightarrow$ (iii). Es gilt $f\cdot 1_{A_n} \to f\cdot 1_{\Omega_0}$ punktweise und $|f\cdot 1_{A_n}| \leq \sup_{x\in\Omega}|f(x)| =: \|f\|_\infty$. Nach dem Satz über majorisierte Konvergenz folgt

$$\int f 1_{A_n} \mathrm{d}P = \sum_{x\in A_n} f(x)P(\{x\}) \longrightarrow \int_{\Omega_0} f \mathrm{d}P = \int_\Omega f \mathrm{d}P\,.$$

Dabei wurde bei der letzten Gleichheit $P(\Omega_0) = 1$ verwendet. Für $n \to \infty$ konvergiert die Summe über $x \in A_n$ gegen $\sum_{x\in\Omega_0} f(x)P(\{x\})$; dabei konvergiert diese Reihe absolut wegen $\sum_{x\in\Omega_0}|f(x)|P(\{x\}) \leq \|f\|_\infty P(\Omega_0)$.

(iii) $\Rightarrow$ (iv). Wähle $f = 1_A$ mit $A \in \mathscr{A}$.

(iv) $\Rightarrow$ (i). Wähle $A = \Omega$ und schreibe $P(A)$ unter Verwendung von (iv). $\square$

Im folgenden wird statt $P(\{\ldots\})$ auch $P\{\ldots\}$ geschrieben.

3. Endliche Wahrscheinlichkeitsräume

In diesem Abschnitt ist $|\Omega| < \infty$ und $\mathscr{A} = \mathscr{P}(\Omega)$. Für die folgenden kombinatorischen Schlußweisen ist das Abzählprinzip nützlich, das man mit Induktion leicht beweisen kann:

Abzählprinzip. Sei Ω eine Menge von Tupeln $(x_1, \dots, x_n)$, wobei x_i das Ergebnis der i-ten Stufe eines n-stufigen Experiments sei. Für alle $i = 1, \dots, n$ sei die Anzahl k_i der möglichen Ausgänge des i-ten Teilexperiments unabhängig von dem bereits realisierten Ergebnis $(x_1, \dots, x_{i-1})$. Dann gilt $|\Omega| = k_1 \cdot \ldots \cdot k_n$.

Definition 3.1. a) Für $k, n \in \mathbb{N}_0$ heißt

$$\binom{n}{k} := \begin{cases} \dfrac{n!}{k!(n-k)!} & \text{falls } k \leq n\,, \\ 0 & \text{sonst,} \end{cases}$$

der Binomialkoffizient „k aus n“ oder „n über k“.

b) Für $n = k_1 + \ldots + k_r$ heißt

$$\binom{n}{k_1, \dots, k_r} := \frac{n!}{k_1! \cdot \ldots \cdot k_r!}$$

Multinomialkoeffizient.

Satz 3.2. *Sei M eine Menge mit $|M| = n < \infty$.*

a) M besitzt genau $\binom{n}{k}$ Teilmengen mit k Elementen.

b) Sei $M = N_1 \dot\cup \ldots \dot\cup N_r$ mit $|N_j| = k_j$. Betrachtet man die Elemente jeder Teilmenge N_j als gleich, so gibt es genau $\binom{n}{k_1, \dots, k_r}$ Möglichkeiten, die Elemente von M anzuordnen.

Beweis. a) Betrachte zunächst die Menge der Tupel $(x_1, \dots, x_k)$ mit $x_i \neq x_j$ für $i \neq j$. Für die Wahl von x_j gibt es $n + 1 - j$ Möglichkeiten ($j = 1, \dots, k$), also insgesamt $n(n-1) \cdot \ldots \cdot (n-k+1)$ Möglichkeiten (Abzählprinzip). Je $k!$ solcher Tupel führen zur selben Menge $\{x_1, \dots, x_k\}$ (Permutationen), also gibt es

$$\frac{n(n-1) \cdot \ldots \cdot (n-k+1)}{k!} = \binom{n}{k}$$

Teilmengen von M mit k Elementen.

b) Von den $n!$ Permutationen von M sind $k_1! \cdot \ldots \cdot k_r!$ als gleich anzusehen, da sie sich nur um Permutationen der Mengen N_j unterscheiden. Somit gibt es

$$\frac{n!}{k_1! \cdot \ldots \cdot k_r!} = \binom{n}{k_1, \dots, k_r}$$

Möglichkeiten. $\square$

Satz 3.3 (Urnenmodelle). *Aus einer Urne mit n Kugeln werden k Kugeln nacheinander gezogen. Dann gibt es folgende Anzahl möglicher Versuchsergebnisse:*

	mit Zurücklegen	ohne Zurücklegen
unter Beachtung der Reihenfolge	n^k	$n(n-1)\cdot\ldots\cdot(n-k+1)$
ohne Beachtung der Reihenfolge	$\binom{n+k-1}{k}$	$\binom{n}{k}$

Beweis. Bis auf die linke untere Ecke ist alles klar nach Satz 3.2 bzw. dessen Beweis. Beschreibe die Ergebnismenge des Versuchs mit Zurücklegen ohne Beachtung der Reihenfolge als

$$\Omega = \{(x_1, \ldots, x_k) \in M^k : x_1 \leq \ldots \leq x_k\},$$

wobei $M = \{1, \ldots, n\}$. Die Abbildung

$$\Omega \to \Omega', \quad (x_1, \ldots, x_k) \mapsto (x_1, x_2 + 1, \ldots, x_k + k - 1)$$

ist eine Bijektion von Ω nach

$$\Omega' := \{(y_1, \ldots, y_k) \in \{1, \ldots, n + k - 1\}^k : y_1 < \ldots < y_k\}.$$

Nach Satz 3.2 a) ist aber $|\Omega'| = \binom{n+k-1}{k}$. □

Satz 3.4. *Eine Urne enthalte n Kugeln, von denen m weiß und $n-m$ schwarz seien. Man zieht nacheinander l Kugeln. Die Größe $X \in \{0, \ldots, l\}$ beschreibe die Anzahl der gezogenen weißen Kugeln. Bei Annahme eines Laplace-Eperiments erhält man folgende Wahrscheinlichkeiten für die Werte von X:*

a) *Hypergeometrische Verteilung: Bei Ziehen ohne Zurücklegen ist die Wahrscheinlichkeit für $X = k$ gegeben durch*

$$H(l; m; n)\{k\} := \frac{\binom{m}{k}\binom{n-m}{l-k}}{\binom{n}{l}} \qquad \text{für } k \in \mathbb{N}_0,\ 0 \leq k \leq \min\{m, l\}.$$

b) *Binomialverteilung: Bei Ziehen mit Zurücklegen ist die Wahrscheinlichkeit für $X = k$ gegeben durch*

$$B(l; p)\{k\} := \binom{l}{k} p^k (1-p)^{l-k} \qquad \text{für } k \in \mathbb{N}_0,\ 0 \leq k \leq l,$$

wobei $p := \frac{m}{n}$ gesetzt wurde.

Beweis. a) Es gibt $\binom{n}{l}$ Möglichkeiten, l Kugeln ohne Zurücklegen zu ziehen. Für das Eintreten von $\{X = k\}$ müssen wir k Kugeln aus den m weißen und $l - k$ Kugeln aus den $n - m$ schwarzen ziehen, dafür gibt es $\binom{n}{k}\binom{n-m}{l-k}$ Möglichkeiten.

b) Bei Ziehen mit Reihenfolge gibt es (mit Zurücklegen) n^k mögliche Ziehungen insgesamt. Für das Eintreten des Ereignisses $\{X = k\}$ muß man zunächst k Ziehungen (von insgesamt l Ziehungen) bestimmen, in welchen weiße Kugeln gezogen werden. Dies ergibt $\binom{l}{k}$ Möglichkeiten. Dann muß man bei diesen k Ziehungen je eine weiße Kugel ziehen (m^k Möglichkeiten), bei den anderen $l - k$ Ziehungen je eine schwarze ($(n - m)^{l-k}$ Möglichkeiten). Insgesamt erhält man

$$\frac{\binom{l}{k} m^k (n-m)^{l-k}}{n^k} = \binom{l}{k} p^k (1-p)^{l-k}$$

als Wahrscheinlichkeit für das Ereignis $\{X = k\}$. □

Definition 3.5. Sei $(\Omega, \mathscr{A}, P)$ ein Wahrscheinlichkeitsraum (nicht notwendig endlich). Für $A, B \in \mathscr{A}$ mit $P(B) > 0$ heißt

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

die bedingte Wahrscheinlichkeit von A unter der Bedingung B.

Bemerkung 3.6. a) Für $P(B) > 0$ ist $P(\,\cdot\,|B)$ wieder ein W-Maß auf $\mathscr{A}$.

b) Sei $\Omega = \dot{\bigcup\limits_{i\in\mathbb{N}}} B_i$ mit $B_i \in \mathscr{A}$. Dann ist

$$P(A) = \sum_{\substack{i\in\mathbb{N} \\ P(B_i)>0}} P(B_i) P(A|B_i) \quad \text{für alle } A \in \mathscr{A}\,,$$

und, falls $A \in \mathscr{A}$ mit $P(A) > 0$,

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum\limits_{\substack{j\in\mathbb{N} \\ P(B_j)>0}} P(B_j)P(A|B_j)}$$

(Formel von Bayes). Dies gilt wegen $P(B_i)P(A|B_i) = P(A \cap B_i)$ und

$$P(A) = P\Big(\dot{\bigcup_j} (A \cap B_j)\Big) = \sum_j P(A \cap B_j)\,.$$

4. Zufallsvariablen, Erwartungswert und Varianz

In den meisten Fällen interessiert man sich nicht für die Ergebnisse $x \in \Omega$ eines Experiments, sondern nur für Funktionen dieser Ergebnisse. Zum Beispiel bei einem dreimaligen Werfen eines Laplace-Würfels mit $\Omega = \{1, \dots, 6\}^3$ für die Anzahl X der Sechser, d.h. für die Werte der Funktion $X\colon \Omega \to \{0, 1, 2, 3\}$ mit $X =$ Anzahl der Sechser.

Allgemein führt dies zu folgender Definition:

Definition 4.1. a) Sei Ω eine Menge. Sei $(X_i)_{i \in I}$ eine Familie von Funktionen $X_i\colon \Omega \to S_i$, wobei $(S_i, \mathscr{S}_i)$ ein Meßraum ist. Dann bezeichnet

$$\sigma((X_i)_{i \in I}) := \sigma\Big(\bigcup_{i \in I} X_i^{-1}(\mathscr{S}_i)\Big)$$

die von $(X_i)_i$ erzeugte σ-Algebra über Ω.

b) Seien $(\Omega, \mathscr{A})$ und $(S, \mathscr{S})$ Meräume. Dann heißt eine Abbildung $X\colon \Omega \to S$ eine Zufallsfunktion, falls X $\mathscr{A}$-$\mathscr{S}$-meßbar ist, d.h. wenn gilt

$$X^{-1}(B) \in \mathscr{A} \qquad \text{für alle } B \in \mathscr{S}\,.$$

Das Bildmaß $P \circ X^{-1}\colon \mathscr{S} \to [0,1]\,,\ B \mapsto P(X^{-1}(B)) = P\{X \in B\}$ heißt die Verteilung oder W-Verteilung von X.

c) Nun sei $(S, \mathscr{S}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ und X $\mathscr{A}$-meßbar. Dann heißt X Zufallsvariable (ZV), und

$$F_X\colon \mathbb{R} \to [0,1]\,,\ F_X(t) := P \circ X^{-1}(-\infty, t] = P\{X \le t\}$$

die Verteilungsfunktion von X. Die Zufallsvariable X heißt diskret verteilt, falls $P \circ X^{-1}$ diskret ist, und stetig verteilt, falls $P \circ X^{-1}$ absolutstetig bzgl. des Lebesgue-Maßes ist, d.h. eine Dichte f_X existiert mit $P \circ X^{-1} = f_X(t)\mathrm{d}t$.

Bemerkung 4.2. a) F_X ist monoton wachsend, rechtstetig (r.c.), d.h. $F_X(t+0) = F_X(t)$, und es gilt $\lim_{t \to -\infty} F_X(t) = 0$ und $\lim_{t \to \infty} F_X(t) = 1$. Außerdem gilt

$$F_X(t) - F_X(t-0) = P \circ X^{-1}\{t\}\,.$$

b) X ist genau dann stetig verteilt, wenn F_X absolutstetig ist, d.h. wenn $F_X(t) = \int_{-\infty}^t f_X(t)\mathrm{d}t$ mit einer meßbaren Funktion $f_X\colon \mathbb{R} \to \mathbb{R}_+$ mit $\int_{\mathbb{R}} f_X(t)\mathrm{d}t = 1$ gilt. Dies ist z.B. dann der Fall, wenn F_X stetig differenzierbar ist.

Satz 4.3. *Sei $(\Omega, \mathscr{A}, P)$ ein W-Raum.*

a) *Falls X und Y Zufallsvariablen sind, dann auch* $\max\{X, Y\}$, $\min\{X, Y\}$, $|X|^r$ *für $r > 0$ und X^r für $r \in \mathbb{N}$.*

b) *Falls $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen ist, dann sind* $\inf_n X_n$, $\sup_n X_n$, $\liminf_n X_n$ *und* $\limsup_n X_n$ *Zufallsfunktionen von Ω nach $\overline{\mathbb{R}}$ (mit σ-Algebra $\mathscr{B}(\overline{\mathbb{R}})$).*

Beweis. a) $(X,Y)\colon \Omega \to \mathbb{R}^2$ ist $\mathscr{A}$-$\mathscr{B}(\mathbb{R}^2)$-meßbar. Die Abbildung $f\colon \mathbb{R}^2 \to \mathbb{R}$, $(x,y) \mapsto \max\{x,y\}$ ist Borel-meßbar, somit ist $f \circ (X,Y)$ eine Zufallsvariable. Der Rest folgt analog.

b) Es gilt

$$\{\sup_n X_n \le \alpha\} = \bigcap_n \{X_n \le \alpha\} \in \mathscr{A}$$

für alle $\alpha \in \mathbb{R}$. Mit Lemma 1.15 folgt die Behauptung für das Supremum. Wegen $\inf X_n = -\sup(-X_n)$ und $\limsup X_n = \inf_n \sup_{k\ge n} X_k$ folgt der Rest daraus. □

Satz 4.4. *Sei $\mathscr{M}$ die Menge aller W-Maße auf $\mathscr{B}(\mathbb{R})$ und $\mathscr{F}$ die Menge aller Abbildungen von $\mathbb{R}$ nach $\mathbb{R}$, welche monoton wachsend und r.c. sind und gegen 0 bzw. 1 konvergieren für $t \to -\infty$ bzw. $t \to \infty$. Dann ist die Abbildung $\mu \mapsto F_\mu$ mit $F_\mu(t) := \mu(-\infty, t]$ eine Bijektion.*

Der Beweis dafür wird hier weggelassen (siehe Maßtheorie-Vorlesung).

Definition 4.5. Zwei ZV X und Y heißen stochastisch äquivalent ($X \sim Y$), falls $P \circ X^{-1} = P \circ Y^{-1}$ gilt (nach Satz 4.4 ist dies genau dann der Fall, falls $F_X = F_Y$ gilt).

Dies bedeutet nicht, daß $X = Y$ oder $P\{X = Y\} = 1$ gilt, wie folgendes Beispiel zeigt:

Beispiel 4.6. Es sei

$$(\Omega, \mathscr{A}, P) = \Big(\{0,1\}, \mathscr{P}(\{0,1\}), \frac{1}{2}(\delta_0 + \delta_1)|_{\mathscr{P}(\{0,1\})}\Big)$$

(d.h. es gilt $P\{0\} = P\{1\} = \frac{1}{2}$). Setzt man $X(t) := t$ und $Y(t) := 1 - t$ für $t \in \Omega$, so ist $P \circ X^{-1} = P \circ Y^{-1} = \frac{1}{2}(\delta_0 + \delta_1)|_{\mathscr{B}(\mathbb{R})}$, aber es gilt $X(t) \ne Y(t)$ für alle $t \in \Omega$.

Nun soll als nächstes der Erwartungswert einer Zufallsvariablen definiert werden. Da dieser als ein Integral definiert wird, muß man etwas über (allgemeine) Lebesgue-Integrale wissen. Diese sind für integrierbare Zufallsvariablen oder für nichtnegative Zufallsvariablen definiert, wobei dann auch der Wert ∞ auftauchen kann.

Satz 4.7. (Was man auch im Schlaf über das Integral wissen muß.) *Seien X und Y Zufallsvariablen über $(\Omega, \mathscr{A}, P)$.*

(i) *X ist genau dann integrierbar, wenn $\int |X| \mathrm{d}P < \infty$.*

(ii) *Seien X und Y integrierbar und $a, b \in \mathbb{R}$. Dann gilt*

$$\int (aX + bY) \mathrm{d}P = a \int X \mathrm{d}P + b \int Y \mathrm{d}P$$

(Linearität).

(iii) *Seien X integrierbar und A_n, $A \in \mathscr{A}$ mit $\dot{\bigcup}_{n\in\mathbb{N}} A_n = A$. Dann ist*

$$\int_A X \mathrm{d}P = \sum_{n\in\mathbb{N}} \int_{A_n} X \mathrm{d}P\,.$$

(iv) *Aus $X \geq 0$ P-f.s. folgt $\int X \mathrm{d}P \geq 0$ (Positivität).*

(v) *Seien X, Y integrierbar mit $X \leq Y$ P-f.s. Dann gilt $\int X \mathrm{d}P \leq \int Y \mathrm{d}P$ (Monotonie).*

(vi) *Sei X integrierbar und $a, b \in \mathbb{R}$ mit $a \leq X \leq b$ P-f.s. auf einer Menge $A \in \mathscr{A}$. Dann gilt*

$$aP(A) \leq \int_A X \mathrm{d}P \leq bP(A)\,.$$

(vii) *Falls X integrierbar ist, gilt $|\int X \mathrm{d}P| \leq \int |X| \mathrm{d}P$.*

(viii) *Seien X_n und X Zufallsvariablen mit $|X_n| \leq Y$ P-f.s. , $X_n \to X$ P-f.s. und $\int Y \mathrm{d}P < \infty$. Dann ist auch X Zufallsvariable, und es gilt*

$$\lim_{n\to\infty} \int X_n \mathrm{d}P = \int X \mathrm{d}P$$

(Satz von der majorisierten Konvergenz).

(ix) *Seien X_n Zufallsvariable mit $\sum_n \int |X_n| \mathrm{d}P < \infty$. Dann gilt*

$$\sum_n |X_n| < \infty \quad P\text{-f.s.}\ ,$$

d.h. $X := \sum_n X_n$ existiert P-f.s. , und es gilt

$$\int X \mathrm{d}P = \sum_n \int X_n \mathrm{d}P$$

(gliedweise Integration).

(x) *Seien $X_n \geq 0$ Zufallsvariablen und X Zufallsvariable mit $X_n \nearrow X$ P-f.s. Dann ist $\lim_n \int X_n \mathrm{d}P = \int X \mathrm{d}P$, wobei auf beiden Seiten der Wert ∞ zugelassen ist (Satz von der monotonen Konvergenz).*

(xi) *Seien $X_n \geq 0$ Zufallsvariablen. Dann ist*

$$\int \liminf_{n\to\infty} X_n \mathrm{d}P \leq \liminf_{n\to\infty} \int X_n \mathrm{d}P$$

(Lemma von Fatou).

Definition 4.8. Sei X Zufallsvariable über $(\Omega, \mathscr{A}, P)$.

a) Falls $X \in L^1(P)$, heißt

$$\mathrm{E}\,X := \int_\Omega X \mathrm{d}P = \int_\mathbb{R} t(P \circ X^{-1})(\mathrm{d}t)$$

der Erwartungswert von X. (Auch sinnvoll falls $X \geq 0$ mit möglichem Wert ∞.)

b) Falls $X \in L^2(P)$ (und damit auch $X \in L^1(P)$), heißt

$$\mathrm{Var}\,X := \mathrm{E}[(X - \mathrm{E}\,X)^2] = \int_\Omega (X - \mathrm{E}\,X)^2 \mathrm{d}P$$

die Varianz von X. Die Zahl $\sigma := \sqrt{\mathrm{Var}\,X}$ heißt die Streuung von X.

c) Der Wert $\mu_n := \mathrm{E}\,X^n$ heßt das n-te Moment von X (falls existent).

Damit gelten alle Eigenschaften des Integrals für den Erwartungswert. So existiert etwa $\mathrm{E}\,X$ genau dann, wenn $\mathrm{E}\,|X| < \infty$ gilt. Man beachte auch $\mathrm{E}\,1 = 1$. Nach dem Transformationslemma gilt $X \in L^1(P)$ genau dann, wenn $\mathrm{id}_\mathbb{R} \in L^1(P \circ X^{-1})$. Die Existenz und der Wert von $\mathrm{E}\,X$ und $\mathrm{Var}\,X$ hängen nur von $P \circ X^{-1}$ ab. Für die Berechnung allgemeiner Integrale verwendet man ebenfalls das Transformationslemma. So gilt für $g \colon \mathbb{R} \to \mathbb{C}$ mit $g \circ X \in L^1(P)$

$$\begin{aligned}\int g \circ X \mathrm{d}P &= \int g\,\mathrm{d}(P \circ X^{-1}) \\ &= \begin{cases} \displaystyle\sum_{x : P\{X = x\}} g(x) P\{X = x\}\,, & \text{falls } X \text{ diskret }, \\ \displaystyle\int_\mathbb{R} g f_X \mathrm{d}\lambda\,, & \text{falls } X \text{ stetig verteilt mit Dichte } f_X\,. \end{cases}\end{aligned}$$

Satz 4.9. *Sei X eine Zufallsvariable auf $(\Omega, \mathscr{A}, P)$ und $f \colon \mathbb{R} \to \mathbb{C}$ stetig mit $f \circ X \in L^1(P)$. Dann gilt*

$$\int_\Omega f \circ X \mathrm{d}P = \int_{-\infty}^{\infty} f(t) \mathrm{d}F_X(t)\,,$$

wobei auf der rechten Seite das uneigentliche Riemann-Stieltjes-Integral steht.

Beweis. Das Integral $\int_{-\infty}^{\infty} f(t)\mathrm{d}F_X(t)$ ist definiert als Limes von $\int_a^b f(t)\mathrm{d}F_X(t)$ für $a \to -\infty$ und $b \to \infty$. Betrachte $\int_a^b f \mathrm{d}F_X := \lim_{|Z| \to 0} \sum_Z f \mathrm{d}F_X$ mit

$$\int_Z f\,\mathrm{d}F_X := \sum_{k=1}^{r} f(t_{k-1})\big(F_X(t_k) - F_X(t_{k-1})\big)$$

(Riemann-Stieltjes-Summe) für eine Zerlegung $Z = (t_0, \ldots, t_r)$ von $[a, b]$ (d.h. $a = t_0 < t_1 < \ldots < t_r = b$). Dabei ist $|Z| := \max_k(t_k - t_{k-1})$.

Zur Zerlegung Z definiere die Treppenfunktion

$$f_Z := \sum_{k=1}^{r} f(t_{k-1}) 1_{]t_k, t_{k-1}]} .$$

Da f stetig ist, gilt $f_Z \to f$ punktweise (hier genügt auch die Voraussetzung, daß f linksseitig stetig ist) für $|Z| \to 0$. Wegen $|f_Z(t)| \leq \sup_{t\in[a,b]} |f(t)|$ kann man majorisierte Konvergenz anwenden und erhält

$$\int_{]a,b]} f_Z \,\mathrm{d}(P \circ X^{-1}) \to \int_{]a,b]} f \mathrm{d}(P \circ X^{-1}) \quad \text{für } |Z| \to 0 .$$

Aber

$$\begin{aligned}\int_{]a,b]} f_Z \mathrm{d}(P \circ X^{-1}) &= \sum_{k=1}^{r} f(t_{k-1})(P \circ X^{-1})]t_{k-1}, t_k] \\ &= \sum_{k=1}^{r} f(t_{k-1})(F_X(t_k) - F_X(t_{k-1})) = \sum_{Z} f \mathrm{d}F_X .\end{aligned}$$

Somit ist $\int_{]a,b]} f \mathrm{d}(P \circ X^{-1}) = \int_a^b f(t) \mathrm{d}F_X(t)$. Wegen $f 1_{]a,b]} \to f$ für $a \to -\infty$ und $b \to \infty$ punktweise und $f \in L^1(P \circ X^{-1})$ folgt mit majorisierter Konvergenz:

$$\int_{\mathbb{R}} f \mathrm{d}P \circ X^{-1} = \lim_{\substack{a\to-\infty \\ b\to+\infty}} \int_{]a,b]} f \mathrm{d}P \circ X^{-1} = \int_{-\infty}^{\infty} f(t) \mathrm{d}F_X(t) .$$

□

Satz 4.10. *Sei X eine Zufallsvariable. Dann ist*

$$\sum_{n\in\mathbb{N}} P\{|X| \geq n\} \leq \mathrm{E}\,|X| \leq 1 + \sum_{n\in\mathbb{N}} P\{|X| \geq n\} .$$

Beweis. Sei $A_n := \{n \leq |X| < n+1\}$ für $n \in \mathbb{N}_0$. Dann gilt nach 4.6 (iii)

$$\mathrm{E}\,|X| = \sum_{n=0}^{\infty} \int_{A_n} |X| \mathrm{d}P .$$

Wegen

$$nP(A_n) \leq \int_{A_n} |X| \mathrm{d}P \leq (n+1)P(A_n)$$

(4.6 (vi)) ist zu zeigen:

$$\sum_{n=0}^{\infty} nP(A_n) = \sum_{n=1}^{\infty} P\{|X| \geq n\} \tag{1}$$

(der Wert ∞ ist möglich). Dazu betrachte für $N \in \mathbb{N}$

$$\begin{aligned}\sum_{n=0}^{N} n\Big[P\{|X| \geq n\} - P\{|X| \geq n+1\}\Big] \\ &= \sum_{n=1}^{N} \big(n-(n-1)\big)P\{|X| \geq n\} - NP\{|X| \geq N+1\} \\ &= \sum_{n=1}^{N} P\{|X| \geq n\} - NP\{|X| \geq N+1\}\,.\end{aligned}$$

Also ist

$$\sum_{n=1}^{N} nP(A_n) \leq \sum_{n=1}^{N} P\{|X| \geq n\} \leq \sum_{n=1}^{N} nP(A_n) + NP\{|X| \geq N+1\}\,.$$

Der letzte Ausdruck ist nicht größer als $\int_{\{|X| \geq N+1\}} |X| \mathrm{d}P$. Falls $\mathrm{E}\,|X| < \infty$, so folgt mit majorisierter Konvergenz, daß $\int_{|X| \geq N+1} |X| \mathrm{d}P \to 0$ für $N \to \infty$, d.h.

$$\sum_{n=0}^{\infty} nP(A_n) = \sum_{n=1}^{\infty} P\{|X| \geq n\}\,,$$

also (1), wobei beide Seiten von (1) endlich sind. Falls $\mathrm{E}\,|X| = \infty$, so folgt

$$\sum_{n=1}^{\infty} nP(A_n) = \infty$$

und damit $\sum_{n=1}^{\infty} P\{|X| \geq n\} = \infty$. □

Lemma 4.11. a) *Für* $X \in L^2(P)$ *ist* $\operatorname{Var} X = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2$.

b) *Es gilt für* $X \in L^2(P)$ *und* $\alpha, \beta \in \mathbb{R}$ *die Gleichheit* $\operatorname{Var}(\alpha X + \beta) = \alpha^2 \operatorname{Var} X$.

Beweis. a) $\mathrm{E}[(X - \mathrm{E}\,X)^2] = \mathrm{E}[X^2 - 2\,\mathrm{E}\,X \cdot X + (\mathrm{E}\,X)^2] = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2$.

b) $\mathrm{E}[(\alpha X + \beta) - \mathrm{E}(\alpha X + \beta)]^2 = \mathrm{E}[\alpha X - \alpha\,\mathrm{E}\,X]^2 = \alpha^2\,\mathrm{E}[X - \mathrm{E}\,X]^2$. □

Satz 4.12 (Ungleichung von Chebyshev). *Sei* $1 \leq p < \infty$ *und* $X \in L^p(P)$. *Dann ist*

$$P\{|X| \geq c\} \leq \frac{1}{c^p}\|X\|_{L^p}^p \qquad \text{für alle } c > 0\,.$$

Insbesondere ist für $X \in L^2(P)$

$$P\{|X - \mathrm{E}\,X| \geq c\} \leq \frac{1}{c^2} \operatorname{Var} X\,.$$

Beweis. Es gilt

$$\|X\|_{L^p}^p = \int |X|^p \mathrm{d}P \geq \int_{\{|X| \geq c\}} |X|^p \mathrm{d}P \geq c^p P\{|X| \geq c\}\,.$$

□

Definition 4.13. Seien $X, Y \in L^1(P)$ und $XY \in L^1(P)$ (etwa $X, Y \in L^2(P)$; dies genügt wegen der Hölderschen Ungleichung). Dann heißt

$$\operatorname{Cov}(X, Y) := \mathrm{E}[(X - \mathrm{E}\,X)(Y - \mathrm{E}\,Y)]$$

die Kovarianz von X und Y. Die Zufallsvariablen X und Y heißen unkorreliert, falls $\operatorname{Cov}(X, Y) = 0$. Für $\operatorname{Var} X > 0$ und $\operatorname{Var} Y > 0$ heißt

$$\rho(X, Y) := \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var} X}\sqrt{\operatorname{Var} Y}}$$

der Korrelationskoeffizient von X und Y.

Abbildung 2 zeigt die gemeinsame Dichte zweier (normalverteilter, siehe Abschnitt 5) Zufallsvariablen mit Korrelation 0. Falls der Korrelationskoeffizient positiv wird, ergibt sich eine „Verschiebung“ in Richtung der Hauptdiagonalen, wie Abbildung 3 zeigt, bei welcher $\rho = 0.8$ ist. Für negativen Korrelationskoeffizienten erhält man eine Verschiebung der Dichte von der Hauptdiagonalen weg, siehe Abbildung 4 mit $\rho = -0.8$.

Abbildung 2: Die Dichte zweier unkorrelierter normalverteilter Zufallsvariablen

Abbildung 3: Die Dichte zweier positiv korrelierter normalverteilter Zufallsvariablen

Abbildung 4: Die Dichte zweier negativ korrelierter normalverteilter Zufallsvariablen

Satz 4.14. *Seien $X_1, \ldots, X_n \in L^2(P)$. Dann ist*

$$\operatorname{Var}\Big(\sum_{i=1}^n X_i\Big) = \sum_{i=1}^n \operatorname{Var} X_i + \sum_{\substack{i,j=1\\ i\neq j}}^n \operatorname{Cov}(X_i, X_j)\,.$$

Falls $(X_1, \ldots, X_n)$ *unkorreliert sind* (*d.h. falls* $\operatorname{Cov}(X_i, X_j) = 0$ *gilt für* $i \neq j$), *so folgt*

$$\operatorname{Var} \sum_{i=1}^n X_i = \sum_{i=1}^n \operatorname{Var} X_i$$

(*Gleichheit von Bienaymé*).

Beweis. Der Beweis erfolgt induktiv, wobei nur der Schritt von n nach $n+1$ zu zeigen ist. Sei $S_n := \sum_{i=1}^n X_i$. Dann gilt

$$\begin{aligned}
\mathrm{E}(S_n + X_{n+1})^2 - {}&[\mathrm{E}(S_n + X_{n+1})]^2 \\
&= \mathrm{E}\,X_n^2 + 2\,\mathrm{E}(S_n X_{n+1}) + \mathrm{E}\,X_{n+1}^2 - (\mathrm{E}\,S_n + \mathrm{E}\,X_{n+1})^2 \\
&= \operatorname{Var} S_n + \operatorname{Var} X_{n+1} + 2\,\mathrm{E}(S_n X_{n+1}) - 2\,\mathrm{E}\,S_n \cdot \mathrm{E}\,X_{n+1} \\
&= \operatorname{Var} S_n + \operatorname{Var} X_{n+1} + 2\operatorname{Cov}(S_n, X_{n+1}) \\
&= \sum_{i=1}^n \operatorname{Var} X_i + \sum_{\substack{i,j=1\\ i\neq j}}^n \operatorname{Cov}(X_i, X_j) + 2\sum_{i=1}^n \operatorname{Cov}(X_i, X_{n+1})
\end{aligned}$$

was zu zeigen war. Bei der letzten Gleichheit wurde die Induktionsvoraussetzung verwendet. □

5. Beispiele wichtiger Wahrscheinlichkeits-Verteilungen

Im folgenden sollen einige wichtige Wahrscheinlichkeitsverteilungen vorgestellt werden. Weitere Beispiele von Verteilungen, die in der Statistik benötigt werden, werden im zugehörigen Kapitel besprochen.

a) Gleichverteilung: Die Gleichverteilung auf $[a,b]$ mit $-\infty < a < b < \infty$ ist definiert als

$$P \circ X^{-1}(B) := \frac{\lambda(B \cap [a,b])}{\lambda([a,b])},$$

wobei λ das eindimensionale Lebesgue-Maß bezeichnet (siehe Abbildungen 5 und 6).

b) Binomialverteilung: Diese ist definiert als

$$P \circ X^{-1} = B(n;p) := \sum_{k=0}^{n} \binom{k}{n} p^k (1-p)^{n-k} \delta_k \qquad \text{für } n \in \mathbb{N}, \quad 0 \leq p \leq 1\,.$$

Es gilt $\mathrm{E}\,X = np$ und $\mathrm{Var}\,X = np(1-p)$. Im Falle $n = 1$ spricht man vom Bernoulli-Experiment; hier ist $P \circ X^{-1} = (1-p)\delta_0 + p\delta_1$ (siehe Abbildungen 7 und 8).

c) Poisson-Verteilung: Hier ist $P \circ X^{-1} = \pi_\lambda|_{\mathscr{B}(\mathbb{R})}$, wobei $\pi_\lambda \colon \mathscr{P}(\mathbb{R}) \longrightarrow [0,1]$ mit Parameter $\lambda > 0$ definiert ist

$$\pi_\lambda := \sum_{k=0}^{\infty} \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \delta_k\,.$$

In diesem Falle ist X diskret verteilt und auf $\mathbb{N}_0$ konzentriert. Es gilt $\mathrm{E}\,X = \mathrm{Var}\,X = \lambda$. Für $\mathrm{E}\,X$ sieht man das folgendermaßen:

$$\begin{aligned}\mathrm{E}\,X &= \int_{\mathbb{R}} \mathrm{id}\,\mathrm{d}(P \circ X^{-1}) = \int_{\Omega} X \mathrm{d}P = \sum_{k=0}^{\infty} k \underbrace{\mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}}_{P\{X=k\}} = \mathrm{e}^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \\ &= \mathrm{e}^{-\lambda} \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda \mathrm{e}^{-\lambda} \mathrm{e}^{\lambda} = \lambda\,;\end{aligned}$$

für $\mathrm{Var}\,X$ ist eine ähnliche Rechnung durchzuführen. Die Poisson-Verteilung ist in den Abbildungen 9 und 10 dargestellt.

d) Exponentialverteilung mit Parameter $\lambda > 0$: Hier ist $P \circ X^{-1} = f_X(t)\mathrm{d}t$ mit $f_X(t) = 1_{\mathbb{R}_+}(t)\lambda \mathrm{e}^{-\lambda t}$ (d.h. X ist stetig verteilt). Für die Verteilungsfunktion erhält man

$$F_X(t) = \begin{cases} 0 & \text{falls } t < 0, \\ 1 - e^{-\lambda t} & \text{falls } t \geq 0. \end{cases}$$

Es gilt $\mathrm{E}\,X = \frac{1}{\lambda}$ und $\operatorname{Var} X = \frac{1}{\lambda^2}$, wie folgende Rechnung zeigt:

$$\begin{aligned}
\mathrm{E}\,X &= \int_{\mathbb{R}} t f_X(t)dt = \int_0^\infty t\lambda \mathrm{e}^{-\lambda t} dt = \frac{1}{\lambda} \\
\mathrm{E}\,X^2 &= \int_{\mathbb{R}} t^2 f_X(t)dt = 2\int_0^\infty t\mathrm{e}^{-\lambda t} dt = \frac{2}{\lambda^2} \\
\operatorname{Var} X = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2 &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}
\end{aligned}$$

(siehe Abbildungen 11 und 12).

e) Normalverteilung zum Mittelwert $\mu \in \mathbb{R}$ mit Streuung $\sigma > 0$: Diese vielleicht berühmteste Verteilung ist definiert als

$$P \circ X^{-1} = N(\mu, \sigma^2) := \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}} \mathrm{d}t\,.$$

Es gilt $\mathrm{E}\,X = \mu$ und $\operatorname{Var} X = \sigma^2$, wie hier nicht nachgerechnet werden soll (vgl. Übungsaufgabe 16). Für $\mu = 0$ und $\sigma = 1$ ist

$$F_X(t) = \Phi(t) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t \mathrm{e}^{-\frac{\tau^2}{2}} \mathrm{d}\tau$$

(Gauß Verteilung). Wegen $N(\mu, \sigma^2)]s, t] = N(0, 1)]\frac{s-\mu}{\sigma}, \frac{t-\mu}{\sigma}]$ (Transformationssatz) gilt: X ist genau dann $N(0, 1)$-verteilt, wenn $\sigma X + \mu$ gemäß $N(\mu, \sigma^2)$-verteilt ist.

Exemplarisch soll gezeigt werden, daß

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-t^2/2} \mathrm{d}t = 1$$

gilt. Dazu schreibt man

$$\begin{aligned}
\left(\int_{\mathbb{R}} \mathrm{e}^{-t^2/2} \mathrm{d}t \right)^2 &= \int_{\mathbb{R}} \mathrm{e}^{-x^2/2} \mathrm{d}x \int_{\mathbb{R}} \mathrm{e}^{-y^2/2} \mathrm{d}y \\
&= \int_{\mathbb{R}^2} \mathrm{e}^{-\frac{x^2+y^2}{2}} \mathrm{d}(x, y)\,.
\end{aligned}$$

Das letzte Integral wird mit Hilfe des Transformationslemmas ausgerechnet. Betrachte dazu

$$J\colon (0, \infty) \times [0, 2\pi) \longrightarrow \mathbb{R}^2 \backslash \{0\}, \quad (r, \varphi) \longmapsto (r\cos\varphi, r\sin\varphi) = (x, y)\,.$$

Es gilt

$$DJ(r, \varphi) = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \varphi} \end{pmatrix} = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}$$

und damit ist

$$\left(\int_{\mathbb{R}} \mathrm{e}^{-t^2/2}\mathrm{d}t\right)^2 = \int_{(0,\infty)\times[0,2\pi)} \mathrm{e}^{-\frac{r^2}{2}} \underbrace{|\det DJ(r,\varphi)|}_{r} d(r,\varphi)$$
$$= \int_{[0,2\pi)} 1\mathrm{d}\varphi \int_{(0,\infty)} r\mathrm{e}^{-\frac{r^2}{2}}\mathrm{d}r = 2\pi\Big[-\mathrm{e}^{-r^2/2}\Big]_0^\infty = 2\pi\,.$$

Die Standard-Normalverteilung ist in den Abbildungen 13 und 14 dargestellt.

f) Cauchy-Verteilung mit Parameter $\alpha > 0$: Diese ist definiert durch

$$P \circ X^{-1} = \frac{\alpha}{\pi}\frac{1}{t^2+\alpha^2}dt$$

Achtung: $\mathrm{E}\,X$ existiert nicht, da $\mathrm{E}\,|X| = \infty$ gilt

(siehe Abbildungen 15 und 16).

Satz 5.1 (Poisson-Näherung für die Binomialverteilung). *Sei* $p_n \longrightarrow 0$ *und* $np_n \longrightarrow \lambda > 0$. *Dann gilt:* $\lim\limits_{n\to\infty} B(n;p_n)\{k\} = \pi_\lambda\{k\}$ *für alle* $k \in \mathbb{N}_0$.

Beweis. Sei $\lambda_n := np_n$. Dann gilt:

$$B(n;p_n)\{k\} = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{n!}{k!(n-k)!}\frac{\lambda_n^k}{n^k}\left(1-\frac{\lambda_n}{n}\right)^{n-k}$$
$$= \underbrace{\frac{n}{n}\cdot\frac{n-1}{n}\cdot\ldots\cdot\frac{n-k+1}{n}}_{\to 1}\cdot\underbrace{\frac{\lambda_n^k}{k!}}_{\to\frac{\lambda^k}{k!}}\cdot\underbrace{\left(1-\frac{\lambda_n}{n}\right)^{-k}}_{\to 1}\cdot\underbrace{\left(1-\frac{\lambda_n}{n}\right)^{n}}_{\to \mathrm{e}^{-\lambda}}$$
$$\longrightarrow \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$$

□

Beispiel 5.2 (Wartezeiten). Sei $\lambda > 0$ und $(X_t)_{t\geq 0}$ eine Familie von Zufallsvariablen $X_t\colon \Omega \longrightarrow \mathbb{N}_0$ mit $X_0 = 0$ und $P\circ X_t^{-1} = \pi_{\lambda t}\quad (t>0)$. Es sei $t \longmapsto X_t(\omega)$ r. c. für alle $\omega \in \Omega$. (Mit einer zusätzlichen Bedingung über Unabhängigkeit und Stationarität definiert dies einen Poisson-Prozeß, der etwa in der Vorlesung über Stochastische Prozesse behandelt wird, vgl. z.B. [7].) X_t beschreibt etwa die Anzahl der Emissionen von α-Teilchen eines radioaktiven Präparats. Man interessiert sich dabei auch für die Wartezeit, bis der erste Impuls gemessen wird. Diese wird gegeben durch

$$Z\colon \Omega \longrightarrow \mathbb{R}_+ \quad \text{mit } Z(\omega) := \inf\{t \geq 0 : X_t(\omega) \geq 1\}\,.$$

Es ist $Z(\omega) > t$ genau dann, wenn $X_t(\omega) = 0$. Also ist $P\{Z > t\} = P\{X_t = 0\} = P \circ X_t^{-1}\{0\} = \pi_{\lambda t}\{0\} = \mathrm{e}^{-\lambda t} \quad (t > 0)$. Somit ist

$$F_Z(t) = \begin{cases} 0, & t < 0, \\ 1 - \mathrm{e}^{-\lambda t}, & t \geq 0, \end{cases}$$

d.h. Z ist exponentialverteilt mit Parameter λ.

Lemma 5.3. *Für $X \in L^2(P)$ sind folgende Aussagen äquivalent:*

(i) $\operatorname{Var} X = 0$,

(ii) $P \circ X^{-1} = \delta_\mu$ *mit* $\mu = \mathrm{E}\, X$,

(iii) $P \circ X^{-1}(\mathscr{B}(\mathbb{R})) = \{0, 1\}$.

Beweis: siehe Übungsaufgabe 17.

Auf den folgenden Seiten werden die Graphen der oben besprochenen Wahrscheinlichkeitsverteilungen wiedergegeben.

6. Konvergenzbegriffe

Bereits in der Analysis wird klar, wie entscheidend es ist, genau zwischen verschiedenen Konvergenzbegriffen, wie etwa punktweiser und gleichmäßiger Konvergenz von Funktionen, zu unterscheiden. Falls es um die Konvergenz von Zufallsgrößen geht, ist dies vielleicht sogar noch wichtiger. Daher sollen in diesem Abschnitt verschiedene Konvergenzarten definiert und analysiert werden; die Aussagen der späteren Abschnitte werden stets auf diese Begriffe zurückgreifen.

Zunächst einmal sei kurz wiederholt, welche Konvergenzarten bereits aus der klassischen Analysis bekannt sind.

(i) Der einfachste Konvergenzbegriff ist wohl der für reelle (oder komplexe) Zahlen. Eine Folge $(z_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ konvergiert genau dann gegen $z \in \mathbb{R}$, falls für alle $\varepsilon > 0$ ein $N \in \mathbb{N}$ existiert mit $|z_n - z| < \varepsilon$ für alle $n \geq N$.

(ii) Etwas komplizierter wird es, falls Funktionen $X_n, X : \Omega \longrightarrow \mathbb{R}$ auf einer Menge Ω gegeben sind. Hier gibt es verschiedene Konvergenzbegriffe:

(a) X_n konvergiert punktweise gegen X, falls $X_n(\omega) \longrightarrow X(\omega)$ für alle $x \in \Omega$ gilt, d.h. falls

$$\forall\,\omega \in \Omega\;\forall\,\varepsilon > 0\;\exists\,N = N(\omega) \in \mathbb{N}\;\forall\,n \geq N : |X_n(\omega) - X(\omega)| < \varepsilon\,.$$

(b) X_n konvergiert gleichmäßig gegen X, falls $||X_n - X||_\infty \longrightarrow 0$ gilt, wobei

$$||X_n - X||_\infty := \sup_{\omega\in\Omega} |X_n(\omega) - X(\omega)|\,,$$

d.h. falls gilt:

$$\forall\,\varepsilon > 0\;\exists\,N \in \mathbb{N}\;\forall\,n \geq N\;\forall\,\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\,.$$

(c) Es gilt $X_n \longrightarrow X$ in $L^p(P)$, falls $||X_n - X||_{L^p} \longrightarrow 0$ (wobei $(\Omega, \mathscr{A}, P)$ Maßraum sei). Dabei ist

$$||X_n - X||_{L^p} := \Big[\int_\Omega |X_n(\omega) - X(\omega)|^p P(d\omega)\Big]^{1/p} = \big[\,\mathrm{E}(|X_n - X|^p)\big]^{1/p}\,.$$

(d) In der Situation von (c) konvergiert X_n P-fast sicher (P-f.s.) gegen X, falls $P\{X_n \not\to X\} = 0$ gilt. Falls P ein W-Maß ist, ist dies äquivalent zu $P\{X_n \to X\} = 1$.

(iii) Ein etwas allgemeinerer Konvergenzbegriff (der in (c) bereits verwendet wurde) ist die Konvergenz im normierten Raum $(E, ||\cdot||_E)$. Hier gilt für $X_n, X \in E$ die Konvergenz $X_n \longrightarrow X$ in E genau dann, wenn $||X_n - X||_E \longrightarrow 0$.

(iv) Falls der Raum nicht normiert ist, sondern nur eine Metrik besitzt, kann man die Konvergenz analog definieren. Sei (E, d_E) ein metrischer Raum. Dann gilt für $X_n, X \in E$ die Konvergenz $X_n \longrightarrow X$ in E, falls $d_E(X_n, X) \longrightarrow 0$.

(v) Falls man nicht einmal eine Metrik zur Verfügung hat, der Raum aber eine topologische Struktur besitzt, so kann man die Konvergenz ebenfalls definieren. Sei also (E, τ_E) ein topologischer Raum. Dann gilt für $X_n, X \in E$ die Konvergenz $X_n \longrightarrow X$ in E (in τ_E), falls für jede Umgebung U von X ein $N \in \mathbb{N}$ existiert, so daß für alle $n \geq N$ gilt: $X_n \in U$.

Sei nun $(\Omega, \mathscr{A}, P)$ ein W-Raum. Von obigen Konvergenzbegriffen sind die beiden $X_n \longrightarrow X$ P-f.s. und $X_n \longrightarrow X$ in $L^p(P)$ wichtig. Das Ziel dieses Abschnitts ist es, weitere (schwächere) Konvergenzbegriffe zu untersuchen, welche etwa bei der Formulierung des zentralen Grenzwertsatzes verwendet werden.

Beachte zu $X_n \longrightarrow X$ P-f.s., daß

$$\begin{aligned}\{X_n \not\to X\} &= \{\omega : \limsup_{n\to\infty} |X_n(\omega) - X(\omega)| > 0\} = \\ &= \bigcup_{m\in\mathbb{N}} \left\{\omega : \limsup_{n\to\infty} |X_n(\omega) - X(\omega)| \geq \frac{1}{m}\right\} \in \mathscr{A}\end{aligned}$$

gilt. Die Folge X_n konvergiert genau dann P-f.s. gegen X, falls für alle $\varepsilon > 0$ gilt $P\{\limsup_{n\to\infty} |X_n - X| > \varepsilon\} = 0$. Dies ist äquivalent dazu, daß für alle $\varepsilon > 0$ gilt

$$P(\limsup_{n\to\infty}\{|X_n - X| > \varepsilon\}) = 0\,,$$

wobei

$$\limsup A_n := \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} A_k\,.$$

Um die Äquivalenz zu sehen, beachte man, daß

$$\begin{aligned}\{\limsup |X_n - X| > \varepsilon\} &= \{\omega \in \Omega : \forall n \in \mathbb{N}\, \exists\, k \geq n : |X_k(\omega) - X(\omega)| > \varepsilon\} \\ &= \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} \{|X_k - X| > \varepsilon\} \\ &= \limsup_{n\to\infty}\{|X_n - X| > \varepsilon\}\,.\end{aligned}$$

Definition 6.1. X_n konvergiert stochastisch gegen X, falls

$$\lim_{n\to\infty} P\{|X_n - X| > \varepsilon\} = 0 \qquad \text{für alle } \varepsilon > 0$$

gilt. Man sagt auch $X_n \longrightarrow X$ in Wahrscheinlichkeit.

Lemma 6.2. a) *Falls $X_n \longrightarrow X$ P-f.s., so folgt $X_n \longrightarrow X$ stochastisch.*

b) *Falls $X_n \longrightarrow X$ in $L^p(P)$, so gilt $X_n \longrightarrow X$ stochastisch.*

Beweis. a) X_n konvergiert gegen X genau dann P-f.s., falls

$$P\Big(\bigcap_{n\in\mathbb{N}} \{\omega : \text{ Es gibt ein } k \geq n \text{ mit } |X_k(\omega) - X(\omega)| > \varepsilon\}\Big) = 0$$

für alle $\varepsilon > 0$ gilt. Dies ist äquivalent zu

$$\lim_{n\to\infty} P\{\omega : \text{ Es gibt ein } k \geq n \text{ mit } |X_k(\omega) - X(\omega)| > \varepsilon\}\Big) = 0 \qquad \text{für alle } \varepsilon > 0\,,$$

woraus für alle $\varepsilon > 0$

$$\lim_{n\to\infty} P\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0$$

und damit die stochastische Konvergenz folgt.

b) Nach Chebyshev gilt

$$P\{|X_n - X| > \varepsilon\} \leq \frac{1}{\varepsilon^p} ||X_n - X||^p_{L^p(P)} \longrightarrow 0\,.$$

□

Satz 6.3 (Teilfolgen-Teilfolgen-Satz). *X_n konvergiert genau dann stochastisch gegen X, falls jede Teilfolge $(X_{n_j})_{j\in\mathbb{N}}$ eine Teilfolge besitzt, die P-f.s. gegen X konvergiert.*

Beweis. „$\Longrightarrow$“. Sei $Y_j := X_{n_j}$. Wähle $j_1 < j_2 < \dots$ mit $P(A_k) < \frac{1}{2^k}$, wobei $A_k := \{|Y_{j_k} - X| > \frac{1}{k}\}$.

Sei $\varepsilon > 0$. Dann gilt für $N > \frac{1}{\varepsilon}$

$$\begin{aligned} P\left(\limsup_{k\to\infty}\{|Y_{j_k} - X| > \varepsilon\}\right) &\leq P\left(\bigcap_{n\geq N}\bigcup_{k\geq n}\{|Y_{j_k} - X| > \varepsilon\}\right) \leq \\ &\leq P\left(\bigcap_{n\geq N}\bigcup_{k\geq n} A_k\right) \leq P\left(\bigcup_{k\geq n} A_k\right) \leq \\ &\leq \sum_{k=N}^{\infty} P(A_k) \leq \sum_{k=N}^{\infty}\frac{1}{2^k} = \frac{1}{2^{N-1}}\,. \end{aligned}$$

Somit ist $P\Big(\limsup_{k\to\infty}\{|Y_{j_k} - X| > \varepsilon\}\Big) = 0$, d.h. $Y_{j_k} \longrightarrow X$ P-f.s.

„$\Longleftarrow$“. Angenommen $X_n \not\to X$ stochastisch . Dann existiert ein $\varepsilon > 0$ mit $P\{|X_n - X| > \varepsilon\} \not\longrightarrow 0$, d.h. es existiert eine Folge $n_1 < n_2 < \dots$ und ein $h > 0$ mit

$$P\{|X_n - X| > \varepsilon\} \geq h \qquad \text{für alle } k \in \mathbb{N}\,. \tag{$*$}$$

Nach Voraussetzung besitzt $(X_{n_k})_k$ eine Teilfolge $(X_{n_{k_j}})_j$, die P-f.s. konvergiert. Wegen Lemma 6.2 a) gilt $X_{n_{k_j}} \longrightarrow X$ stochastisch , d.h. $P\{|X_{n_{k_j}} - X| > \varepsilon\} \xrightarrow{j\to\infty} 0$ für alle $\varepsilon > 0$ im Widerspruch zu $(*)$. □

Korollar 6.4. a) *Falls* $X_n \longrightarrow X$ *stochastisch , so existiert eine Teilfolge* $\{X_{n_k}\} \subset \{X_n\}$ *mit* $X_{n_k} \longrightarrow X$ *P-f.s.*

b) *Falls* $X_n \longrightarrow X$ *in* $L^p(P)$ *mit* $(1 \leq p < \infty)$*, so existiert eine Teilfolge, die P-f.s. gegen* X *konvergiert.*

c) *Aus* $X_n \longrightarrow X$ *und* $Y_n \longrightarrow Y$ *stochastisch folgt* $X_n \pm Y_n \longrightarrow X \pm Y$ *stochastisch und* $X_n \cdot Y_n \longrightarrow X \cdot Y$ *stochastisch .*

Bemerkung 6.5. a) Stochastische Konvergenz ist eine Konvergenz im Sinne der Topologie, d.h. es existiert eine Topologie τ auf $\mathscr{L} := \{X : \Omega \to \mathbb{C} \,|\, X\ \mathscr{A}\text{-meßbar}\}$, so daß $X_n \overset{n\to\infty}{\longrightarrow} X$ in τ genau dann, wenn $X_n \longrightarrow X$ stochastisch .

Man kann als Umgebungsbasis von $X \in \mathscr{L}$ alle Mengen der Form

$$\Big\{Y \in \mathscr{L} : P\Big\{|X - Y| > \frac{1}{m}\Big\} < \frac{1}{n}\Big\} \text{ mit } m, n \in \mathbb{N}$$

wählen. Die Topologie kann sogar durch eine (Halb-)Metrik beschrieben werden, siehe Korollar 6.7.

b) Die P-f.s. Konvergenz läßt sich nicht durch eine Topologie schreiben. Denn angenommen es existiert eine Topologie $\hat{\tau}$ mit $X_n \longrightarrow X$ P-f.s. $\Longleftrightarrow X_n \longrightarrow$ in τ.
Sei $X_n \longrightarrow X$ stochastisch , aber nicht P-f.s. ($X_n \not\longrightarrow X$ P-f.s.) (siehe Übung 21). Dann existiert eine Umgebung U von X und eine Teilfolge $\{X_{n_k}\}_k$ mit $X_{n_k} \notin U$ für alle k. Aber $\{X_{n_k}\}_k$ besitzt nach Satz 6.3 eine Teilfolge, die P-f.s. (und damit in $\hat{\tau}$) gegen X konvergiert. Somit müssen die Elemente dieser Teilfolge (für große Indizes) in U liegen, was aber ein Widerspruch zu $X_{n_k} \notin U$ ist.

Satz 6.6 (dominierte Konvergenz). *Sei* $1 \leq p < \infty$ *und* $X_n, X, Y \in L^p(P)$ *mit* $|X_n| \leq Y$ *P-f.s. für alle* n*. Falls* $X_n \longrightarrow X$ *stochastisch , so gilt* $X_n \longrightarrow X$ *in* $L^p(P)$*.*

Beweis. Mit X_n ist auch $X_n - X$ dominiert (durch $Y + |X|$)

$$\begin{aligned} \mathrm{E}(|X_n - X|^p) &= \int_{\{|X_n - X| < \varepsilon\}} |X_n - X|^p dP + \int_{\{|X_n - X| \geq \varepsilon\}} |X_n - X|^p dP \leq \\ &\leq \varepsilon^p + \int_{\{|X_n - X| \geq \varepsilon\}} (Y + |X|)^p dP \end{aligned}$$

Wegen $P\{|X_n - X| \geq \varepsilon\} \longrightarrow 0$, da $X_n \longrightarrow X$ stochastisch , folgt (mit Übung 8) $\lim\limits_{n\to\infty} \mathrm{E}(|X_n - X|^p) \leq \varepsilon^p$. Da $\varepsilon > 0$ beliebig war, gilt: $\lim\limits_{n\to\infty} \mathrm{E}(|X_n - X|^p) = 0$, d.h. $X_n \longrightarrow X$ in $L^p(P)$. $\square$

Korollar 6.7. (*vgl. auch Übung* 9). *Es gilt* $X_n \longrightarrow X$ *stochastisch genau dann, wenn*

$$\mathrm{E}\left(\frac{|X_n - X|}{1 + |X_n - X|}\right) \longrightarrow 0\,.$$

Beweis. o.E. sei $X = 0$ (beachte Korollar 6.4 c). Da $\frac{|X_n|}{1+|X_n|} \leq 1$ und $1 \in L^p(P)$ gilt, liefert Satz 6.6: $\frac{|X_n|}{1+|X_n|} \longrightarrow 0$ in $L^p(P)$ genau dann, wenn $\frac{|X_n|}{1+|X_n|} \longrightarrow 0$ stochastisch . Wegen

$$P\{|X_n| > \varepsilon\} = P\Big\{\frac{|X_n|}{1+|X_n|} > \frac{\varepsilon}{1+\varepsilon}\Big\}$$

ist dies äquivalent zu $X_n \longrightarrow 0$ stochastisch . □

Nun kommen noch zwei Konvergenzbegriffe hinzu, die beim Beweis des zentralen Grenzwertsatzes eine wichtige Rolle spielen:

Definition 6.8. a) $X_n \longrightarrow X$ in Verteilung $:\Longleftrightarrow F_{X_n}(t) \longrightarrow F_X(t)$ für alle t, für welche F_X stetig ist an der Stelle t.

b) Sei $\mathscr{K} \subset C(\mathbb{R})$ eine Familie von stetigen Funktionen. Dann konvergiert $X_n \longrightarrow X$ schwach bzgl. $\mathscr{K}$ genau dann, wenn

$$\int_{\mathbb{R}} f \mathrm{d}P \circ X_n^{-1} \longrightarrow \int_{\mathbb{R}} f \mathrm{d}P \circ X^{-1} \quad \text{für alle } f \in \mathscr{K}\,.$$

Als Familie $\mathscr{K}$ wählt man häufig eine der folgenden Funktionenklassen:

$$\begin{aligned} C_c(\mathbb{R}) &:= \{f : \mathbb{R} \to \mathbb{C} \mid f \text{stetig, supp} f \text{ kompakt}\}\,, \\ C_b(\mathbb{R}) &:= \{f : \mathbb{R} \to \mathbb{C} \mid f \text{ stetig, beschränkt}\}\,, \\ \mathscr{D}(\mathbb{R}) &:= C_c^\infty(\mathbb{R}) := C_c(\mathbb{R}) \cap C^\infty(\mathbb{R})\,. \end{aligned}$$

Lemma 6.9. *Es gilt $X_n \to X$ in Verteilung genau dann, wenn eine dichte Teilmenge $D \subset \mathbb{R}$ existiert mit $F_{X_n}(t) \longrightarrow F_X(t)$ für alle $t \in D$.*

Beweis. „$\Longrightarrow$“. $D := \{t \in \mathbb{R} \mid F_X$ stetig an der Stelle $t\}$ ist dicht wegen $\mathbb{R}\backslash D$ abzählbar nach 4.2 a) und 2.5.

„$\Longleftarrow$“. Sei F_X stetig an der Stelle t. Wähle $\varepsilon > 0$ und $t_1, t_2 \in D$ mit

$$t - \varepsilon < t_1 < t < t_2 < t + \varepsilon\,.$$

Wegen $F_{X_n}(t_1) \longrightarrow F_X(t_1)$ und $F_{X_n}(t_2) \longrightarrow F_X(t_2)$ gilt für ein $n_0 \in \mathbb{N}$:

$$|F_{X_n}(t_i) - F_X(t_i)| \leq \varepsilon \quad \text{für alle } n \geq n_0,\ i = 1, 2 \tag{$*$}$$

und damit

$$\begin{aligned} F_X(t-\varepsilon) \leq F_x(t_1) &\overset{(*)}{\leq} F_X(t_1) + \varepsilon \leq F_{X_n}(t) + \varepsilon \leq F_{X_n}(t_2) + \varepsilon \\ &\overset{(*)}{\leq} F_X(t_2) + 2\varepsilon \leq F_X(t+\varepsilon) + 2\varepsilon \end{aligned}$$

Somit gilt $F_X(t-\varepsilon) - \varepsilon \leq F_{X_n}(t) \leq F_X(t+\varepsilon) + \varepsilon$. Für $\varepsilon \searrow 0$ konvergiert die linke und die rechte Seite der letzten Ungleichung gegen $F_X(t)$; also konvergiert auch der Ausdruck in der Mitte gegen $F_X(t)$. □

Ein großer Vorteil der Konvergenz in Verteilung liegt in der Folgenkompaktheit, welche im nächsten Satz zum Ausdruck kommt:

Satz 6.10. *Sei $\{\mu_n\}$ eine Folge von Maßen auf $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ mit $\mu_n(\mathbb{R}) \leq 1$. Dann existiert ein Maß μ auf $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ mit $\mu(\mathbb{R}) \leq 1$ und eine Teilfolge $\{\mu_{n_k}\}$ mit*

$$\mu_{n_k}(-\infty, t] \stackrel{k\to\infty}{\longrightarrow} \mu(-\infty, t] \quad \textit{für alle } t \in \mathbb{R} \ \textit{ mit } \ \mu\{t\} = 0\,.$$

Beweis. Definiere $F_n(t) := \mu_n(-\infty, t]$ für $n \in \mathbb{N}, t \in \mathbb{R}$. Sei D abzählbare dichte Teilmenge von $\mathbb{R}$, etwa $D = \mathbb{Q}$. Schreibe $D = \{r_1, r_2, \dots\}$.

Die Folge $\{F_n(r_1)\}_{n\in\mathbb{N}} \subset \mathbb{R}$ ist beschränkt (da in [0,1]); also existiert eine konvergente Teilfolge $\{F_{1k}(r_1)\}_{k\in\mathbb{N}}$ mit $\lim\limits_{k\to\infty} F_{1k}(r_1) =: l_1 \in [0, 1]$.

Die Folge $\{F_{1k}(r_2)\}_{k\in\mathbb{N}} \subset [0, 1]$ besitzt wieder eine konvergente Teilfolge $\{F_{2k}(r_2)\}_{k\in\mathbb{N}}$ mit $F_{2k}(r_2) \longrightarrow l_2 \in [0, 1]$.

Dieser Prozeß wird fortgesetzt; man erhält im j-ten Schritt die Teilfolge $\{F_{jk}\}_k$. Betrachte nun die Diagonalfolge $\{F_{kk}\}_{k\in\mathbb{N}}$. Sei $r_j \in D$ beliebig. Dann gilt $F_{kk}(r_j) \longrightarrow l_j$, da für $k \geq j$ die Diagonalfolge $\{F_{kk}\}$ eine Teilfolge von $\{F_{jk}\}_k$ ist. Insgesamt haben wir also eine Teilfolge $\{F_{n_k}\}_{k\in\mathbb{N}} := \{F_{kk}\}$ von $\{F_n\}$ und eine Funktion $G : D \longrightarrow \mathbb{R};\ G(r_j) := l_j$ mit

$$F_{n_k}(r) \longrightarrow G(r) \quad \text{für alle } r \in D\,.$$

Die Funktion G ist monoton wachsend wegen

$$\begin{array}{ccccc} F_{n_k}(r) & \leq & F_{n_k}(r') & & (r < r') \\ \big\downarrow & & \big\downarrow & & \\ G(r) & \leq & G(r') & & \end{array}$$

Definiere $F(t) := \inf\{G(r) : r > t, r \in D\}$ für alle $t \in \mathbb{R}$. Dann ist F monoton wachsend, r.c. (vgl. auch Maßtheorie-Vorlesung) und definiert durch $\mu(-\infty, t] = F(t)$ ein eindeutiges Maß μ auf $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ mit $\mu(\mathbb{R}) \leq 1$.
Genauso wie im Beweis von Lemma 6.9, Teil „$\Leftarrow$", sieht man, daß $F_{n_k}(t) \longrightarrow F(t)$ für alle t mit F stetig an der Stelle t. □

Der folgende Satz zeigt den Zusammenhang zwischen Konvergenz in Verteilung und schwacher Konvergenz:

Satz 6.11. *Seien $\{X_n\}, X$ Zufallsvariable. Dann sind äquivalent:*

(i) $X_n \longrightarrow X$ *in Verteilung.*

(ii) $X_n \longrightarrow X$ *schwach bzgl.* $\mathscr{D}(\mathbb{R})$.

(iii) $X_n \longrightarrow X$ *schwach bzgl.* $C_b(\mathbb{R})$.

(iv) $X_n \longrightarrow X$ *schwach bzgl.* $\mathscr{K} := \{t \longmapsto \mathrm{e}^{ixt} : x \in \mathbb{R}\}$.

Der Beweis verwendet folgende Approximationsaussage:

Lemma 6.12 (Approximationslemma). *Sei $I \subset \mathbb{R}$ ein kompaktes Intervall, $f\colon I \to \mathbb{R}$ stetig, $D \subset \mathbb{R}$ dicht und $\varepsilon > 0$. Dann existiert eine Treppenfunktion $f_\varepsilon\colon I \to \mathbb{R}$ der Form $f_\varepsilon = \sum_{j=1}^m c_j 1_{(a_j,a_{j+1}]}$ mit $a_1 < \cdots < a_{m+1}$, $a_j \in D$ und*

$$\sup_{t\in I} |f(t) - f_\varepsilon(t)| < \varepsilon\,.$$

Beweis von Lemma 6.12. Da I kompakt ist, ist f gleichmäßig stetig. Zu gegebenem $\varepsilon > 0$ wähle ein $\delta > 0$ so, daß für alle x, x' mit $|x - x'| < \delta$ gilt $|f(x) - f(x')| < \varepsilon$. Nun wähle $a_j \in D$, $a_1 < \cdots < a_{m+1}$ so, daß $I = [a_1, a_{m+1}]$ und $|a_{j+1} - a_j| < \delta$ gilt. Wählt man $x_j \in (a_j, a_{j+1})$, so gilt die Behauptung mit $c_j := f(x_j)$. □

Beweis von Satz 6.11. Wir setzen $\mu := P \circ X^{-1}$ und $\mu_n := P \circ X_n^{-1}$.

(i)$\Longrightarrow$(ii). Für $D = \{t \in \mathbb{R} : \mu\{t\} = 0\}$ gilt

$$\mu_n(a,b] \to \mu(a,b] \qquad \text{für } a < b,\ a, b \in D\,.$$

Damit gilt für jede Treppenfunktion $f = \sum_{j=1}^m c_j 1_{(a_j,a_{j+1}]}$ mit $a_j \in D$:

$$\int f \mathrm{d}\mu_n = \sum_{j=1}^m c_j \mu_n(a_j, a_{j+1}] \to \int f \mathrm{d}\mu \quad (n \to \infty)\,.$$

Sei nun $f \in \mathscr{D}(\mathbb{R})$ und $\varepsilon > 0$. Wähle f_ε wie im Approximationslemma. Dann gilt

$$\Big|\int f\mathrm{d}\mu_n - \int f\mathrm{d}\mu\Big| \le \int |f - f_\varepsilon|\mathrm{d}\mu + \Big|\int f_\varepsilon\mathrm{d}\mu_n - \int f_\varepsilon \mathrm{d}\mu\Big| + \int |f - f_\varepsilon|\mathrm{d}\mu\,,$$

wobei das erste und das letzte Integral auf der rechten Seite nicht größer als ε sind und der mittlere Ausdruck für $n \to \infty$ gegen 0 geht. Damit gilt

$$\limsup_{n\to\infty} \Big|\int f\mathrm{d}\mu_n - \int f\mathrm{d}\mu\Big| \le 2\varepsilon\,.$$

Da ε beliebig war, folgt $\int f\mathrm{d}\mu_n \to \int f\mathrm{d}\mu$.

(ii)$\Longrightarrow$(iii). Sei $[a,b] \subset \mathbb{R}$ mit $\mu([a,b]^c) < \varepsilon$ und $u \in \mathscr{D}(\mathbb{R})$ mit $0 \le u \le 1$ und $u \equiv 1$ in $[a,b]$. Dann gilt

$$\int (1 - u)\mathrm{d}\mu \le \mu([a,b]^c) < \varepsilon\,.$$

Wegen $\int u \mathrm{d}\mu_n \to \int u \mathrm{d}\mu$ folgt für hinreichend großes n, daß $\int (1-u)\mathrm{d}\mu_n < \varepsilon$. Sei nun $f \in C_b(\mathbb{R})$. Dann ist

$$\Big|\int f \mathrm{d}\mu_n - \int f \mathrm{d}\mu\Big| \le \int |f|(1-u)\mathrm{d}\mu_n + \Big|\int fu \mathrm{d}\mu_n - \int fu \mathrm{d}\mu\Big| + \int |f|(1-u)\mathrm{d}\mu\,.$$

Dabei sind das erste und letzte Integral nicht größer als $C\varepsilon$ für eine geeignete Konstante C, und der mittlere Ausdruck konvergiert für $n \to \infty$ gegen 0. Genauso wie im Schritt (i)$\Longrightarrow$(ii) folgt daraus $\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu$.

(iii)$\Longrightarrow$(i). Sei F_X stetig an der Stelle t. Wähle $f_1, f_2 \in C_b(\mathbb{R})$ mit

$$1_{(-\infty,t-\varepsilon]} \le f_1 \le 1_{(-\infty,t]} \le f_2 \le 1_{(-\infty,t+\varepsilon]}\,.$$

Dann gilt

$$\int f_1 \mathrm{d}\mu_n \le F_{X_n}(t) \le \int f_2 \mathrm{d}\mu_n\,.$$

Für $n \to \infty$ konvergiert das Integral auf der linken Seite gegen $\int f_1 \mathrm{d}\mu$ und das Integral auf der rechten Seite gegen $\int f \mathrm{d}\mu_2$. Somit erhält man

$$\int f_1 \mathrm{d}\mu \le \liminf_n F_{X_n}(t) \le \limsup_n F_{X_n}(t) \le \int f_2 \mathrm{d}\mu\,.$$

Andererseits gilt $\int f_1 \mathrm{d}\mu \le F_X(t) \le \int f_2 \mathrm{d}\mu$ und

$$\int f_2 \mathrm{d}\mu - \int f_1 \mathrm{d}\mu < \mu(t-\varepsilon, t+\varepsilon)\,.$$

Da F_X an der Stelle t stetig ist, konvergiert $\mu(t-\varepsilon, t+\varepsilon)$ für $\varepsilon \to 0$ gegen 0, und daher gilt $F_{X_n}(t) \to F_X(t)$.

(iii)$\Longrightarrow$(iv) ist trivial.

(iv)$\Longrightarrow$(ii). Sei $f \in \mathscr{D}(\mathbb{R})$. Dann existiert ein $g \in \mathscr{S}(\mathbb{R}) \subset L^1(\mathbb{R})$ mit $f(x) = \int \mathrm{e}^{\mathrm{i}xt} g(t)\mathrm{d}t$; hierbei bezeichnet $\mathscr{S}(\mathbb{R})$ den Schwartz-Raum. Somit gilt

$$\begin{aligned}\int f \mathrm{d}\mu_n &= \int \Big[\int \mathrm{e}^{\mathrm{i}xt} g(t)\mathrm{d}t\Big]\mu_n(\mathrm{d}x)\\ &= \int g(t)\Big[\int \mathrm{e}^{\mathrm{i}xt}\mu_n(\mathrm{d}x)\Big]\mathrm{d}t\\ &\to \int g(t)\Big[\int \mathrm{e}^{\mathrm{i}xt}\mu(\mathrm{d}x)\Big]\mathrm{d}t = \int f \mathrm{d}\mu\,.\end{aligned}$$

Dabei wurden der Satz von Fubini und der Satz über majorisierte Konvergenz verwendet. Man beachte dazu, daß $g \in L^1$ und $|\int \mathrm{e}^{\mathrm{i}xt}\mu_n(\mathrm{d}x)| \le 1$ gilt. □

Korollar 6.13. *Falls $X_n \longrightarrow X$ stochastisch, folgt $X_n \longrightarrow X$ in Verteilung.*

Beweis. Sei $f \in \mathscr{D}(\mathbb{R})$ und $\varepsilon > 0$. Da f gleichmäßig stetig ist, existiert ein $\delta > 0$ mit $|f(x) - f(x')| < \varepsilon$ für alle $|x - x'| < \delta$. Somit folgt aus $|f(X(\omega)) - f(X_n(\omega))| > \varepsilon$, daß $|X(\omega) - X_n(\omega)| \geq \delta$ und damit insbesondere $|X(\omega) - X_n(\omega)| > \frac{\delta}{2}$. Daher ist

$$P\{|f(X) - f(X_n)| > \varepsilon\} \leq \underbrace{P\{|X - X_n| > \tfrac{\delta}{2}\}}_{\to 0, \text{ da } X_n \to X \text{ stochastisch}} .$$

Man erhält $F(X_n) \longrightarrow f(X)$ stochastisch . Wegen $|f(X_n(\omega))| \leq \sup\limits_{t \in \mathbb{R}} |f(t)| < \infty$ gilt $f(X_n) \longrightarrow f(X)$ in $L^1(P)$ und damit

$$\begin{aligned} \Big| \int f(X_n) dP - \int f(X) dP \Big| &\leq \int |f(X_n) - f(X)| dP \\ &= ||f(X_n) - f(X)||_{L^1(P)} \stackrel{n \to \infty}{\longrightarrow} 0 \,. \end{aligned}$$

Somit $X_n \longrightarrow X$ schwach bzgl $\mathscr{D}(\mathbb{R})$, und nach Satz 6.11 folgt $X_n \longrightarrow X$ in Verteilung. $\square$

$$
\begin{tikzcd}
X_n \to X \text{ in } L^p(P) \arrow[rr, "\text{für eine Teilfolge}", "6.4\text{ b}"'] \arrow[dr, Rightarrow, "6.2\text{ b}"] & & X_n \to X \quad P\text{-f.s.} \arrow[dl, Rightarrow, "6.2\text{a}"'] \\
& \begin{matrix} X_n \to X \\ \text{stochastisch} \end{matrix} \arrow[ul, bend left, "\begin{matrix}\text{falls dominiert}\\ 6.6\end{matrix}"] \arrow[ur, bend right, "\begin{matrix}\text{für Teilfolge}\\ 6.4\text{ a}\end{matrix}"'] \arrow[d, Rightarrow, "6.13"] & \\
& \begin{matrix} X_n \to X \\ \text{in Verteilung} \end{matrix} \arrow[d, Leftrightarrow, "6.11"] & \\
& \begin{matrix} X_n \to X \\ \text{schwach bzgl. } \mathscr{D}(\mathbb{R}) \end{matrix} \arrow[d, Leftrightarrow, "6.11"] & \\
& \begin{matrix} X_n \to X \\ \text{schwach bzgl.} \\ \mathscr{K} = \{t \mapsto \mathrm{e}^{itx} : x \in \mathbb{R}\} \end{matrix} &
\end{tikzcd}
$$

Abbildung 17: Konvergenzarten für eine Folge von Zufallsvariablen

7. Stochastische Unabängigkeit

Wiederholung: Für $A, B \in \mathscr{A}$ mit $P(B) > 0$ war

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

Somit:

$$P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A)P(B).$$

Dies motiviert einen allgemeinen Begriff von Unabängigkeit:

Definition 7.1. Sei $(\Omega,\mathscr{A},P)$ ein Wahrscheinlichkeitsraum, $I \neq \emptyset$ Indexmenge und $X_i \colon (\Omega, \mathscr{A}) \longrightarrow (S_i, \mathscr{S}_i)$ für $i \in I$ Zufallsfunktionen (d.h. X_i $\mathscr{A}$-$\mathscr{S}_i$-meßbar). Dann heißt $(X_i)_{i\in I}$ (stochastisch) unabhängig, falls für jede endliche Teilmenge $I_0 \subset I$ gilt:

$$P\Big(\bigcap_{i\in I_0}\{X_i \in A_i\}\Big) = \prod_{i\in I_0} P\{X_i \in A_i\} \qquad \text{für alle } A_i \in \mathscr{S}_i\,.$$

Beispiel 7.2. Zwei Zufallsvariablen $\{X_1, X_2\}$ sind genau dann unabhängig, wenn

$$P(\{X_1 \in A_1\} \cap \{X_2 \in A_2\}) = P\{X_1 \in A_1\}P\{X_2 \in A_2\}$$

für alle $A_1 \in \mathscr{S}_1, A_2 \in \mathscr{S}_2$.

Bemerkung 7.3. (Siehe auch Maßtheorievorlesung.) Sei $(S_i, \mathscr{S}_i) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Betrachte die Wahrscheinlichkeitsräume $(S_i, \mathscr{S}_i, P \circ X_i^{-1}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}), P \circ X_i^{-1})$. Dann ist dazu das Produkt $(S,\mathscr{S},Q)$ definiert, wobei

$$\begin{aligned}
S &:= \prod_{i\in I} S_i && \text{(kartesisches Produkt)},\\
\mathscr{S} &:= \bigotimes_{i\in I} \mathscr{S}_i && \text{(Produkt von σ-Algebren)},\\
Q &:= \bigotimes_{i\in I} P \circ X_i^{-1} && \text{(Produkt von Maßen)}.
\end{aligned}$$

Die Produkt-σ-Algebra $\mathscr{S}$ kann man einfach beschreiben, wenn man die Projektionen betrachtet. Dazu sei für $J \subset I$ die Abbildung X_J definiert durch

$$\begin{aligned}
X_J : \Omega &\longrightarrow \prod_{j\in J} S_j,\\
\omega &\mapsto (X_j(\omega))_{j\in J},
\end{aligned}$$

d.h. X_J ist die Projektion von $X_I = (X_i(\omega))_{i\in I}$ auf die Menge J. Mit Hilfe der Projektion läßt sich X_j schreiben als

$$X_j(\omega) = pr_j((X_i(\omega))_{i\in I}),$$

wobei pr_j die Projektion auf die j-te Komponente ist,

$$pr_j : \underbrace{\prod_{i \in I} S_i}_{S} \longrightarrow S_j \,.$$

Dann ist

$$\mathscr{S} := \sigma\Big(\bigcup_{j \in I} pr_j^{-1}(\mathscr{S}_j)\Big) = \sigma(\{pr_j : j \in I\}).$$

Die Definition des Maßes Q gehört in die Maßtheorie-Vorlesung, es sei hier nur erwähnt, daß Q als Maß auf $\mathscr{S}$ bereits durch die Angabe aller Werte auf Mengen der Form

$$\prod_{i \in I_0} A_i \times \prod_{i \in I \setminus I_0} S_i$$

mit $I_0 \subset$ I endlich und $A_i \in \mathscr{S}_i$ festgelegt ist. Es gilt: $(X_i)_{i \in I}$ ist genau dann unabhängig, wenn

$$P \circ X_{I_0}^{-1} = \bigotimes_{i \in I_0} P \circ X^{-1} \quad \text{für alle } I_0 \subset I \text{ endlich.}$$

Dies ist äquivalent zu

$$P \circ X_I^{-1} = \bigotimes_{i \in I} P \circ X_i^{-1} \,.$$

Sei nun $I = \{1, \ldots, n\}$ und $(X_1, \ldots, X_n)$ Zufallsvariable. Betrachte

$$(\Omega, \mathscr{A}, P) \stackrel{(X_1,\ldots,X_n)}{\longrightarrow} (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \underbrace{P \circ (X_1, \ldots, X_n)^{-1}}_{\text{gem. Vert. v.}(X_1,\ldots,X_n)}). \tag{$*$}$$

$(X_1, \ldots, X_n)$ ist genau dann unabhängig, wenn $P \circ (X_1, \ldots, X_n)^{-1} = \bigotimes_{i=1}^n P \circ X_i^{-1}$. Da

$$\{A_1 \times \cdots \times A_n : A_i \in \mathscr{B}(\mathbb{R})\}$$

ein Erzeugenden-System von $\mathscr{B}(\mathbb{R}^n) = \mathscr{B}(\mathbb{R}) \otimes \cdots \otimes \mathscr{B}(\mathbb{R})$ ist, gilt $(*)$ genau dann, wenn

$$\underbrace{P \circ (X_1, \ldots, X_n)^{-1}(A_1 \times \cdots \times A_n)}_{P\{X_1 \in A_1, \ldots, X_n \in A_n\}} = \underbrace{\bigotimes_{i=1}^n (P \circ X^{-1})(A_1 \times \cdots \times A_n)}_{\prod_{i=1}^n P \circ X_i^{-1}(A_i)} \,.$$

Beispiele 7.4. a) Sei $\Omega = \{0,1\}^2$ mit dem zweimaligen Werfen einer (Laplace-) Münze als zugehörigem Experiment. Für

$$X_1((\omega_1, \omega_2)) := \omega_1 \,, \quad X_2((\omega_1, \omega_2)) := \omega_1 + \omega_2$$

gilt

$$\begin{aligned} P \circ X_1^{-1} &= \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1, \\ P \circ X_2^{-1} &= \frac{1}{4}\delta_0 + \frac{1}{2}\delta_1 + \frac{1}{4}\delta_2 \,. \end{aligned}$$

Sei A=$\{(0,2)\} \in \mathscr{B}(\mathbb{R}^2)$, d.h. $A = A_1 \times A_2$ mit $A_1 = \{0\}$, $A_2 = \{2\}$. Dann ist

$$\begin{aligned} P \circ (X_1, X_2)(A) = P\{X_1 \in A_1, X_2 \in A_2\} = P\{X_1 = 0, X_2 = 2\} = 0\,, \\ P \circ X_1^{-1}(A_1) = P\{X_1 \in A_1\} = P\{X_1 = 0\} = \frac{1}{2}\,, \\ P \circ X_2^{-1}(A_2) = P\{X_1 \in A_2\} = P\{X_2 = 2\} = \frac{1}{4}\,. \end{aligned}$$

Somit gilt

$$P \circ (X_1, X_2)^{-1}(A) = 0 \neq \frac{1}{8} = (P \circ X_1^{-1} \otimes P \circ X_2^{-1})(A),$$

d.h. (X_1, X_2) ist abhängig.

b) Es gilt: (X, X) ist genau dann unabhängig, wenn $P \circ X^{-1} = \delta_a$ mit $a \in \mathbb{R}$. Denn sei zunächst (X, X) unabhängig. Dann folgt

$$P \circ X^{-1}(A) = (P \circ X^{-1}(A))^2 \Rightarrow P \circ X^{-1}(\mathscr{B}(\mathbb{R})) = \{0, 1\} \overset{\text{Lemma 5.3}}{\Longleftrightarrow} P \circ X^{-1} = \delta_a$$

mit $a = \mathrm{E}\, X$.

Falls andererseits $P \circ X^{-1} = \delta_a$ ist, so gilt

$$\begin{aligned} P\{X \in A_1, X \in A_2\} &= P \circ X^{-1}(A_1 \cap A_2) = \delta_a(A_1 \cap A_2) = \delta_a(A_1)\delta_a(A_2) \\ &= P\{X \in A_1\}P\{X \in A_2\}. \end{aligned}$$

Satz 7.5 (Multiplikationssatz). *Seien $X_1, \ldots, X_n \in L^1(P)$ unabhängig. Dann ist $\prod_{i=1}^n X_i \in L^1(P)$ und*

$$\mathrm{E}\Big(\prod_{i=1}^n X_i\Big) = \prod_{i=1}^n \mathrm{E}\, X_i \,.$$

Beweis. Sei

$$M : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad (x_1, \ldots, x_n) \longmapsto \mid x_1 \cdot \cdots \cdot x_n \mid \,.$$

Dann ist

$$\begin{aligned} \mathrm{E}(\mid X_1 \cdot \ldots \cdot X_n \mid) &= \int_\Omega M \circ (X_1, \ldots, X_n) \mathrm{d}P = \int_{\mathbb{R}^n} M \mathrm{d}P \circ (X_1, \ldots, X_n)^{-1} \\ &= \int_{\mathbb{R}^n} M \mathrm{d} \bigotimes_{i=1}^n P \circ X_i^{-1} \end{aligned}$$

$$\overset{\text{Fubini}}{=} \int_{\mathbb{R}^n} \cdots \int_{\mathbb{R}^n} |\, x_1 \cdot \ldots \cdot x_n \,|\, P \circ X_1^{-1}(\mathrm{d}x_1) \cdot \ldots \cdot P \circ X_n^{-1}(\mathrm{d}x_n)$$
$$= \prod_{i=1}^{n} \underbrace{\int_{\mathbb{R}} |\, X_i \,|\, (P \circ X_i^{-1})(\mathrm{d}x_i)}_{E(|X_i|)}$$
$$= \prod_{i=1}^{n} E(|\, X_i \,|) < \infty\,.$$

Somit ist $X_1 \cdot \ldots \cdot X_1 \in L^1(P)$. Die gleiche Rechnung ohne Betragsstriche zeigt

$$\mathrm{E}(X_1 \cdot \ldots \cdot X_n) = \mathrm{E}\, X_1 \cdot \ldots \cdot \mathrm{E}\, X_n.$$

□

Korollar 7.6. *Seien* $X_1, \ldots, X_n \in L^2(P)$. *Falls* $(X_1, \ldots, X_n)$ *unabhängig ist, so sind* $X_1, \ldots, X_n$ *unkorreliert, d.h. es gilt* $\mathrm{Cov}(X_i, X_j) = 0$ *für* $i \neq j$. *Insbesondere gilt dann*

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}\, X_1 + \cdots + \mathrm{Var}\, X_n.$$

Beweis. Es gilt

$$\mathrm{Cov}(X_i, X_j) = \mathrm{E}((X_i - \mathrm{E}\, X_i)(X_j - \mathrm{E}\, X_j)) \overset{\mathrm{E}(\mathrm{E}\, X) = \mathrm{E}\, X}{=}$$
$$= \underbrace{\mathrm{E}(X_i X_j)}_{= \mathrm{E}\, X_i\, \mathrm{E}\, X_j} - 2\, \mathrm{E}\, X_i\, \mathrm{E}\, X_j + \mathrm{E}\, X_i\, \mathrm{E}\, X_j = 0\,.$$

Der Rest folgt mit Satz 4.14. □

Achtung: Die Umkehrung gilt nicht! Sei $\Omega = \{1, 2, 3\}$, Laplace-Experiment, und X und Y definiert durch

ω	1	2	3
$X(\omega)$	1	0	-1
$Y(\omega)$	0	1	0

Dann gilt $\mathrm{E}\, X = 0$, $\mathrm{E}(XY) = 0$ und damit $\mathrm{Cov}(X, Y) = 0$, d.h. X und Y sind unkorreliert. Aber

$$P\{X = 1, Y = 1\} = 0 \neq P\{X = 1\} P\{Y = 1\} = \frac{1}{9}\,,$$

d.h. (X, Y) ist **nicht** unabhängig.

Definition 7.7. Sei$(X_n)_{n \in \mathbb{N}} \subset L^1(P)$. Dann genügt $(X_n)_n$ dem schwachen Gesetz der großen Zahlen, falls für $S_n := X_1 + \cdots + X_n$ gilt:

$$\frac{S_n - ES_n}{n} \longrightarrow 0 \quad \text{stochastisch.}$$

Satz 7.8 (Schwaches Gesetz der großen Zahlen). *Sei* $(X_n)_n \subset L^2(P), (X_n)_n$ *unabhängig. Es existiere eine Folge* $\beta_n > 0$ *mit* $\beta_n \longrightarrow \infty$ *für* $(n \longrightarrow \infty)$ *und*

$$\frac{1}{\beta_n^2} \sum_{i=1}^{n} \operatorname{Var} X_i \longrightarrow 0 \quad (n \longrightarrow \infty).$$

Dann gilt

$$\frac{1}{\beta_n^2}(S_n - \operatorname{E} S_n) \longrightarrow 0 \qquad \text{in } L^2(P)$$

und damit auch stochastisch.

Beweis. Dies folgt wegen

$$\begin{aligned}\left\| \frac{1}{\beta_n}(S_n - \operatorname{E} S_n) \right\|_{L^2(P)}^2 &= \frac{1}{\beta_n^2} \int_\Omega (S_n - \operatorname{E} S_n)^2 \mathrm{d}P \overset{X_n \text{ unabhängig}}{=} \\ &= \frac{1}{\beta_n^2} \operatorname{E}(S_n - \operatorname{E} S_n)^2 = \frac{1}{\beta_n^2} \operatorname{Var} S_n \overset{7.6}{=} \\ &= \frac{1}{\beta_n^2} \sum_{i=1}^{n} \operatorname{Var} X_i \longrightarrow 0 \quad (n \longrightarrow \infty).\end{aligned}$$

□

Definition und Satz 7.9. *Seien* $\mu_1, \dots, \mu_n$ *endliche Maße auf* $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ *und* $A_n : \mathbb{R}^n \to \mathbb{R}$ *definiert durch* $A_n(x_1, \dots, x_n) = x_1 + \cdots + x_n$*. Dann heißt*

$$\mu_1 * \cdots * \mu_n := (\mu_1 \otimes \cdots \otimes \mu_n) \circ A_n^{-1}$$

das Faltungsprodukt von $\mu_1, \dots, \mu_n$*. Dies ist ein endliches Maß auf* $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$*. Falls* $\mu_i = P \circ X_i^{-1}$ *mit unabhängigen Zufallsvariablen* $(X_1, \dots, X_n)$*, so ist* $\mu_1 * \cdots * \mu_n$ *die W-Verteilung von* $S_n := X_1 + \cdots + X_n$*, d.h.*

$$P \circ S_n^{-1} = (P \circ X_1^{-1}) * \cdots * (P \circ X_n^{-1}).$$

Beweis. Wegen $S_n = A_n \circ (X_1, \dots, X_n)$ gilt mit dem Transformationslemma

$$\begin{aligned}P \circ S_n^{-1} &= P \circ [A_n \circ (X_1, \dots, X_n)]^{-1} = \underbrace{(P \circ (X_1, \dots, X_n)^{-1})}_{(P \circ X_1^{-1}) \otimes \cdots \otimes (P \circ X_n^{-1})} \circ A_n^{-1} \\ &= [(P \circ X_1^{-1}) \otimes \cdots \otimes (P \circ X_n^{-1})] \circ A_n^{-1} = (P \circ X_1^{-1}) * \cdots * (P \circ X_n^{-1}).\end{aligned}$$

□

Definition 7.10 (Unabhängige Mengensysteme). Sei $(\Omega, \mathscr{A}, P)$ W-Raum und $(\mathscr{E}_i)_{i \in I}$ eine Familie von Teilmengensystemen $\mathscr{E}_i \subset \mathscr{A}$. Dann heißt $(\mathscr{E}_i)_{i \in I}$ unabhängig, falls für alle endlichen $I_0 \subset I$ gilt:

$$P\Big(\bigcap_{i \in I_0} A_i \Big) = \prod_{i \in I_0} P(A_i) \qquad \text{für alle } A_i \in \mathscr{E}_i.$$

Bemerkung 7.11. a) Falls $\mathscr{E}_i = \{A_i\}$, so heißen (A_i) unabhängig. Falls die Gleichheit in 7.10 für alle I_0 mit $| I_0 |= 2$ gilt, so heißen $(A_i)_{i\in I}$ paarweise unabhängig. Dies ist nicht äquivalent zur Unabhängigkeit von $(A_i)_{i\in I}$. Dafür, daß $(A_1,\dots,A_n)$ unabhängig ist, genügt **nicht**, daß $P(A_1\cap\dots\cap A_n) = P(A_1)\cdot\ldots\cdot P(A_n)$.

b) Mit Definition 7.1 und 7.10 gilt: $(X_i)_{i\in I}$ ist genau dann unabhängig, wenn die erzeugten σ-Algebren $(\sigma(X_i))_{i\in I}$ unabhängig sind.

c) $(\mathscr{E}_i)_{i\in I}$ ist genau dann unabhängig, wenn für alle endlichen Teilmengen $\tilde{I}\subset I$ das System $(\mathscr{E})_{i\in\tilde{I}}$ unabhängig ist.

d) Sei $(\mathscr{E}_i)_{i\in I}$ unabhängig und $\mathscr{E}_i'\subset\mathscr{E}_i$ für alle i. Dann ist auch $(\mathscr{E}_i')_{i\in I}$ unabhängig.

e) Seien $(X_i)_{i\in I}$ unabhängige Zufallsfunktionen $X_i : (\Omega,\mathscr{A},P)\longrightarrow(S_i,\mathscr{S}_i)$ und $f_i : (S_i,\mathscr{S}_i)\longrightarrow(T_i,\mathscr{S}_i)$ meßbar. Dann ist $(f_i\circ X_i)_{i\in I}$ unabhängig.

Satz 7.12. *Seien $(\mathscr{E}_i)_{i\in I},\mathscr{E}_i\subset\mathscr{A}$ unabhängig.*

a) *Für die zugehörigen Dynkin-Systeme gilt: $(\mathscr{D}(\mathscr{E}_i))_{i\in i}$ unabhängig.*

b) *Falls $\mathscr{E}_i$ $\cap$-stabil ist für jedes $i\in I$, so ist $(\sigma(\mathscr{E}_i))_{i\in I}$ unabhängig.*

Beweis. a) Zu zeigen ist, daß für jedes endliche $\tilde{I}\subset I$ gilt: $(\mathscr{D}(\mathscr{E}_i))_{i\in\tilde{I}}$ unabhängig. Sei $i_0\in\tilde{I}$. Setze: $\mathscr{D}_{i_0} := \{E\in\mathscr{A}\mid(\tilde{\mathscr{E}}_i(E))_{i\in\tilde{I}}$ unabhängig$\}$, wobei

$$\tilde{\mathscr{E}}_i(E) := \begin{cases} \mathscr{E}_i, & i\neq i_0,\\ \{E\}, & i = i_0.\end{cases}$$

Dann ist $\mathscr{D}_{i_0}$ ein Dynkinsystem, denn

(i) $\Omega\in\mathscr{D}_{i_0}$ wegen

$$P(A_{i_1}\cap\dots\cap A_{i_n}\cap\Omega) = P(A_{i_1}\cap\dots\cap A_{i_n}) = P(A_{i_1})\cdot\ldots\cdot P(A_{i_n})\cdot\underbrace{P(\Omega)}_{=1}$$

für alle $I_0 = \{i_1,\dots,i_n\}\cup\{i_0\}\subset\tilde{I}$.

(ii) $E\subset\mathscr{D}_{i_0}\Rightarrow E^c\in\mathscr{D}_{i_0}$, denn

$$\begin{aligned} P(A_{i_1}\cap\dots\cap A_{i_n}\cap E^C) &= P(A_{i_1}\cap\dots\cap A_{i_n}\cap\Omega) - P(A_{i_1}\cap\dots\cap A_{i_n}\cap E)\\ &= P(A_{i_1})\cdot\ldots\cdot P(A_{i_n}) - P(A_{i_1})\cdot\ldots\cdot P(A_{i_n})\cdot P(E)\\ &= P(A_{i_1})\cdot\ldots\cdot P(A_{i_n})\underbrace{(1-P(E))}_{P(E^C)}.\end{aligned}$$

(iii) $(E_k)_{k\in\mathbb{N}}\subset\mathscr{D}_{i_0}, E_k$ disjunkt $\Rightarrow\dot{\bigcup\limits_{k\in\mathbb{N}}}E_k\in\mathscr{D}_{i_0}$, denn

$$P\Big(A_{i_1}\cap\dots\cap A_{i_n}\cap\Big(\dot{\bigcup_{k\in\mathbb{N}}}E_k\Big)\Big)$$

$$
\begin{aligned}
&= P\Big(\dot{\bigcup_{k\in\mathbb{N}}} A_{i_1}\cap\dots\cap A_{i_n}\cap E_k\Big)\\
&= \sum_{i=1}^{\infty} P(A_{i_1})\cdot\ldots\cdot P(A_{i_n})\cdot P(E_k)\\
&= P(A_{i_1})\cdot\ldots\cdot P(A_{i_n})\cdot \sum_{i=1}^{\infty} P(E_k)\\
&= P(A_{i_1})\cdot\ldots\cdot P(A_{i_n})\cdot P\Big(\dot{\bigcup_{k\in\mathbb{N}}} E_k\Big).
\end{aligned}
$$

Nach Definition ist die Familie $(\mathscr{E}_i')_{i\in\tilde{I}}$ mit

$$
\mathscr{E}_i' := \begin{cases} \mathscr{E}_i, & \text{falls } i\neq i_0,\\ \mathscr{D}_{i_0}, & \text{falls } i=i_0,\end{cases}
$$

unabhängig. Aber $\mathscr{E}_{i_0}\subset\mathscr{D}_{i_0}$ und damit $\mathscr{D}(\mathscr{E}_{i_0})\subset\mathscr{D}_{i_0}$. Iteration bzgl. $I_0\in\tilde{I}$ (endlich oft) ergibt: $(\mathscr{D}(\mathscr{E}_i))_{i\in\tilde{I}}$ unabhängig.

b) Dies folgt sofort wegen $\sigma(\mathscr{E}_i)=\mathscr{D}(\mathscr{E}_i)$ für $\cap-$stabile $\mathscr{E}_i$ nach dem Dynkin-Lemma (Lemma 1.3). □

Der folgende Satz heißt auch Satz über erweiterte Unabhängigkeit.

Satz 7.13 (Zusammenfassen unabhängiger $\sigma-$Algebren). *Seien $(\mathscr{E}_i)_{i\in I}$ unabhängig, $\mathscr{E}_i$ $\cap$-stabil, $I=\dot{\bigcup\limits_{j\in J}} I_j$ eine Partition von I. Dann ist $(\sigma(\bigcup_{i\in I_j}\mathscr{E}_i))_{j\in J}$ unabhängig.*

Beweis. Sei

$$
\tilde{\mathscr{E}}_j := \Big\{\bigcup_{i\in I_0} E_i \mid E_i\in\mathscr{E}_i, I_0\subset I_j \text{ endlich}\Big\}.
$$

Dann ist $\tilde{\mathscr{E}}_j$ $\cap-$stabil und $(\tilde{\mathscr{E}}_j)_{j\in J}$ unabhängig, da $(\mathscr{E}_i)_{i\in I}$ unabhängig. Wegen $\sigma(\tilde{\mathscr{E}}_i)=\sigma(\bigcup_{i\in I_j}\mathscr{E}_i)$ folgt die Behauptung aus Satz 7.12.b). □

Korollar 7.14. *Sei $(X_n)_{n\in\mathbb{N}}$ unabhängige Folge von Zufallsvariablen, $1<n_1<n_2<\ldots$ und $f_k:\mathbb{R}^{n_k-n_{k-1}}\longrightarrow\mathbb{R}$ meßbar. Dann ist $(f_k\circ(X_{n_{k-1}+1},\ldots,X_{n_k}))_{k\in\mathbb{N}}$ unabhängig.*

Beweis. $(X_{n_{k-1}+1},\ldots,X_{n_k})\colon\Omega\longrightarrow\mathbb{R}^{n_k-n_{k-1}}$ ist

$$
\sigma(X_{n_{k-1}+1},\ldots,X_{n_k})\text{-}\mathscr{B}(\mathbb{R}^{n_k-n_{k-1}})\text{-meßbar},
$$

also ist $\sigma(f_k\circ(X_{n_{k-1}+1},\ldots,X_{n_k}))\subset\sigma(X_{n_{k-1}+1},\ldots,X_{n_k})$, aber nach Satz 7.13 ist $\sigma(X_{n_{k-1}+1},\ldots,X_{n_k})_{k\in\mathbb{N}}$ unabhängig. Die Behauptung folgt mit 7.11.d). □

Beispiel 7.15. Viermaliges Werfen eines Würfels. Sei Y_1 die Augensumme der ersten beiden Würfe und Y_2 die Augensumme der letzten beiden Würfe. Dann ist (Y_1, Y_2) unabhängig.

8. Null-Eins-Gesetze

Definition 8.1. a) Seien $\mathscr{A}$ und $(\mathscr{A}_n)_{n\in\mathbb{N}}$ σ-Algebren mit $\mathscr{A}_n \subset A$. Dann heißt

$$\mathscr{A}_\infty := \limsup_{n\to\infty} \mathscr{A}_n := \bigcap_{k\in\mathbb{N}} \sigma\Big(\bigcup_{n\geq k} \mathscr{A}_n\Big)$$

die σ-Algebra der terminalen Ereignisse.

b) Sei $\mathscr{A}_n = \sigma(X_n)$ für eine Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen. Dann heißt T terminale Funktion von $(X_n)_{n\in\mathbb{N}}$, falls $T : (\Omega, \mathscr{A}_\infty) \to (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ meßbar ist.

Satz 8.2 (Null-Eins-Gesetz von Kolmogorov). *Sei $(\mathscr{A}_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen σ-Algebren $\mathscr{A}_n \subset \mathscr{A}$. Dann ist jedes $A \in \mathscr{A}_\infty$ deterministisch, es gilt also entweder $P(A) = 0$ oder $P(A) = 1$.*

Beweis. Sei $A \in \mathscr{A}_\infty$ und $\mathscr{D} := \{D \in \mathscr{A} : P(A \cap D) = P(A)P(D)\}$. Wie man leicht nachprüft, ist $\mathscr{D}$ ein Dynkin-System. Nach Satz 7.13 ist

$$\Big(\sigma(\mathscr{A}_1 \cup \ldots \cup \mathscr{A}_n), \sigma\Big(\bigcup_{k>n} \mathscr{A}_k\Big)\Big)$$

unabhängig. Somit gilt wegen $A \in \sigma(\bigcup_{k>n} \mathscr{A}_k)$, daß $\mathscr{E} := \bigcup_{n\in\mathbb{N}} \sigma(\mathscr{A}_1 \cup \ldots \cup \mathscr{A}_n) \subset \mathscr{D}$. Da zwei Mengen aus $\mathscr{E}$ in einer gemeinsamen σ-Algebra $\sigma(\mathscr{A}_1 \cup \ldots \cup \mathscr{A}_n)$ liegen, ist $\mathscr{E}$ $\cap$-stabil. Nach dem Dynkin-Lemma ist $\sigma(\mathscr{E}) = \mathscr{D}(\mathscr{E}) \subset \mathscr{D}$ und folglich

$$\mathscr{A}_\infty \subset \sigma\Big(\bigcup_{n\in\mathbb{N}} \mathscr{A}_n\Big) \subset \sigma(\mathscr{E}) \subset \mathscr{D}.$$

Insbesondere folgt $A \in \mathscr{D}$, das heißt $P(A) = P(A)^2$, also $P(A) \in \{0, 1\}$. $\square$

Korollar 8.3. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Zufallsvariablen und $T : \Omega \to \overline{\mathbb{R}}$ eine terminale Funktion (von $(X_n)_{n\in\mathbb{N}}$). Dann existiert ein $\alpha \in \overline{\mathbb{R}}$ mit $T = \alpha$ P-f.s.*

Beweis. Da T $\mathscr{A}_\infty$-$\mathscr{B}(\overline{\mathbb{R}})$-meßbar ist, gilt nach Satz 8.2 $P\{T \leq \gamma\} \in \{0, 1\}$ für jedes $\gamma \in \overline{\mathbb{R}}$. Sei $\alpha := \inf\{\gamma \in \overline{\mathbb{R}} : P\{T \leq \gamma\} = 1\}$ (∞ liegt sicher in letzterer Menge, so daß diese nicht leer ist). Aufgrund der Rechtsstetigkeit der Verteilungsfunktion folgt $P\{T \leq \alpha\} = 1$, $P\{T < \alpha\} = 0$, somit $T = \alpha$ P-f.s. $\square$

Beispiele 8.4. Sei $t : \mathbb{R}^\mathbb{N} \to \overline{\mathbb{R}}$ mit $t\big((x_n)_{n\in\mathbb{N}}\big) = t\big((x_{n+m})_{n\in\mathbb{N}}\big)$ für alle $m \in \mathbb{N}$ und $T := t \circ (X_n)_{n\in\mathbb{N}}$ meßbar bezüglich $\sigma(\{X_n : n \in \mathbb{N}\})$ (man sagt, t ist ein terminales Funktional). Dann ist T terminale Funktion, beispielsweise:

a) $T := 1_{\{\omega : X_n(\omega) \text{ konvergiert}\}}$ wegen $\{X_n \text{ konvergiert}\} = \{\limsup X_n = \liminf X_n\} \in \sigma(\{X_n : n \in \mathbb{N}\})$.

b) $T := \limsup_{n\in\mathbb{N}} \frac{1}{n}\sum_{k=1}^{n} X_k$ ist terminale Funktion, da

$$\frac{1}{n}\sum_{k=1}^{n} X_{k+m} = \frac{n+m}{n}\,\frac{1}{n+m}\sum_{k=1}^{n+m} X_k - \frac{1}{n}\sum_{k=1}^{m} X_k,$$

und mit der Konvergenz von $\frac{n+m}{n}$ gegen 1 sowie des rechten Summanden gegen 0 folgt obige Eigenschaft.

Ein solches T ist deswegen terminal, weil $T = t(X_1, X_2, \ldots) = t(X_{1+m}, X_{2+m}, \ldots)$ meßbar ist bezüglich $\sigma(\{X_{n+m} : n \in \mathbb{N}\})$ für jedes $m \in \mathbb{N}$, also auch meßbar bezüglich $\bigcap_{m\in\mathbb{N}} \sigma(\{X_{n+m} : n \in \mathbb{N}\}) = \mathscr{A}_\infty$.

Analog zur σ-Algebra der terminalen Ereignisse kann man auch Mengenfolgen betrachten: Zu $A_n \in \mathscr{A}$ definiert man

$$\limsup_{n\to\infty} A_n := \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} A_k.$$

Es ist

$$\begin{aligned}\limsup A_n &= \{\omega \in \Omega : \omega \in A_n \text{ für unendlich viele } n\}\\ &= \{\omega \in \Omega : \omega \in A_n \text{ immer wieder}\},\end{aligned}$$

daher schreibt man manchmal $P(\limsup A_n) = P(A_n \text{ unendlich oft})$.

Bemerkung 8.5. a) $1_{\limsup A_n} = \limsup 1_{A_n}$.

b) Mit $\mathscr{A}_n := \sigma(A_n)$ folgt $\limsup A_n \in \mathscr{A}_\infty$, denn für alle $n \in \mathbb{N}$ gilt $\bigcup_{k\geq n} A_k \in \sigma\Big(\bigcup_{k\geq n} \mathscr{A}_k\Big)$, also $\limsup A_n \in \sigma\Big(\bigcup_{k\geq n} \mathscr{A}_k\Big)$. Falls die Folge $(A_n)_{n\in\mathbb{N}}$ unabhängig ist, folgt nach dem Null-Eins-Gesetz 8.2, daß $P(\limsup A_n) \in \{0, 1\}$.

Satz 8.6 (Borel-Cantelli-Lemma). *Sei $(\Omega, \mathscr{A}, P)$ ein Wahrscheinlichkeitsraum und $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$. Dann gilt*

$$\sum_{n\in\mathbb{N}} P(A_n) < \infty \Longrightarrow P(\limsup A_n) = 0.$$

Falls die $(A_n)_{n\in\mathbb{N}}$ paarweise unabhängig sind (insbesondere also bei Unabhängigkeit der Folge $(A_n)_{n\in\mathbb{N}}$), gilt auch die Umkehrung, genauer sogar

$$\sum_{n\in\mathbb{N}} P(A_n) = \infty \Longrightarrow P(\limsup A_n) = 1.$$

Beweis. a) Sei $\sum_{n\in\mathbb{N}} P(A_n) < \infty$. Dann ist

$$P(\limsup A_n) \leq P\Big(\bigcup_{k\geq n} A_k\Big) \leq \sum_{k\geq n} P(A_k), \quad n \in \mathbb{N},$$

und der letzte Ausdruck strebt gegen 0 mit $n \to \infty$.

b) Sei $\sum_{n\in\mathbb{N}} P(A_n) = \infty$ und die $(A_n)_{n\in\mathbb{N}}$ paarweise unabhängig. Definiere $A := \limsup A_n$, $I_n := 1_{A_n}$, $S_n := \sum_{j=1}^n I_j$ sowie $S := \sum_{j=1}^\infty I_j$. Da die $(A_n)_{n\in\mathbb{N}}$ paarweise unabhängig sind, sind es auch die $(I_n)_{n\in\mathbb{N}}$, folglich sind sie unkorreliert. Nach Satz 4.13 folgt

$$\operatorname{Var} S_n = \sum_{j=1}^n \operatorname{Var} I_j = \sum_{j=1}^n \operatorname{E}(I_j^2) - (\operatorname{E} I_j)^2 = \operatorname{E} S_n - \sum_{j=1}^n (\operatorname{E} I_j)^2 \leq \operatorname{E} S_n,$$

wobei $I_j^2 = I_j$ verwendet wurde. Nach Voraussetzung ist daher

$$\sum_{n\in\mathbb{N}} \operatorname{E} I_n = \sum_{n\in\mathbb{N}} P(A_n) = \infty,$$

somit folgt $\operatorname{E} S = \lim_{n\in\mathbb{N}} \operatorname{E} S_n = \infty$. Wegen

$$\omega \in \limsup A_n \Leftrightarrow \omega \in A_n \text{ unendlich oft} \Leftrightarrow S(\omega) = \infty$$

ist zu zeigen, daß $P\{S = \infty\} = 1$.

Nach Chebyshev gilt $P\{|S_n - \operatorname{E} S_n| \leq \alpha\} \geq 1 - \frac{\operatorname{Var} S_n}{\alpha^2}$. Wegen $\operatorname{E} S_n \to \infty$ sei ohne Einschränkung $\operatorname{E} S_n > 0$ für alle n. Damit ist

$$P\Big\{S_n \geq \frac{1}{2}\operatorname{E} S_n\Big\} \geq P\Big\{|S_n - ES_n| \geq \frac{1}{2}\operatorname{E} S_n\Big\} \geq 1 - \frac{\operatorname{Var} S_n}{4(\operatorname{E} S_n)^2}.$$

Aber $\frac{\operatorname{Var} S_n}{4(\operatorname{E} S_n)^2} \to 0$ wegen $\frac{\operatorname{Var} S_n}{\operatorname{E} S_n} \leq 1$ und $\operatorname{E} S_n \to \infty$. Damit ergibt sich: Für alle $\varepsilon > 0$ existiert $n_0 \in \mathbb{N}$ derart, daß für alle $n \geq n_0$ gilt $P\{S_n \geq \frac{1}{2}\operatorname{E} S_n\} \geq 1 - \varepsilon$. Somit ist $P\{S \geq \frac{1}{2}\operatorname{E} S_n\} \geq P\{S_n \geq \frac{1}{2}\operatorname{E} S_n\} \geq 1 - \varepsilon$ für alle $n \geq n_0$ (beachte $S \geq S_n$) und folglich $P\{S = \infty\} \geq 1 - \varepsilon$. Da ε beliebig war, muß $P\{S = \infty\} = 1$ gelten. □

Korollar 8.7. *Falls die $(A_n)_{n\in\mathbb{N}}$ paarweise unabhängig sind, folgt $P(\limsup A_n) \in \{0, 1\}$ (vgl. Bemerkung 8.5 b).*

Definition 8.8. Eine Folge $(X_n)_{n\in\mathbb{N}} : (\Omega, \mathscr{A}) \to (S, \mathscr{S})$ heißt identisch verteilt, falls $P \circ X_n^{-1} = P \circ X_1^{-1}$ für alle $n \in \mathbb{N}$ ist, und i.i.d. (independent and identically distributed), falls die $(X_n)_{n\in\mathbb{N}}$ unabhängig und identisch verteilt sind.

Satz 8.9. *Sei $(X_n)_{n\in\mathbb{N}}$ eine i.i.d.-Folge nichtkonstanter Zufallsvariablen. Dann ist $P\{X_n \text{ konvergiert}\} = P\{S_n \text{ konvergiert}\} = 0$.*

Beweis. a) Nach Beispiel 8.4 a) ist $\{X_n \text{ konvergiert}\} \in \mathscr{A}_\infty := \limsup \sigma(X_n)$, also folgt nach Satz 8.2 $P\{X_n \text{ konvergiert}\} \in \{0, 1\}$.
Angenommen, $P\{X_n \text{ konvergiert}\} = 1$. Nach Korollar 8.3 ist $\omega \mapsto \limsup X_n(\omega)$

P-f.s. konstant, d.h. es gibt ein $c \in \mathbb{R}$ mit $P\{X_n \to c\} = 1$. Sei $k \in \mathbb{N}$ fest. Da die $(X_n)_{n\in\mathbb{N}}$ unabhängig sind, gilt für

$$A_{n,k} := \left\{ |X_n - c| \geq \frac{1}{k} \right\} \in \sigma(X_n),$$

daß auch $(A_{n,k})_{n\in\mathbb{N}}$ ein unabhängiges System bilden. Nach Bemerkung 8.5 a) ist $P(\limsup_{n\to\infty} A_{n,k}) \in \{0,1\}$. Wegen $P\{X_n \to c\} = 1$ und $P(\limsup_{n\to\infty} A_{n,k}) = P(A_{n,k}$ unendlich oft) gilt $P(\limsup A_{n,k}) = 0$. Nach Satz 8.6 (Borel-Cantelli) ist dann $\sum_{n\in\mathbb{N}} P(A_{n,k}) < \infty$. Wegen $P(A_{n,k}) = P(A_{1,k})$ aufgrund der identischen Verteilung der X_n folgt also $P(A_{1,k}) = 0$, somit $P\{|X_1 - c| \geq \frac{1}{k}\} = 0$ für alle $k \in \mathbb{N}$. Demzufolge wäre $X_1 = c$ P-f.s. im Widerspruch zur Nichtkonstanz von X_1.

b) $\{S_n$ konvergiert$\} \subset \{X_n \to 0\} \subset \{X_n$ konvergiert$\}$. $\square$

9. Starke Gesetze der großen Zahlen

Wir wissen nach Satz 7.8, daß für unabhängige $(X_n)_{n\in\mathbb{N}} \subset L^2(P)$ das schwache Gesetz der großen Zahlen gilt, d.h. es gilt $n^{-1}(S_n - \mathrm{E}\, S_n) \to 0$ stochastisch. Ziel ist es nun, dasselbe P-f.s. zu zeigen. Dazu brauchen wir einiges über die Konvergenz von $S_n := \sum_{i=1}^n X_i$.

Lemma 9.1 (Skorokhod-Lemma). *Seien* $(X_1, \dots, X_n)$ *unabhängige Zufallsvariable und* $\delta > 0$. *Falls*

$$\gamma := \max_{1\le k\le n} P\Big\{\Big|\sum_{i=k}^n X_i\Big| > \frac{\delta}{2}\Big\} < 1\,,$$

so gilt

$$P\Big\{\max_{1\le k\le n}\Big|\sum_{i=1}^n X_i\Big| > \delta\Big\} \le \frac{\gamma}{1-\gamma}$$

(*ein Beispiel sogenannter Maximalungleichungen*).

Beweis. Sei

$$A := \Big\{\max_{1\le k\le n}\Big|\sum_{i=1}^k X_i\Big| > \delta\Big\}.$$

Für $\omega \in A$ definiere

$$T(\omega) := \min\Big\{k \in \{1, \dots, n\} : \Big|\sum_{i=1}^k X_i(\omega)\Big| > \delta\Big\}.$$

Dann gilt $A = \bigcup_{k=1}^n A_k$ mit $A_k := \{\omega : T(\omega) = k\} \in \sigma(X_1, \dots, X_k)$. Wegen

$$1-\gamma \le P\Big\{\Big|\sum_{i=k+1}^n X_i\Big| \le \frac{\delta}{2}\Big\}$$

gilt

$$(1-\gamma)P(A) \le (1-\gamma)\sum_{k=1}^n P(A_k) \le \sum_{k=1}^n P(A_k)P\Big\{\Big|\sum_{i=k+1}^n X_i\Big| \le \frac{\delta}{2}\Big\}.$$

In jedem Term der letzten Summe ist der erste Faktor in $\sigma(X_1, \dots, X_k)$ und der zweite Faktor in $\sigma(X_{k+1}, \dots, X_n)$. Aufgrund der Unabhängigkeit ergibt sich daher

$$\begin{aligned}(1-\gamma)P(A) &\le \sum_{k=1}^n P\Big(A_k \cap \Big\{\Big|\sum_{i=k+1}^n X_i\Big| \le \frac{\delta}{2}\Big\}\Big)\\ &\le \sum_{k=1}^n P\Big(A_k \cap \Big\{|S_n| > \frac{\delta}{2}\Big\}\Big) = P\Big\{|S_n| > \frac{\delta}{2}\Big\} \le \gamma\,.\end{aligned}$$

Bei der letzten Ungleichung ist dabei zu beachten, daß

$$A_k \cap \Big\{\Big|\sum_{i=k+1}^{n} X_i\Big| \le \frac{\delta}{2}\Big\} \subset A_k \cap \Big\{|S_n| > \frac{\delta}{2}\Big\}$$

gilt. □

Satz 9.2 (Skorokhod-Lévy). *Sei $(X_n)_n$ unabhängig. Dann konvergiert die Reihe $\sum_{n\in\mathbb{N}} X_n$ genau dann stochastisch, wenn sie P-f.s. konvergiert.*

Beweis. Zu zeigen ist nur die Richtung von stochastischer Konvergenz nach P-fast sicherer Konvergenz. Sei $N \in \mathbb{N}$ fest. Da $\sum_n X_n$ stochastisch konvergiert, gilt

$$P\{|S_n - S_m| > (2N)^{-1}\} \to 0 \quad \text{für } m \to \infty \text{ und } n \ge m\,.$$

Insbesondere ist

$$\gamma_{nm} := \max_{m\le k\le n} P\{|S_n - S_m| > (2N)^{-1}\} < \frac{1}{2}$$

für $m \ge m_0$ und $n \ge m$, falls m_0 geeignet gewählt wird. Wende Lemma 9.1 an auf $(X_{m+1}, \ldots, X_{m+n})$ und $\delta = N^{-1}$:

$$P\Big\{\max_{m+1\le k\le m+n}\Big|\sum_{i=m+1}^{k} X_i\Big| > \frac{1}{N}\Big\} \le \frac{\gamma_{nm}}{1-\gamma_{nm}}\,.$$

Wegen $\gamma_{nm} \to 0$ für $m \to \infty$ und $n \ge m$ gilt für

$$A_N := \bigcap_{m\in\mathbb{N}} \Big\{\sup_{n>m}\Big|\sum_{i=m+1}^{n} X_i\Big| > \frac{1}{N}\Big\}$$

die Abschätzung

$$P(A_N) \le P\Big\{\sup_{n>m}\Big|\sum_{i=m+1}^{n} X_i\Big| > \frac{1}{N}\Big\}\,.$$

Da die rechte Seite für $m \to \infty$ gegen Null konvergiert, erhält man $P(A_N) = 0$ für alle $N \in \mathbb{N}$ und damit $P(\bigcup_{N\in\mathbb{N}} A_N) = 0$. Somit ergibt sich

$$\begin{aligned}
P\{S_n \text{ konvergiert }\} &= P\Big\{\omega \in \Omega : \forall\, N \in \mathbb{N}\ \exists\, m \in \mathbb{N} : \sup_{n>m} |S_n(\omega) - S_m(\omega)| \le \frac{1}{N}\Big\} \\
&= P\Big(\bigcap_{N\in\mathbb{N}} \bigcup_{m\in\mathbb{N}} \Big\{\sup_{n>m} |S_n - S_m| \le \frac{1}{N}\Big\}\Big) \\
&= P\Big(\bigcup_{N\in\mathbb{N}} A_N\Big)^c = 1\,.
\end{aligned}$$

□

Lemma 9.3 (Kronecker-Lemma). *Seien* $(x_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ *und* $(a_n)_{n\in\mathbb{N}} \subset (0,\infty)$ *mit* $a_n \nearrow \infty$. *Dann gilt*

$$\sum_{n\in\mathbb{N}} \frac{x_n}{a_n} < \infty \implies \frac{1}{a_n}\sum_{j=1}^{n} x_j \to 0\,.$$

Beweis. Setze $b_n := \sum_{j=1}^{n} \frac{x_j}{a_j}$ und $a_0 := b_0 := 0$. Dann ist $x_n = a_n(b_n - b_{n-1})$ und („partielle Summation“)

$$\begin{aligned}\frac{1}{a_n}\sum_{j=1}^{n} x_j &= \frac{1}{a_n}\sum_{j=1}^{n} a_j(b_j - b_{j-1})\\ &= b_n - \frac{1}{a_n}\sum_{j=0}^{n-1} b_j(a_{j+1}-a_j) = \frac{1}{a_n}\sum_{j=0}^{n-1}(b_n - b_j)(a_{j+1}-a_j)\,.\end{aligned}$$

Beim letzten Gleichheitszeichen verwendeten wir

$$\frac{1}{a_n}\sum_{j=0}^{n-1}(a_{j+1}-a_j) = 1\,. \tag{$*$}$$

Sei $\varepsilon > 0$. Wähle $m \in \mathbb{N}$ mit $\sup_{n,j\geq m}|b_n - b_j| < \frac{\varepsilon}{2}$. Dann gilt für $n > m$:

$$\frac{1}{a_n}\Big|\sum_{j=1}^{n} x_j\Big| \leq \frac{1}{a_n}\Big|\sum_{j=0}^{m-1}(b_n-b_j)(a_{j+1}-a_j)\Big| + \Big[\frac{1}{a_n}\sum_{j=m}^{n-1}(a_{j+1}-a_j)\Big]\cdot\frac{\varepsilon}{2}\,.$$

Die erste Summe ist bezüglich n beschränkt, der Ausdruck in eckigen Klammern ist wegen ($*$) nicht größer als 1. Somit wird die rechte Seite der letzten Ungleichung kleiner oder gleich ε für hinreichend großes n. Damit erhält man

$$\lim_{n\to\infty}\frac{1}{a_n}\sum_{j=1}^{n} x_j = 0\,.$$

□

Satz 9.4 (Erstes starkes Gesetz der großen Zahlen von Kolmogorov). *Sei* $(X_n)_{n\in\mathbb{N}} \subset L^2(P)$ *eine unabhängige Folge und* $0 < \beta_n \nearrow \infty$ *eine Zahlenfolge. Falls*

$$\sum_{n\in\mathbb{N}} \frac{1}{\beta_n^2}\operatorname{Var} X_n < \infty\,,$$

so gilt

$$\frac{S_n - \operatorname{E} X_n}{\beta_n} \to 0 \qquad P\text{-f.s.}$$

Beweis. Sei $Y_n := X_n - \mathrm{E}\, X_n$. Da $(Y_n)_n$ unabhängig ist, folgt

$$\left\langle \frac{Y_n}{\beta_n}, \frac{Y_k}{\beta_k} \right\rangle = \frac{1}{\beta_n \beta_k} \mathrm{E}(Y_n Y_k) = \frac{1}{\beta_n \beta_k} \mathrm{E}\, Y_n \, \mathrm{E}\, Y_k = 0$$

für alle $n \neq k$, wobei $\langle \cdot, \cdot \rangle$ das L_2-Skalarprodukt bezeichnet. Die Folge $\left(\sum_{k=1}^{n} \frac{Y_k}{\beta_k} \right)_{n \in \mathbb{N}}$ ist Cauchy-Folge in $L^2(P)$ wegen

$$\Big\| \sum_{k=n+1}^{m} \frac{Y_k}{\beta_k} \Big\|_{L^2}^2 = \sum_{k=n+1}^{m} \Big\| \frac{Y_k}{\beta_k} \Big\|_{L^2}^2 = \sum_{k=n+1}^{m} \frac{\operatorname{Var} X_k}{\beta_k^2} \to 0 \,, \quad n, m \to \infty \,.$$

Da $L^2(P)$ vollständig ist, existiert $Y := \sum_{k \in \mathbb{N}} \frac{Y_k}{\beta_k} \in L^2(P)$, und es gilt $\sum_{k=1}^{n} \frac{Y_k}{\beta_k} \to Y$ in $L^2(P)$, also stochastisch, also (mit Satz 9.2) P-f.s. Nach dem Kronecker-Lemma (9.3) folgt $\frac{1}{\beta_n} \sum_{k=1}^{n} Y_k \to 0$ P-f.s. □

In Satz 9.4 war $X_n \in L^2(P)$ vorausgesetzt, obwohl in der Behauptung nur der Erwartungswert auftaucht. Ziel ist es nun, auf die Voraussetzung $X_n \in L^2(P)$ zu verzichten (für i.i.d.-Folgen). Dazu (und nicht nur dazu) folgende Definition:

Definition 9.5. Zwei Folgen $(X_n)_{n \in \mathbb{N}}$ und $(Y_n)_{n \in \mathbb{N}}$ heißen äquivalent, falls

$$\sum_{n \in \mathbb{N}} P\{X_n \neq Y_n\} < \infty \,.$$

Lemma 9.6. *Falls $(X_n)_n$ und $(Y_n)_n$ äquivalent sind, gilt: $\sum_n (X_n - Y_n)$ konvergiert P-f.s. , und für $a_n \nearrow \infty$ gilt*

$$\frac{1}{a_n} \sum_{k=1}^{n} (X_k - Y_k) \to 0 \qquad P\text{-f.s.} \;.$$

Beweis. Nach Borel-Cantelli (8.6) gilt

$$P(\limsup_n \{X_n \neq Y_n\}) = 0 \,,$$

d.h. es existiert eine Menge N mit $P(N) = 0$, so daß für alle $\omega \in \Omega \backslash N$ gilt:

$$\omega \in (\limsup_n \{X_n \neq Y_n\})^c = \Big(\bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} \{X_k \neq Y_k\} \Big)^c = \bigcup_{n \in \mathbb{N}} \bigcap_{k \geq n} \{X_k = Y_k\} \,.$$

Somit existiert für jedes $\omega \in \Omega \backslash N$ ein $n_0(\omega)$, so daß für alle $k \geq n_0(\omega)$ gilt: $X_k(\omega) = Y_k(\omega)$. Daher besteht $\sum_{n \in \mathbb{N}} (X_n(\omega) - Y_n(\omega))$ nur aus endlich vielen von Null verschiedenen Summanden. Somit konvergiert diese Reihe und $\frac{1}{a_n} \sum_{k=1}^{n} (X_k(\omega) - Y_k(\omega)) \to 0$ für alle $\omega \in \Omega \backslash N$. □

Satz 9.7 (Zweites starkes Gesetz der großen Zahlen von Kolmogorov). *Sei $(X_n)_{n\in\mathbb{N}}$ eine i.i.d.-Folge von Zufallsvariablen mit $X_1 \in L^1(P)$. Dann gilt*

$$\frac{S_n}{n} \to \mathrm{E}\, X_1 \ P\text{-f.s.}$$

Beweis. Definiere

$$Y_n(\omega) := \begin{cases} X_n(\omega)\,, & \text{falls } |X_n(\omega)| \le n\,, \\ 0\,, & \text{sonst} \end{cases}$$

(Abschneidung von X_n). Wir wollen Satz 9.4 auf Y_n anwenden. Es gilt

$$\begin{aligned} \sum_{n\in\mathbb{N}} \frac{1}{n^2} \operatorname{Var} Y_n &\le \sum_{n\in\mathbb{N}} \frac{\mathrm{E}\, Y_n^2}{n^2} = \sum_{n\in\mathbb{N}} \frac{1}{n^2} \int\limits_{[-n,n]} t^2 P \circ X_1^{-1}(\mathrm{d}t) \\ &= \sum_{j=1}^{n} \int\limits_{\{j-1<|t|\le j\}} t^2 P \circ X_1^{-1}(\mathrm{d}t) \cdot \sum_{n=j}^{\infty} \frac{1}{n^2} \\ &\le C \sum_{j=1}^{\infty} j \int\limits_{\{j-1<|t|\le j\}} |t| P \circ X_1^{-1}(\mathrm{d}t) \cdot \frac{1}{j} \\ &= C \int_{\mathbb{R}} |t| P \circ X_1^{-1}(\mathrm{d}t) = C\, \mathrm{E}\, |X_1| < \infty\,. \end{aligned}$$

Dabei wurde $\sum_{n=j}^{\infty} \frac{1}{n^2} \le \frac{C}{j}$ verwendet (siehe unten). Nach Satz 9.4 gilt $\frac{1}{n}\sum_{j=1}^{n}(Y_j - \mathrm{E}\, Y_j) \to 0 \ \ P$-f.s. . Wegen $\mathrm{E}\, Y_n \to \mathrm{E}\, X_1$ gilt $\frac{1}{n}\sum_{j=1}^{n} \mathrm{E}\, Y_j \to \mathrm{E}\, X_1$ (siehe unten), d.h. $\frac{1}{n}\sum_{j=1}^{n} Y_j \to \mathrm{E}\, X_1 \ \ P$-f.s. . Nach Lemma 9.6 folgt

$$\frac{1}{n}\sum_{j=1}^{n} X_j = \frac{S_n}{n} \to \mathrm{E}\, X_1 \qquad P\text{-f.s.}$$

Lemma 9.6 ist anwendbar, da

$$\sum_{n\in\mathbb{N}} P\{X_n \ne Y_n\} = \sum_{n\in\mathbb{N}} P\{|X_n| > n\} = \sum_{n\in\mathbb{N}} P\{|X_1| > n\} \le \mathrm{E}\, |X_1| < \infty$$

nach Satz 4.10. □

In obigem Beweis wurden folgende elementare Tatsachen aus der Analysis verwendet:

1) Es gilt $\sum_{n=j}^{\infty} \frac{1}{n^2} \le \frac{C}{j}$ für eine Konstante C. Dies ist der Fall wegen

$$\sum_{n=j}^{\infty} \frac{1}{n^2} \le \int_j^{\infty} \frac{1}{(t-1)^2} \mathrm{d}t = \frac{1}{j-1} \le \frac{C}{j}\,.$$

2) Falls $a_n \to a$ für $n \to \infty$ für eine komplexe Zahlenfolge $(a_n)_n$, so gilt $\frac{1}{n}\sum_{j=1}^n a_j \to a$ für $n \to \infty$. Dies sieht man folgendermaßen: Für alle $\varepsilon > 0$ existiert ein n_0, so daß für alle $n \ge n_0$ gilt $|a_n - a| < \varepsilon$. Damit folgt für alle $n \ge n_0$

$$\Big|\frac{1}{n}\sum_{j=1}^{n}(a_j - a)\Big| \le \frac{1}{n}\sum_{j=1}^{n_0-1}|a_j - a| + \frac{1}{n}\sum_{j=n_0}^{n}|a_j - a|\,.$$

Die erste Summe auf der rechten Seite ist beschränkt für $n \to \infty$, während die zweite Summe nicht größer als $(n - n_0 + 1)\varepsilon$ wird. Damit geht der Ausdruck auf der linken Seite der letzten Ungleichung für $n \to \infty$ gegen Null.

Korollar 9.8. *Sei $(Y_n)_{n\in\mathbb{N}}$ eine i.i.d.-Folge von Zufallsfunktionen $Y_n\colon (\Omega, \mathscr{A}, P) \to (S, \mathscr{S})$ und $f\colon S \to \mathbb{R}$ meßbar mit $f \circ Y \in L^1(P)$. Dann gilt*

$$\frac{1}{n}\sum_{j=1}^{n} f \circ Y_j \to \mathrm{E}(f \circ Y_1) \qquad P\text{-f.s.}$$

Beweis. $(f \circ Y_n)_n$ ist i.i.d.-Folge von L^1-Zufallsvariablen. □

Satz 9.9 (Hauptsatz der mathematischen Statistik).
a) *Sei $(Y_n)_n$ wie in Korollar 9.8. Dann gilt für jedes $B \in \mathscr{S}$:*

$$\frac{1}{n}\Big|\Big\{j \in \{1,\dots,n\} : Y_j \in B\Big\}\Big| \to P\{Y_1 \in B\} \qquad P\text{-f.s.}$$

Auf der linken Seite steht die relative Häufigkeit des Ereignisses B. Die in „P-f.s.“ steckende Nullmenge kann dabei von B abhängen.

b) *Sei $(Y_n)_n$ eine i.i.d.-Folge von Zufallsvariablen. Definiere die empirische Verteilungsfunktion F_n durch*

$$F_n(\omega, t) := \frac{1}{n}\Big|\Big\{j \in \{1,\dots,n\} : Y_j \le t\Big\}\Big| \quad (\omega \in \Omega\,,\ t \in \mathbb{R})\,.$$

Dann gilt

$$F_n(\omega, t) \to F_{Y_1}(t) \qquad P\text{-f.s.}\ ,$$

wobei die Nullmenge von t abhängen kann.

Beweis. a) Setze in Korollar 9.8 $f = 1_B$ (damit ist $f \circ Y_n \in L^\infty(P)$).

b) Setze in a) $B = (-\infty, t]$. □

Bemerkung 9.10. Es gilt sogar (Lemma von Glivenko-Cantelli): Es gibt ein $N \in \mathscr{A}$ mit $P(N) = 0$ und

$$F_n(\omega, t) \to F_{Y_1}(t) \quad \text{gleichmäßig in } t$$

für alle $\omega \in \Omega\backslash N$.

Beispiel 9.11 (Monte-Carlo-Methode). Sei $(S, \mathscr{S}, \mu)$ ein W-Raum und $(Y_n)_n$, f wie in Korollar 9.8 mit $P \circ Y_1^{-1} = \mu$. Dann gilt $\frac{1}{n}\sum_{j=1}^n f \circ Y_j \to \int f \mathrm{d}\mu$ P-f.s. Dies kann zur Approximation des Integrals verwendet werden. Dazu ist insbesondere zu beachten, daß

$$\frac{1}{n}\sum_{j=1}^{n} f \circ Y_j = \frac{1}{n}\Big((n-1)\Big[\frac{1}{n-1}\sum_{j=1}^{n-1} f \circ Y_j\Big] + f \circ Y_n\Big)$$

gilt, d.h. daß beim Schritt von $n-1$ auf n die Funktion f lediglich einmal ausgewertet werden muß (der Ausdruck in $[\ldots]$ ist genau das Ergebnis des $(n-1)$-ten Schritts).

Die Monte-Carlo-Methode ist ein beliebtes Verfahren, um Integrale approximativ zu berechnen. Der Vorteil hierbei ist etwa, daß (falls f eine charakteristische Funktion ist) man lediglich entscheiden können muß, ob die beobachtete Zufallsfunktion in einer Menge liegt oder nicht. Falls man jedoch einen expliziten Ausdruck für das zu approximierende Integral kennt, gibt es wesentlich bessere numerische Verfahren. Die Konvergenz der Monte-Carlo-Methode ist i.a. relativ langsam.

Die Abbildungen 18–**??** zeigen die Anwendung der Monte-Carlo-Methode zur Berechnung von

$$\ln 2 = \int_0^1 \frac{1}{1+x}\mathrm{d}x$$

und

$$\pi = 4\int_0^1 \sqrt{1-x^2}\,\mathrm{d}x\,.$$

Man erkennt, daß selbst für $N = 100$ der approximierende Wert noch relativ stark schwankt (Abbildungen 18 und 20). Als Vergleich dazu ist in den Abbildungen 19 und **??** die Berechnung als Riemann-Summe mit äquidistanten Stützstellen zu sehen. Natürlich ist auch dies kein numerisch akzeptables Verfahren. Berechnet man etwa das Integral für $\ln 2$ mit Hilfe einer Newton-Cotes-Formel durch Auswertung an 7 Stützstellen, den Wert von $\ln 2 = 0.6931471806\ldots$ auf 5 Stellen genau.

Abbildung 20: Monte-Carlo-Methode zur Berechnung von $\pi/4$.

Abbildung 19: Berechnung von $\ln 2$ durch Riemann-Summen.

Abbildung 18: Monte-Carlo-Methode zur Berechnung von $\ln 2$.

Dennoch liefert die Monte-Carlo-Methode ein interessantes und für manche Zwecke nützliches Verfahren zur Approximation von Integralen.

10. Charakteristische Funktion und zentraler Grenzwertsatz

Definition 10.1. Sei X eine Zufallsvariable auf dem W-Raum $(\Omega, \mathscr{A}, P)$. Dann heißt

$$\varphi_X(t) := \mathrm{E}(\mathrm{e}^{\mathrm{i}tX}) = \int_\Omega \mathrm{e}^{\mathrm{i}tX(\omega)} P(\mathrm{d}\omega) = \int_\mathbb{R} \mathrm{e}^{\mathrm{i}tx} (P \circ X^{-1})(\mathrm{d}x)$$

die charakteristische Funktion von X. Für ein endliches Borelmaß μ auf $\mathbb{R}$ heißt

$$\hat{\mu}(t) := \int_\mathbb{R} \mathrm{e}^{\mathrm{i}tx} \mu(\mathrm{d}t)$$

die Fourier-Transformierte von μ. (Damit ist $\varphi_X(t) = (P \circ X^{-1})(t)$.)

Lemma 10.2. *Sei X eine Zufallsvariable.*

a) *Es gilt* $|\varphi_X(t)| \leq 1 = \varphi_X(0)\,,\ \varphi_X(-t) = \overline{\varphi_X(t)}$ *für alle* $t \in \mathbb{R}$ (*dabei ist* $\overline{\varphi}$ *die komplexe Konjugation*).

b) φ_X *ist gleichmäßig stetig auf* $\mathbb{R}$.

c) *Es gilt* $\varphi_{aX+b}(t) = \varphi_X(at)\mathrm{e}^{\mathrm{i}tb}$ *für alle* $a, b \in \mathbb{R}$.

d) *Seien* $(X_1, \ldots, X_n)$ *unabhängige Zufallsvariable und* $S_n := X_1 + \ldots + X_n$. *Dann gilt*

$$\varphi_{S_n}(t) = \prod_{j=1}^n \varphi_{X_j}(t) \quad (t \in \mathbb{R})\,.$$

Beweis. a) ist offensichtlich.

b) Es gilt

$$\begin{aligned} |\varphi_X(t+h) - \varphi_X(t)| &= \left| \int (\mathrm{e}^{\mathrm{i}(t+h)x} - \mathrm{e}^{\mathrm{i}tx})(P \circ X^{-1})(\mathrm{d}x) \right| \\ &\leq \int |\mathrm{e}^{\mathrm{i}hx} - 1| (P \circ X^{-1})(\mathrm{d}x)\,. \end{aligned}$$

Da $|\mathrm{e}^{\mathrm{i}hx} - 1| \leq 2$, ist der Satz über majorisierte Konvergenz anwendbar, und es folgt $|\varphi_X(t+h) - \varphi_X(t)| \to 0$. Da außerdem $\int |\mathrm{e}^{\mathrm{i}hx} - 1|(P \circ X^{-1})(\mathrm{d}x)$ von t unabhängig ist, ist die Konvergenz gleichmäßig.

c) Es gilt $\varphi_{aX+b}(t) = \mathrm{E}(\mathrm{e}^{\mathrm{i}t(aX+b)}) = \mathrm{E}(\mathrm{e}^{\mathrm{i}atX}\mathrm{e}^{\mathrm{i}tb}) = \mathrm{e}^{\mathrm{i}tb}\varphi_X(at)$.

d) Für unabhängige $(X_1, \ldots, X_n)$ ist nach Bemerkung 7.11 e) auch $(\mathrm{e}^{\mathrm{i}tX_1}, \ldots, \mathrm{e}^{\mathrm{i}tX_n})$ unabhängig. Damit ist

$$\mathrm{E}(\mathrm{e}^{\mathrm{i}t(X_1 + \ldots + X_n)}) = \mathrm{E}(\mathrm{e}^{\mathrm{i}tX_1} \cdot \ldots \cdot \mathrm{e}^{\mathrm{i}tX_n}) = \prod_{j=1}^n \mathrm{E}(\mathrm{e}^{\mathrm{i}tX_j})\,.$$

□

Wir haben bereits die Faltung $\mu_1 * \mu_2$ von Maßen kennengelernt. Die Fourier-Transformierte davon ist besonders einfach zu beschreiben:

Lemma 10.3. *Seien μ_1, μ_2 endliche Maße auf $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.*

a) Für $h \in L^1(\mu_1 * \mu_2)$ gilt

$$\int_{\mathbb{R}} h \mathrm{d}(\mu_1 * \mu_2) = \int_{\mathbb{R}} \int_{\mathbb{R}} h(x_1 + x_2) \mu_1(\mathrm{d}x_1) \mu_2(\mathrm{d}x_2) .$$

b) Es gilt $(\mu_1 * \mu_2)\hat{}(t) = \hat{\mu}_1(t)\hat{\mu}_2(t)$.

Beweis. a) Sei $A_2(x_1, x_2) := x_1 + x_2$. Dann folgt mit Fubini

$$\begin{aligned}\int_{\mathbb{R}} h \mathrm{d}(\mu_1 * \mu_2) &= \int_{\mathbb{R}} h \mathrm{d}[(\mu_1 \otimes \mu_2) \circ A_2^{-1}] \\ &= \int_{\mathbb{R}^2} h \circ A_2 \mathrm{d}(\mu_1 \otimes \mu_2) = \int_{\mathbb{R}} \int_{\mathbb{R}} h(x_1 + x_2) \mu_1(\mathrm{d}x_1) \mu_2(\mathrm{d}x_2) .\end{aligned}$$

b) Wegen a) gilt

$$\int \mathrm{e}^{\mathrm{i}tx} (\mu_1 * \mu_2)(\mathrm{d}x) = \int \int \mathrm{e}^{\mathrm{i}t(x_1 + x_2)} \mu_1(\mathrm{d}x_1) \mu_2(\mathrm{d}x_2) = \hat{\mu}_1(t)\hat{\mu}_2(t) .$$

□

Beispiele 10.4. a) Es gilt $\hat{\delta}_a = \mathrm{e}^{\mathrm{i}at}$.

b) Sei $\mu = q\delta_0 + p\delta_1$ mit $p = 1 - q \in [0, 1]$ (Bernoulli-Verteilung). Dann ist $\hat{\mu}(t) = q + p\mathrm{e}^{\mathrm{i}t}$.

c) Binomial-Verteilung: Sei $\mu = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \delta_k$. Dann ist

$$\hat{\mu}(t) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \mathrm{e}^{\mathrm{i}kt} = (q + p\mathrm{e}^{\mathrm{i}t})^n .$$

d) Poisson-Verteilung: Für $\pi_\lambda = \mathrm{e}^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \delta_n$ gilt

$$\hat{\pi}_\lambda(t) = \mathrm{e}^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \mathrm{e}^{\mathrm{i}nt} = \exp(\lambda(\mathrm{e}^{\mathrm{i}t} - 1)) .$$

e) Exponentialverteilung: Für $\mu = \lambda \mathrm{e}^{-\lambda x} 1_{[0,\infty)} \mathrm{d}x$ ist $\hat{\mu}(t) = \frac{\lambda}{\lambda - it}$.

f) Normalverteilung: Für $N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) \mathrm{d}x$ ist

$$N(\mu, \sigma^2)\hat{}(t) = \mathrm{e}^{\mathrm{i}\mu t - (\sigma^2 t^2)/2} .$$

Insbesondere ist $N(0, 1)\hat{}(t) = \mathrm{e}^{-t^2/2}$ (vgl. Übungsaufgabe 38).

Definition 10.5 (Faltung von Funktionen). Seien $f_1, f_2 \in L^1(\mathbb{R})$. Dann heißt

$$f_1 * f_2(x) := \int_{\mathbb{R}} f_1(x-y) f_2(y) \mathrm{d}y$$

das Faltungsprodukt von f_1 und f_2. (Dieses existiert fast überall nach Fubini.)

Satz 10.6. *Seien X_1, X_2 unabhängig und stetig verteilt mit Dichten f_1, f_2. Dann ist $X_1 + X_2$ stetig verteilt mit Dichte $f_1 * f_2$.*

Beweis. Nach Satz 7.9 ist $P \circ (X_1 + X_2)^{-1} = (P \circ X_1^{-1}) * (P \circ X_2^{-1})$. Bezeichnet man $P \circ X_i^{-1}$ mit μ_i, so ist mit Lemma 10.3 a)

$$\begin{aligned} \mu_1 * \mu_2 &= \int_{\mathbb{R}} \int_{\mathbb{R}} 1_A(x_1 + x_2) \mu_1(\mathrm{d}x_1) \mu_2(\mathrm{d}x_2) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} 1_A(x_1 + x_2) f_1(x_1) f_2(x_2) \mathrm{d}x_1 \mathrm{d}x_2 \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} 1_A(y) f_1(y - x_2) f_2(x_2) \mathrm{d}y \mathrm{d}x_2 \\ &= \int_{\mathbb{R}} 1_A(y) \left[\int_{\mathbb{R}} f_1(y - x_2) f_2(x_2) \mathrm{d}x_2 \right] \mathrm{d}y \\ &= \int_A (f_1 * f_2)(y) \mathrm{d}y \,, \end{aligned}$$

d.h. es ist $\mu_1 * \mu_2 = (f_1 * f_2)(y)\mathrm{d}y$. □

Satz 10.7 (Eindeutigkeitssatz). *Die Abbildung $\mu \mapsto \hat{\mu}$ ist injektiv, d.h. für zwei endliche Maße μ_1, μ_2 auf $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ mit $\mu_1(t) = \mu_2(t)$ für alle $t \in \mathbb{R}$ gilt $\mu_1 = \mu_2$.*

Der Beweis dieses Satzes wird in der Maßtheorie-Vorlesung durchgeführt.

Korollar 10.8 (Faltungshalbgruppen). *Es gilt $\pi_\lambda * \pi_\mu = \pi_{\lambda+\mu}$ und $N(\mu_1, \sigma_1^2) * N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Insbesondere gilt: Sind X_i normalverteilt mit Erwartungswert μ_i und Varianz σ_i^2 für $i = 1, \ldots, n$ und ist $(X_1, \ldots, X_n)$ unabhängig, so ist $S_n := X_1 + \ldots + X_n$ ebenfalls normalverteilt mit Erwartungswert $\mu_1 + \ldots + \mu_n$ und Varianz $\sigma_1^2 + \ldots + \sigma_n^2$.*

Beweis. Man rechnet die charakteristische Funktion aus. Es folgt

$$\begin{aligned} (\pi_\lambda * \pi_\mu)\hat{}(t) &= \hat{\pi}_\lambda(t) \hat{\pi}_\mu(t) = \exp(\lambda(\mathrm{e}^{\mathrm{i}t} - 1)) \exp(\mu(\mathrm{e}^{\mathrm{i}t} - 1)) \\ &= \exp((\lambda + \mu)(\mathrm{e}^{\mathrm{i}t} - 1)) = \hat{\pi}_{\lambda+\mu} \,. \end{aligned}$$

Die Behauptung folgt nun aus Satz 10.7. Genauso geht's für $N(\mu_i, \sigma_i^2)$. □

Satz 10.9 (Ableiten unter dem Integral). *Sei $X \in L^n(P)$ für ein $n \in \mathbb{N}$. Dann ist φ_X n-mal stetig differenzierbar, und*

$$\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)^k \varphi_X(t) = \int (\mathrm{i}x)^k \mathrm{e}^{\mathrm{i}tx} (P \circ X^{-1})(\mathrm{d}x) \quad (k = 1, \ldots, n)\,.$$

Insbesondere ist $\mu_k = \mathrm{E}\, X^k = (-\mathrm{i})^k \varphi_X^{(k)}(0)$.

Beweis. Wegen $L^n(P) \subset L^k(P)$ für $k \leq n$ folgt $\mathrm{E}\,|X|^k < \infty$ für $k = 1, \ldots, n$. Betrachte

$$\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = \mathrm{E}\left[\mathrm{e}^{\mathrm{i}tx}\Big(\frac{\mathrm{e}^{\mathrm{i}hX} - 1}{h}\Big)\right].$$

Wegen $|\frac{\mathrm{e}^{\mathrm{i}hx}-1}{h}| \leq |x|$ für alle $x \in \mathbb{R}$ und $\mathrm{E}\,|X| < \infty$ folgt mit majorisierter Konvergenz

$$\lim_{h \to 0} \mathrm{E}\left[\mathrm{e}^{\mathrm{i}tX}\Big(\frac{\mathrm{e}^{\mathrm{i}tX} - 1}{h}\Big)\right] = \mathrm{E}(\mathrm{i}X\mathrm{e}^{\mathrm{i}tX}) = \mathrm{i}\int_{\mathbb{R}} x\mathrm{e}^{\mathrm{i}tx}(P \circ X^{-1})(\mathrm{d}x)\,.$$

Dies zeigt die Behauptung für $k = 1$. Der Rest folgt mit Induktion. □

Definition 10.10. Seien $(X_n)_{n\in\mathbb{N}}$ und X Zufallsvariable. Dann konvergiert $X_n \to X$ schwach, falls $\varphi_{X_n}(t) \to \varphi_X(t)$ gilt für alle $t \in \mathbb{R}$.

Dies bedeutet in der Sprechweise von Kapitel 6, daß $X_n \to X$ schwach bezüglich $\{x \mapsto \mathrm{e}^{\mathrm{i}xt} : t \in \mathbb{R}\}$ konvergiert. Bei allen betrachteten Konvergenzarten spricht man auch von der entsprechenden Konvergenz der zugehörigen Maße ($P \circ X_n^{-1} \to P \circ X^{-1}$ schwach etc.).

Nach Satz 6.10 gilt $X_n \to X$ schwach genau dann, wenn $X_n \to X$ schwach bezüglich $\mathscr{D}(\mathbb{R})$. Dies ist wiederum äquivalent zu $X_n \to X$ in Verteilung.

Der folgende Satz, einer der wichtigsten Sätze der Wahrscheinlichkeitstheorie, wird auch der Satz von de Moivre-Laplace genannt.

Satz 10.11 (Zentraler Grenzwertsatz für i.i.d.-Folgen). *Sei $(X_n)_{n\in\mathbb{N}}$ eine i.i.d.-Folge mit $X_1 \in L^2(P)$ nicht konstant. Dann genügt $(X_n)_n$ dem zentralen Grenzwertsatz, d.h. für die standardisierten Partialsummen*

$$S_n^* := \frac{S_n - \mathrm{E}\,S_n}{\sqrt{\operatorname{Var} S_n}} \ \Big(= \frac{S_n - n\,\mathrm{E}\,X_1}{\sqrt{n \operatorname{Var} X_1}}\Big)$$

gilt

$$P \circ (S_n^*)^{-1} \to N(0,1) \quad \text{ in Verteilung für } n \to \infty\,.$$

Damit folgt

$$P \circ (S_n^*)^{-1}(a,b] \to N(0,1)(a,b] = \frac{1}{\sqrt{2\pi}} \int_a^b \mathrm{e}^{-x^2/2}\mathrm{d}x \quad \text{ für alle } a < b\,.$$

Beweis. Wegen $S_n - \mathrm{E}\,S_n = \sum_{k=1}^n (X_k - \mathrm{E}\,X_k)$ sei ohne Einschränkung $\mathrm{E}\,X_k = 0$. Setze $\sigma^2 := \operatorname{Var} X_1$ und $\mu := P \circ X_1^{-1}$. Nach Satz 10.9 ist $\hat{\mu} \in C^2(\mathbb{R})$ mit $\hat{\mu}(0) = 1$, $\hat{\mu}'(0) = 0$ und $\hat{\mu}''(0) = \sigma^2$. Die Taylorreihe für $\hat{\mu}$ an der Stelle 0 ergibt somit

$$\hat{\mu}(t) = 1 - \frac{t^2}{2}\sigma^2 + r(t) \quad \text{mit } \frac{|r(t)|}{t^2} \to 0 \ (t \to 0)\,.$$

Nach Satz 7.9 und Lemma 10.3 b) ist

$$(P \circ S_n^{-1})\hat{}(t) = (\mu * \ldots * \mu)\hat{}(t) = (\hat{\mu}(t))^n\,,$$

also

$$\begin{aligned}
[P \circ (S_n^*)^{-1}]\hat{}(t) &= \Big[\hat{\mu}\Big(\frac{t}{\sqrt{n}\sigma}\Big)\Big]^n \\
&= \Big(1 - \frac{\sigma^2}{2}\,\frac{t^2}{n\sigma^2} + r\Big(\frac{t}{\sqrt{n}\sigma}\Big)\Big)^n \\
&\to \mathrm{e}^{-t^2/2} \quad \text{für } n \to \infty\,.
\end{aligned}$$

Dabei wurde $nr(\frac{t}{\sqrt{n}\sigma}) \to 0$ und die Tatsache verwendet, daß für eine komplexe Zahlenfolge z_n mit $z_n \to z$ gilt: $(1 + \frac{z_n}{n})^n \to \mathrm{e}^z$. Damit konvergiert $P \circ (S_n^*)^{-1}$ schwach gegen $N(0,1)$, und nach Satz 6.10 auch in Verteilung. □

Bemerkung 10.12. Man kann (unter zusätzlichen Bedingungen) die Konvergenz von $F_{S_n^*}(t)$ gegen $\varphi(t)$ qualitativ abschätzen. Es gilt der Satz von Berry-Esséen:

$$\sup_{t\in\mathbb{R}} |F_{S_n^*}(t) - \varphi(t)| \le \frac{6}{\sqrt{n}\sqrt{\operatorname{Var} X}^3}\, \mathrm{E}(|X_1 - \mathrm{E}\,X_1|^3)\,,$$

falls $X_1 \in L^3(P)$ und ansonsten die Situation von Satz 10.11 vorliegt.

Es zeigt sich, daß die Schranke aus dem Satz von Berry-Esséen in vielen Fällen viel zu grob ist. Wie man an den Abbildungen auf den nächsten Seiten sehen kann, konvergiert die Verteilung relativ schnell gegen die Normalverteilung. In der Praxis verwendet man oft ab einer Wiederholung der Länge 20–30 als Approximation die Normalverteilung.

In den Abbildungen **??**–**??** ist die Ausgangsverteilung eine Gleichverteilung. Die Verteilung bei zweifacher Wiederholung besitzt als Dichte eine Dreiecksform, und bereits diese wird durch die Dichte der Standard-Normalverteilung überraschend gut approximiert (Abbildung **??**). Auf den folgenden Abbildungen **??** und **??** sieht man die relativ schnelle Konvergenz (der Dichten).

Geht man von einer diskreten, eventuell auch noch unsymmetrisch verteilten Zufallsvariablen aus, so ist die Konvergenz aus dem Zentralen Grenzwertsatz naturgemäß langsamer. Dennoch sieht man in den Abbildungen **??**–**??**, daß die Standard-Normalverteilung als eine erste Näherung sinnvoll sein kann. In den Abbildungen ist die Ausgangsdichte die Bernoulli-Verteilung (Binomialverteilung $B(n, p)$ mit $n = 1$) mit Parameter $p = 0.6$. Man beachte, wie die ursprüngliche Asymmetrie durch die Faltung mehr und mehr verschwindet (die Grenzverteilung ist ja selbst symmetrisch).

11. Parameter-Punktschätzung

Bis jetzt war immer ein W-Feld $(\Omega, \mathscr{A}, P)$ gemeinsam mit einer Zufallsvariablen X gegeben. Aber in Anwendungen kennt man das W-Feld nicht, man beobachtet lediglich den Wert von X. Dies führt zu Fragestellungen der Statistik. Hier hat man folgende Situation:

Sei $(\Omega, \mathscr{A})$ ein Meßraum und $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ eine Familie von Maßen. Das wahre Maß P_0 ist nicht bekannt; es wird aber stets

$$P_0 \in \mathscr{P}$$

vorausgesetzt, d.h. es gilt $P_0 = P_{\theta_0}$ mit einem zu suchenden Parameter $\theta_0 \in \Theta$. Für eine Zufallsvariable $X_0 : \Omega \to \mathbb{R}$ hat man die Möglichkeit, eine n-fache unabhängige Wiederholung $(X_1, \dots, X_n)$ von X_0 zu beobachten, d.h. $(X_1, \dots, X_n)$ ist i.i.d.-Folge mit (unbekannter) Verteilung $P_0 \circ X_j^{-1} = P_0 \circ X_0^{-1}$. Aufgrund des beobachteten Wertes $(x_1, \dots, x_n) \in \mathbb{R}^n$ (einer Realisierung von $(X_1, \dots, X_n)$) will man Informationen über P_0 gewinnen, genauer: über $P_0 \circ X_0^{-1}$.

Im folgenden sei $P^X := P \circ X^{-1}$ das Bildmaß und $\mathscr{P}^X := \{P^X : P \in \mathscr{P}\}$.

Beispiel 11.1. Gesucht ist die mittlere Körpergröße von Studenten in cm, d.h. der Erwartungswert der Zufallsvariable X_0 = Körpergröße. Es werden 5 Personen gemessen mit dem Ergebnis $(x_1, \dots, x_5) = (175, 170, 165, 180, 185)$. Gesucht ist

$$\mathrm{E}\, X_0 = \int_\Omega X_0 \mathrm{d}P_0 = \int_{\mathbb{R}} x {P_0}^{X_0}(\mathrm{d}x)\,.$$

Dabei ist die Verteilung ${P_0}^{X_0}$ unbekannt. Der Erwartungswert ist ein Parameter (von vielen) des unbekannten Maßes; allgemein hat man eine „Kenngröße“ $\xi\colon \Theta \to \mathbb{R}$, welche jedem Element $P_\theta^{X_0}$ von $\mathscr{P}^{X_0}$ einen Parameter $\xi(P_\theta^{X_0}) = \xi(\theta)$ zuweist. In unserem Beispiel ist

$$\xi(\theta) = \mathrm{E}_\theta\, X_0 = \int \mathrm{id}_{\mathbb{R}}\, \mathrm{d}P_\theta^{X_0}\,.$$

Die Wahl von $\mathscr{P}$ hängt von der Vorinformation ab; ohne zusätzliche Information wird man

$$\mathscr{P} := \{P : P \text{ W-Maß auf } (\Omega, \mathscr{A}) \text{ mit } X_0 \in L^1(P)\}$$

wählen. Im wesentlichen gibt es drei interessante Fragen bzw. Typen von Antworten, die an unserem Beispiel erläutert werden sollen:

- Parameter-Punktschätzung: Hier wird ein Schätzwert für $\xi(\theta) = \mathrm{E}_\theta\, X_0$ angegeben, d.h. man hat eine Abbildung $T\colon \mathbb{R}^n \to \mathbb{R}$ und gibt als Antwort bei Beobachtung von $(x_1, \dots, x_n) \in \mathbb{R}^n$ den Wert $T(x_1, \dots, x_n)$ an.

Als Beispiel für einen Parameter-Punktschätzer sei etwa $T(x_1, \dots, x_n) := \frac{1}{n}(x_1 + \dots + x_n)$ erwähnt (das Stichprobenmittel), das bei uns den Wert 175 ergibt.

Man beachte, daß etwa bei $x = (175, 170, 165, 1800, 185)$ ein unsinniger Wert als Antwort herauskommt; es gibt andere Schätzer und Methoden, bei denen man auch solche offensichtliche Druckfehler mit berücksichtigen kann (etwa durch Ignorieren des größten und des kleinsten Wertes oder bei Verwendung des Medians).

- Konfidenzintervalle: Hier wird ein Intervall angegeben, in dem die gesuchte Kenngröße $\xi(\theta)$ vermutlich liegt. Eine typische Antworten ist dabei

  Mit Wahrscheinlichkeit 95% liegt $\mathrm{E}_{P_0} X_0$ im Intervall $[171, 179]$.

  Hier hat man eine Abbildung $(T_1, T_2)\colon \mathbb{R}^n \to \mathbb{R}^2$.

- Hypothesentests: Hier wird die Gültigkeit einer Hypothese untersucht, etwa der Hypothese $H_0\colon\ \mathrm{E}_{\theta_0} \leq 170$. Man ist in der Praxis daran interessiert, die Hypothese mit hinreichend großer Wahrscheinlichkeit zu verwerfen (Signifikanztest). Mögliche Antworten sind hier d_1: „H_0 wird abgelehnt" und d_2: „H_0 wird nicht abgelehnt". Somit hat man eine Abbildung $D\colon \mathbb{R}^n \to \{d_1, d_2\}$.

Die entscheidende Frage ist jeweils: Welche Abbildung T / (T_1, T_2) / D soll man wählen?

Definition 11.2. Gegeben sei ein Meßraum $(\Omega, \mathscr{A})$, eine Familie $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ von W-Maßen auf $(\Omega, \mathscr{A})$, eine Zufallsvariable $X_0\colon \Omega \to \mathbb{R}$ und eine zu schätzende Kenngröße $\xi : \Theta \to \mathbb{R}$.

a) Eine i.i.d.-Folge $X = (X_1, \dots, X_n)$ mit $P \circ X_i^{-1} = P \circ X_0^{-1}$ heißt Stichprobe vom Umfang n.

b) Eine Abbildung $T_n : \mathbb{R}^n \to \mathbb{R}$ heißt (Parameter-Punkt-)Schätzer (auch Statistik), falls T_n Borel-meßbar ist. $T_n \circ X$ heißt Schätzvariable. Die Abbildung T_n ist so zu interpretieren, daß bei Beobachtung einer Realisierung $(x_1, \dots, x_n)$ von $(X_1, \dots, X_n)$ die Antwort $T_n(x_1, \dots, x_n)$ als Schätzwert für die Kenngröße $\xi(\theta_0)$ gewählt wird.

c) Sei $T_n \circ X \in L^1(P_\theta)$ für alle $\theta \in \Theta$. Dann heißt der Schätzer T_n und die Schätzvariable $T_n \circ X$ erwartungstreu für $\xi(\theta)$, falls gilt

$$\mathrm{E}_\theta(T_n \circ X) \left[= \int_\Omega T_n \circ X \,\mathrm{d}P_\theta\right] = \xi(\theta) \qquad \text{für alle } \theta \in \Theta\,.$$

Im folgenden sei stets ein Schätzproblem, d.h. $(\Omega, \mathscr{A})$, $\mathscr{P}$, ξ und X_0 wie oben, gegeben.

Bemerkung 11.3 (Kanonische Wahl des W-Raumes). Die gemeinsame Verteilung einer Stichprobe $X = (X_1, \dots, X_n)$ ist das Produktmaß, d.h. man erhält den W-Raum

$$(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), (P_\theta \circ X_0^{-1})^{\otimes n}),$$

falls das Maß P_θ zugrundeliegt. Dabei wurde die abkürzende Schreibweise

$$(P_\theta \circ X_0^{-1})^{\otimes n} := (P_\theta \circ X_0^{-1}) \otimes \ldots \otimes (P_\theta \circ X_0^{-1})$$

(n Faktoren) verwendet. Geht man andererseits von diesem W-Raum aus und wählt die Koordinatenprojektionen $X_k(x) := x_k$ für $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ als Zufallsvariablen, so ist $(X_1, \dots, X_n)$ i.i.d. und wie X_0 verteilt, also eine Stichprobe vom Umfang n zu X_0.

Satz 11.4. a) *Sei $X_0 \in L^1(P_\theta)$ für alle $\theta \in \Theta$. Dann ist das Stichprobenmittel*

$$\overline{X}_{(n)} := \frac{1}{n} \sum_{i=1}^{n} X_i$$

eine erwartungstreue Schätzvariable für $\xi(\theta) = \mathrm{E}_\theta\, X_0$.

b) *Sei $X_0 \in L^2(P_\theta)$ für alle $\theta \in \Theta$. Dann ist die Stichprobenvarianz*

$$s_{(n)}^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_{(n)})^2$$

eine erwartungstreue Schätzvariable für $\xi(\theta) = \mathrm{Var}_\theta\, X_0$.

Beweis. a) Es gilt

$$\mathrm{E}_\theta\, \overline{X}_{(n)} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}_\theta\, X_i = \mathrm{E}_\theta\, X_0\,.$$

b) Sei $\mu_\theta := \mathrm{E}_\theta\, X_0$. Dann gilt

$$\begin{aligned}
\mathrm{E}_\theta\, s_{(n)}^2 &= \frac{1}{n-1}\, \mathrm{E}_\theta \Big(\sum_{i=1}^{n} (X_i - \overline{X}_{(n)})^2 \Big) \\
&= \frac{1}{n-1}\, \mathrm{E}_\theta \Big(\sum_{i=1}^{n} [(X_i - \mu_\theta) - (\overline{X}_{(n)} - \mu_\theta)]^2 \Big) \\
&= \frac{1}{n-1}\, \mathrm{E}_\theta \Big(\sum_{i=1}^{n} (X_i - \mu_\theta)^2 + \sum_{i=1}^{n} (\overline{X}_{(n)} - \mu_\theta)^2 \\
&\qquad - 2 \sum_{i=1}^{n} (X_i - \mu_\theta)(\overline{X}_{(n)} - \mu_\theta) \Big)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n-1}\,\mathrm{E}_\theta\Big(\sum_{i=1}^n (X_i-\mu_\theta)^2 + \sum_{i=1}^n (\overline{X}_{(n)}-\mu_\theta)^2 \\
&\qquad - 2\sum_{i,j=1}^n (X_i-\mu_\theta)(X_j-\mu_\theta)\Big) \\
&= \frac{1}{n-1}\Big(n\operatorname{Var}_\theta X_0 + n\operatorname{Var}_\theta \overline{X}_{(n)} - \frac{2}{n}\sum_{i,j=1}^n \operatorname{Cov}_\theta(X_i,X_j)\Big) \\
&= \frac{1}{n-1}\Big(n+1-\frac{2}{n}\cdot n\Big)\operatorname{Var}_\theta X_0 \\
&= \operatorname{Var}_\theta X_0\,.
\end{aligned}
$$

Dabei wurde verwendet, daß $\operatorname{Var}_\theta \overline{X}_{(n)} = \frac{1}{n}\operatorname{Var}_\theta X_0$ und $\operatorname{Cov}_\theta(X_i,X_j) = \delta_{ij}\operatorname{Var}_\theta X_0$ gilt. □

Definition 11.5. Sei $\mathscr{P} = \{P_\theta : \theta\in\Theta\}$, $\xi\colon \Theta\to\mathbb{R}$ und $T\colon \mathbb{R}^n\to\mathbb{R}$ ein Parameter-Punktschätzer für die Kenngröße $\xi(\theta)$.

a) Die Risikofunktion $R_T\colon \Theta\to\overline{\mathbb{R}}$ ist definiert durch

$$R_T(\theta) := \mathrm{E}_\theta(T\circ X-\xi(\theta))^2\ \Big(=\operatorname{Var}_\theta T\circ X\quad \text{falls } T \text{ erwartungstreu}\Big).$$

b) Sei $\mathscr{T}$ eine Familie von erwartungstreuen Schätzern. Ein Schätzer T^* heißt Minimum-Varianz-Schätzer (oder simultan bester Schätzer) in $\mathscr{T}$, falls T^* erwartungstreu ist und

$$R_{T^*}(\theta)\le R_T(\theta)\quad \text{für alle } \theta\in\Theta \text{ und } T\in\mathscr{T}$$

gilt.

c) T^* heißt Minimax-Schätzer in $\mathscr{T}$, falls

$$\sup_{\theta\in\Theta} R_{T^*}(\theta) = \min_{T\in\mathscr{T}}\sup_{\theta\in\Theta} R_T(\theta)\,.$$

Beispiel 11.6. Sei $\xi(\theta) = \mathrm{E}_\theta X_0$ und

$$\mathscr{T} = \{T\colon \mathbb{R}^n\to\mathbb{R}\ \big|\ T \text{ linear, erwartungstreu}\}\,.$$

Für $T\in\mathscr{T}$ gilt $T(x)=\sum_{i=1}^n a_ix_i$ mit $\sum_{i=1}^n a_i = 1$ (wegen Linearität und Erwartungstreue). Somit erhält man

$$
\begin{aligned}
R_T(\theta) &= \mathrm{E}_\theta(T\circ X-\mathrm{E}_\theta X_0)^2 \\
&= \mathrm{E}_\theta\Big(\sum_{i=1}^n a_i(X_i-\mathrm{E}_\theta X_0)\Big)^2
\end{aligned}
$$

$$= \Big(\sum_{i=1}^{n} a_i^2\Big) \operatorname{Var}_\theta X_0 \,.$$

Damit ist $R_T(\theta)$ genau dann für alle $\theta \in \Theta$ minimal, wenn $\sum_{i=1}^n a_i^2$ minimal ist, was äquivalent ist zu $a_i = \frac{1}{n}$ für alle $i = 1, \dots, n$. Somit ist der (einzige) Minimum-Varianz-Schätzer in $\mathscr{T}$ das Stichprobenmittel $T = \overline{X}_{(n)}$.

Bemerkung 11.7. a) Die Risikofunktion ist von Bedeutung, da etwa (für erwartungsteues T)

$$P_\theta\{|T \circ X - \xi(\theta)| \geq \delta\} \leq \frac{1}{\delta^2} R_T(\theta)$$

gilt (wegen Chebyshev). Es gilt für $T = \overline{X}_{(n)}$

$$R_{\overline{X}_{(n)}}(\theta) = \operatorname{Var}_\theta \overline{X}_{(n)} = \frac{1}{n} \operatorname{Var}_\theta X_0 \,,$$

d.h. $R_{\overline{X}_{(n)}}$ nimmt mit wachsendem Stichprobenumfang ab.

b) Statt $R_T(\theta) = \mathrm{E}_\theta(T \circ X - \xi(\theta))^2$ verwendet man auch $\mathrm{E}_\theta \, |T \circ X - \xi(\theta)|^k$, insbesondere für $k = 1$.

Um Schätzer zu konstruieren, kann auch der Maximum-Likelihood-Ansatz verwendet werden, der zunächst an einem Beispiel erläutert werden soll.

Beispiel 11.8. In einer Urne befinden sich sechs schwarze und weiße Kugeln, wobei die Zahl k der schwarzen Kugeln unbekannt ist. Bei einer Stichprobe werden drei Kugeln entnommen, davon sind zwei schwarz. Zu schätzen ist k.

Ein möglicher Ansatz besteht darin, für $P = P_k \in \mathscr{P} := \{H(3; k; 6) : k \in 0, \dots, 6\}$ die Wahrscheinlichkeit $P_k\{X = 2\}$ zu bestimmen, wobei X die Anzahl der gezogenen schwarzen Kugeln bezeichne. Dabei ist H die aus 3.4 bekannte hypergeometrische Verteilung. Nach Satz 3.4 a) ist

$$P_k\{X = 2\} = \frac{\binom{k}{2}\binom{6-k}{1}}{\binom{6}{3}} = \frac{k(k-1)(6-k)}{2 \cdot 4 \cdot 5} \,.$$

Diese Werte sind in der folgenden Tabelle ersichtlich:

k	0	1	2	3	4	5	6
$P_k\{X = 2\}$	0	0	0.2	0.45	0.6	0.5	0

Als Schätzwert für k wird nun diejenige Zahl genommen, für welche das beobachtete Ergebnis die größte Wahrscheinlichkeit hat (in unserem Beispiel ergibt sich als Schätzwert für k der Wert 4).

Definition 11.9. Sei $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ mit $\Theta \subset \mathbb{R}$ und $\xi(\theta) = \theta$. Sei außerdem jedes $P_\theta^{X_0}$ stetig verteilt mit Dichte f_θ bzgl. des Lebesgue-Maßes oder jedes $P_\theta^{X_0}$ diskret verteilt mit Dichte f_θ bzgl. eines diskreten Maßes (d.h. es gilt entweder $P_\theta^{X_0} = f_\theta(x)\mathrm{d}x$ für alle $\theta \in \Theta$ oder alle $P_\theta^{X_0}$ sind diskret und auf derselben abzählbaren Menge konzentriert).

a) Die Abbildung $L\colon \Theta \times \mathbb{R}^n \to \mathbb{R}$, definiert durch

$$L(\theta, x_1, \dots, x_n) := f_\theta(x_1) \cdot \ldots \cdot f_\theta(x_n)\,,$$

heißt Likelihood-Funktion.

b) Ein Schätzer $T^*\colon \mathbb{R}^n \to \Theta$ heißt Maximum-Likelihood-Schätzer, falls gilt

$$L(T^*(x), x) = \max_{\theta\in\Theta} L(\theta, x) \qquad \text{für alle } x \in \mathbb{R}^n\,.$$

Beispiel 11.10. Sei $\mathscr{P}^{X_0} = \{\frac{1}{\theta}1_{[0,\theta]}(x)\mathrm{d}x : \theta \in (0,\infty)\}$ (Menge der Gleichverteilungen auf dem Intervall $[0,\theta]$). Dann ist

$$\begin{aligned} L(\theta, x) &= \begin{cases} (\frac{1}{\theta})^n\,, & \text{falls } x \in [0,\theta]^n\,, \\ 0\,, & \text{sonst,} \end{cases} \\ &= \begin{cases} (\frac{1}{\theta})^n\,, & \text{falls } \theta \geq \max\{x_1, \dots, x_n\} \text{ und } x_i \geq 0\,, \\ 0\,, & \text{sonst.} \end{cases} \end{aligned}$$

Da die Abbildung $\theta \mapsto (\frac{1}{\theta})^n$ streng monoton fallend auf $\Theta = (0,\infty)$ ist, erhält man für festes $x \in \mathbb{R}^n$: Es gilt

$$L(\hat\theta, x) = \max_{\theta\in\Theta} L(\theta, x)$$

genau dann, wenn

$$\hat\theta = \max\{x_1, \dots, x_n\}\,.$$

Also ist $T^*(x) = \hat\theta(x) = \max\{x_1, \dots, x_n\}$ der einzige Maximum-Likelihood-Schätzer. Man beachte, daß T^* nicht erwartungstreu ist.

Im allgemeinen existiert kein Maximum-Likelihood-Schätzer, oder er ist nicht mit dem Minimum-Varianz-Schätzer identisch. Aber für wichtige Klassen von Verteilungen ist dies doch der Fall.

Definition 11.11. Eine Familie von W-Maßen $\{Q_\theta : \theta \in \Theta\}$ auf einem Meßraum $(\Omega, \mathscr{A})$ heißt eine k-parametrige Exponentialfamilie, falls

$$Q_\theta = c_\theta \exp(\langle C(\theta), S(x)\rangle)\; \mu(\mathrm{d}x)$$

gilt, wobei $C\colon \Theta \to \mathbb{R}^k$ eine Abbildung, $S\colon \Omega \to \mathbb{R}^k$ meßbar, $\langle\cdot,\cdot\rangle$ das Standard-Skalarprodukt im $\mathbb{R}^k$, μ ein σ-endliches Maß auf $(\Omega, \mathscr{A})$ und

$$c_\theta := \left(\int \exp(\langle C(\theta), S(x)\rangle)\mu(\mathrm{d}x)\right)^{-1}$$

die Normierungskonstante ist.

Bemerkung 11.12. a) Aus $Q_\theta \neq Q_{\theta'}$ folgt auch $C(\theta) \neq C(\theta')$.

b) Falls $\{Q_\theta : \theta \in \Theta\}$ eine k-parametrige Exponentialfamilie auf $(\Omega, \mathscr{A})$ ist, dann ist das n-fache Produktmaß $(Q_\theta)^{\otimes n}$ eine k-parametrige Exponentialfamilie auf dem W-Raum $(\Omega^n, \mathscr{A}^{\otimes n})$, denn es gilt

$$(Q_\theta)^{\otimes n} = (c_\theta)^n \exp(\langle nC(\theta), \overline{S}(x)\rangle)\, \mu^{\otimes n}(\mathrm{d}x)\,, \quad x = (x_1, \ldots, x_n)\,,$$

wobei $\overline{S}(x) := \frac{1}{n}\sum_{i=1}^n S(x_i)$ gesetzt wurde.

Beispiele 11.13. a) Es sei $\sigma > 0$ fest und $\theta = \alpha \in \Theta = \mathbb{R}$. Dann ist

$$N(\alpha, \sigma^2) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\alpha^2}{2\sigma^2}\right)}_{c_\theta} \cdot \exp\left(\frac{\alpha}{\sigma^2}\cdot x\right) \cdot \underbrace{\exp\left(-\frac{x^2}{2\sigma^2}\right)\mathrm{d}x}_{\mu}$$

eine einparametrige Exponentialfamilie mit $C(\theta) = \frac{\theta}{\sigma^2}$ und $S(x) = x$.

b) Nun sei α und σ unbekannt, d.h. es sei $\theta = (\alpha, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$. Dann ist

$$N(\alpha, \sigma^2) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\alpha^2}{2\sigma^2}\right)}_{c_\theta} \cdot \exp\left(\frac{\alpha}{\sigma^2}\cdot x - \frac{1}{2\sigma^2}\cdot x^2\right) \cdot \underbrace{\mathrm{d}x}_{\mu}$$

eine 2-parametrige Exponentialfamilie mit

$$C(\theta) = \begin{pmatrix} \alpha/\sigma^2 \\ -1/(2\sigma^2) \end{pmatrix} \text{ und } S(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix}.$$

c) Die Exponentialverteilung mit Parmeter $\theta = \lambda$ hat die Verteilung

$$\underbrace{\lambda}_{c_\theta} \cdot \exp(-\lambda\cdot x) \cdot \underbrace{1_{\mathbb{R}_+}\mathrm{d}x}_{\mu}\,,$$

ist also eine einparametrige Exponentialfamilie mit $C(\theta) = -\theta$ und $S(x) = x$.

d) Die Poisson-Verteilung mit Parameter $\theta = \lambda \in (0, \infty)$ hat die Form

$$\pi_\lambda = \mathrm{e}^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\delta_n = \mathrm{e}^{-\lambda}\lambda^x\mu(\mathrm{d}x) = \mathrm{e}^{-\lambda}\mathrm{e}^{\ln\lambda\cdot x}\mu(\mathrm{d}x)\,,$$

wobei $\mu := \sum_{n=0}^{\infty}\frac{1}{n!}\delta_n$. Somit ist π_λ eine einparametrige Exponentialfamilie mit $C(\theta) = \ln\theta$ und $S(x) = x$.

Das folgende Lemma wird hier ohne Beweis angegeben.

Lemma 11.14. *Sei* $\{Q_\theta : \theta \in \Theta\}$ *eine einparametrige Exponentialfamilie mit den Bezeichnungen aus Definition* 11.11. *Sei* $\Theta \subset \mathbb{R}$ *ein offenes Intervall und* $C\colon \Theta \to \mathbb{R}$ *stetig differenzierbar mit* $C'(\theta) \neq 0$ *für alle* $\theta \in \Theta$ *(vgl. dazu auch Bemerkung* 11.12 a)*. Dann gilt* $S \in L^1(Q_\theta)$ *für alle* $\theta \in \Theta$ *und*

a) $\frac{\mathrm{d}}{\mathrm{d}\theta} \ln c_\theta = -C'(\theta)\,\mathrm{E}_\theta\, S$.

b) *Sei* $f \in L^1(Q_\theta)$ *für alle* $\theta \in \Theta$. *Dann gilt*

$$\frac{\mathrm{d}}{\mathrm{d}\theta} = C'(\theta)\,\mathrm{E}_\theta[(S - \mathrm{E}_\theta\, S) \cdot f]\,.$$

Satz 11.15. *Sei die Voraussetzung von Lemma* 11.14 *erfüllt mit* $Q_\theta = P_\theta \circ X_0^{-1}$. *Dann gilt:*

a) $\overline{S}$ *ist Minimum-Varianz-Schätzer für* $\xi(\theta) = \mathrm{E}_\theta(S \circ X_0)$.

b) *Falls* μ *absolutstetig bzgl. des Lebesgue-Maßes oder eines diskreten Maßes ist und* T^* *ein Maximum-Likelihood-Schätzer für* θ *ist, so gilt* $\overline{S}(X) = \xi(T^*(X))$.

Beweis. a) Nach Bemerkung 11.12 b) sei o.E. $N = 1$. Es sei $T\colon \mathbb{R} \to \mathbb{R}$ mit $\mathrm{E}_\theta\, T \circ X = \theta$ und $T \circ X \in L^2(P_\theta)$ für alle $\theta \in \Theta$. Dann gilt nach Lemma 11.14 b)

$$\begin{aligned} \xi'(\theta) &= \frac{\mathrm{d}}{\mathrm{d}\theta}\,\mathrm{E}_\theta\, T = C'(\theta)\,\mathrm{E}_\theta[(S - \mathrm{E}_\theta\, S) \cdot T] \\ &= C'(\theta)\,\mathrm{Cov}_\theta(S, T)\,. \end{aligned}$$

Dabei wurde bei der letzten Gleichheit verwendet, daß $\mathrm{E}_\theta[(S - \mathrm{E}_\theta\, S)\,\mathrm{E}_\theta\, T] = 0$ und damit

$$\mathrm{E}_\theta[(S - \mathrm{E}_\theta\, S) \cdot T] = \mathrm{Cov}_\theta(S, T)$$

gilt. Somit ist also

$$|\xi'(\theta)|^2 \leq C'(\theta)^2\,\mathrm{Var}_\theta\, S \cdot \mathrm{Var}_\theta\, T\,.$$

Andererseits gilt auch nach obiger Rechnung (mit $T = S$) die Gleichheit

$$\xi'(\theta) = C'(\theta)\,\mathrm{Var}_\theta\, S\,,$$

also erhält man insgesamt

$$\mathrm{Var}_\theta\, S \geq \mathrm{Var}_\theta\, S\,,$$

was zeigt, daß S ein Minimum-Varianz-Schätzer ist.

b) Nach Definition des Maximum-Likelihood-Schätzers gilt

$$L^*(T(x), x) = \max_{\theta \in \Theta} L(\theta, x)$$

für alle $x \in \mathbb{R}^n$, wobei

$$L(\theta, x) = f_\theta(x_1) \cdot \ldots \cdot f_\theta(x_n) = (c_\theta)^n \exp(nC(\theta)\overline{S}(x)) .$$

Daher ist auch

$$\ln L(T^*(x), x) = \max_{\theta \in \Theta} \ln L(\theta, x) ,$$

d.h. es gilt

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}\theta} \ln L(\theta, x)|_{\theta = T^*(x)} = \frac{\mathrm{d}}{\mathrm{d}\theta} \Big[n(\ln c_\theta + C(\theta)\overline{S}(x) \Big] \Big|_{\theta = T^*(x)} \\ &= n \Big[- C'(\theta) \operatorname{E}_\theta S + C'(\theta)\overline{S}(x) \Big] \Big|_{\theta = T^*(x)} \end{aligned}$$

nach Lemma 11.14 a). Da nach Voraussetzung $C'(\theta) \neq 0$ war, folgt somit

$$\overline{S}(x) = \operatorname{E}_\theta S|_{\theta = T^*(x)} = \xi(T^*(x)) .$$

□

Beispiele 11.16. Angewendet auf die Beispiele 11.13 erhält man mit der dortigen Bezeichnung folgende Aussagen:

a) Hier ist $S(x) = x$, d.h. es ist $\xi(\theta) = \operatorname{E}_\theta X_0$. Ein Minimum-Varianz-Schätzer und der einzige Kandidat für einen Maximum-Likelihood-Schätzer für $\xi(\theta) = \theta$ ist somit das Stichprobenmittel $\overline{X}_{(n)}$.

b) Hier liefert der Satz keine Aussage, da er sich nur auf einparametrige Exponentialfamilien bezieht.

c) Hier ist $S(x) = s$ und $\xi(\theta) = \operatorname{E}_\theta X_0 = \frac{1}{\theta}$. Damit ist $\overline{X}_{(n)}$ ein Minimum-Varianz-Schätzer für $\operatorname{E}_\theta X_0$, und $(\overline{X}_{(n)})^{-1}$ ist der einzige Kandidat für einen Maximum-Likelihood-Schätzer für θ.

d) genauso wie a).

12. Signifikanztests

Eine Punktschätzung ist eine Abbildung $\mathbb{R}^n \to \mathbb{R}$, d.h. man gibt aufgrund des beobachteten Ergebnisses einen Schätzwert für den unbekannten Parameter an. Bei Signifikanztests geht es um die Beurteilung einer Hypothese, etwa $H_0 : \theta_0 \leq 1$, wobei θ_0 der zum wahren W-Maß gehörige Parameter ist. Dabei sind zwei Antworten möglich:

d_1: Die Hypothese H_0 wird abgelehnt.

d_2: Die Hypothese H_0 wird nicht abgelehnt.

Die Antwort d_2 heißt nicht, daß H_0 als wahr angenommen wird! H_0 kann nur durch die beobachteten Daten nicht verworfen werden.

Definition 12.1. Sei $(\Omega, \mathscr{A})$ ein Meßraum, $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ eine Familie von Maßen auf $(\Omega, \mathscr{A})$, $X_0 \colon \Omega \to \mathbb{R}$ eine Zufallsvariable, P_{θ_0} das (unbekannte) wahre Maß und H_0 eine Hypothese über θ_0. (Bei uns ist meistens $H_0 \subset \Theta$ in dem Sinn, daß die Hypothese die Form $\theta_0 \in H_0 \subset \Theta$ hat.) Ein Signifikanztest ist eine meßbare Abbildung $D \colon \mathbb{R}^n \to \{d_1, d_2\}$ mit folgender Interpretation: Zu einer Realisierung $\hat{x} = (x_1, \dots, x_n)$ einer Stichprobe vom Umfang n wird H_0 abgelehnt, falls $D(\hat{x}) = d_1$, d.h. falls

$$\hat{x} \in K := D^{-1}(\{d_1\})\,;$$

sonst wird H_0 nicht abgelehnt. Die Menge $K \in \mathscr{B}(\mathbb{R}^n)$ heißt kritischer Bereich des Testes D.

Ab sofort sei stets die Situation von Definition 12.1 gegeben, wobei $H_0 \subset \Theta$ sei.

Die einzige Frage zur Konstruktion eines Tests D ist, wie der kritische Bereich K zu wählen ist. Wie bereits bei Parameter-Punktschätzern gibt es auch hier keine eindeutige Antwort.

Definition 12.2. Die Funktion

$$g \colon \Theta \to \mathbb{R}\,, \quad \theta \mapsto (P_\theta \circ X^{-1})(K) = P_\theta\{X \in K\}$$

heißt die Gütefunktion des Tests D.

Für $\theta \in H_0$ heißt $g(\theta)$ die Fehler-Wahrscheinlichkeit erster Art; für $\theta \in \Theta \backslash H_0$ heißt $1 - g(\theta) = P_\theta\{X \notin K\}$ die Fehler-Wahrscheinlichkeit zweiter Art.

Eine typische Gütefunktion (bei einem einseitigen Test) hat die in Abbildung **??** skizzierte Form.

Die Gütefunktion bestimmt die Qualität eines Tests praktisch vollständig. Man beachte, daß in der Anwendung die beiden Fehlerarten nicht gleichwertig sind. Man will eine Aussage „statistisch signifikant nachweisen“, indem man die Nullhypothese ablehnt. Die Nullhypothese verkörpert in gewisser Weise den „Normalzustand“, bei dem keine Besonderheiten auftreten. Als Beispiel einer typischen Nullhypothese sei etwa genannt: Die Klausurergebnisse der Teilnehmer der ersten Übungsgruppen unterscheiden sich im Durchschnitt nicht von denen der zweiten Übungsgruppe. Bei der Formulierung einer solchen Nullhypothese möchte man untersuchen, ob die zwei Übungsgruppen einen signifikanten Unterschied aufweisen, das Ziel ist es also, einen solchen Unterschied durch die erhobenen Daten (in diesem Fall die Klausurergebnisse) zu belegen.

Entscheidend ist dabei der Fehler erster Art, man will vor allem keine Fehlentscheidung in der Richtung treffen, daß die Nullhypothese zu Unrecht abgelehnt wird. Daher wird in der Praxis versucht, den Fehler erster Art unter einer gewissen Schranke zu halten, während der Fehler zweiter Art als Gütekriterium verwendet wird („Schärfe“ des Tests).

Definition 12.3. Sei $0 < \alpha < 1$. Ein Test D heißt Signifikanztest zum Signifikanzniveau α, falls für seine Gütefunktion gilt:

$$\sup_{\theta \in H_0} g(\theta) \leq \alpha \,.$$

Auf der linken Seite dieser Ungleichung steht das Supremum über die Fehler-Wahrscheinlichkeit erster Art. Üblich für Werte von α sind 0.1, 0.05 und 0.01, vor allem $\alpha = 0.05 = 5\%$.

Es bleibt die Frage, wie man einen solchen Signifikanztest D konstruiert. Dies soll an einem Beispiel erläutert werden:

Beispiel 12.4. (Einseitiger Gaußtest) Durch das Hören einer Vorlesung über Wahrscheinlichkeitstheorie soll der Intelligenzquotient der Studenten gesteigert werden. Zu Beginn der Vorlesung beträgt der durchschnittliche IQ 100, es soll „gezeigt“ werden, daß dieser Wert am Ende des Semesters auf mindestens 120 gestiegen ist. Ende des Semesters wird der IQ von 10 Studenten getestet, wobei sich ein durchschnittlicher Wert von 128 ergibt. Nun soll entschieden werden, ob die gewünschte Verbesserung auf 120 eingetreten ist.

Zuerst muß man eine mathematische Modellierung vornehmen. In diesem Beipsiel wird man den IQ als eine Zufallsvariable X_0 interpretieren, der gesuchte Parameter ist dann der durchschnittliche IQ, d.h. der Erwartungswert von X_0. Wir haben also $\theta = \mathrm{E}_\theta \, X_0$ mit $\Theta = (0, \infty)$.

(i) Als ersten Schritt muß man sich über die Verteilungsannahmen klarwerden, die im konkreten Fall getroffen werde sollen. Es sei hier etwa angenommen, daß X_0

normalverteilt ist, was aus dem Zentralen Grenzwertsatz eine gewisse Begründung erfährt (und in den meisten Fällen in der Praxis auch so gemacht wird). Dabei sei die Varianz von X_0 aus früheren Messungen bekannt mit $\operatorname{Var} X_0 = \sigma_0^2 = 400$. Weiterhin geht man davon aus, daß man eine Stichprobe im Sinn von Definition 11.2 vorliegen hat, insbesondere, daß die Messungen unabhängig sind. Unter diesen Annahmen erhält man

$$\mathscr{P}^{X_0} = \{N(\theta, \sigma_0^2) : \theta \in (0, \infty)\} = \{P_\theta : \theta \in \Theta\}\,.$$

(ii) Der nächste Schritt besteht darin, die Nullhypothese zu formulieren. Wie oben bereits erwähnt, nimmt man als Nullhypothese das Gegenteil dessen, was man nachweisen will. In unserem Fall ergibt sich $H_0 : \theta \leq 120$, d.h. $H_0 = (0, 120] \subset \Theta$.

(iii) Einen sinnvollen Test kann man konstruieren, indem man einen vernünftigen Punktschätzer für θ verwendet und damit den kritischen Bereich festlegt. Wie aus dem Abschnitt über Parameter-Punktschätzung bekannt ist, ist das Stichprobenmittel $T(X) = \overline{X}_{(n)}$ ein guter Punktschätzer für $\theta = \mathrm{E}_\theta\, X_0$. Große Werte von T werden zu einer Ablehnung von H_0 führen. Daher wählt man den Ablehnungsbereich K des Tests in der Form

$$K = \{x \in \mathbb{R}^n : T(x) > c\}\,,$$

wobei die Konstante c noch zu bestimmen ist. Für die weitere Rechnung benötigt man die Verteilung von $T \circ X = \overline{X}_{(n)}$. Nach Korollar 10.8 ist $X_1 + \cdots + X_n$ normalverteilt mit Erwartungswert $n\theta$ und Varianz $n\sigma_0^2$, falls X_0 $N(\theta, \sigma_0^2)$-verteilt ist. Daher ist $\overline{X}$ $N(\theta, \sigma_0^2/n)$-verteilt. Meistens verwendet man die Standardisierung

$$\frac{\overline{X}_{(n)} - \theta}{\sqrt{\frac{\sigma_0^2}{n}}}\,,$$

welche $N(0,1)$-verteilt ist.

(iv) Nun ist noch die Konstante c zu bestimmen. Dies soll so geschehen, daß

$$\sup_{\theta \in H_0} P_\theta\{X \in K\} = \alpha$$

gilt. Dabei ist „$\leq$“ notwendig, um einen Signifikanztest zum Niveau α zu erhalten. Man wird aber auch keinen kleineren Wert für das obige Supremum nehmen, da sonst der Fehler zweiter Art zu groß wird. Für $\theta \in H_0$ ist

$$\begin{aligned} P_\theta\{X \in K\} &= P_\theta\{T \circ X > c\} \\ &= P_\theta\left\{\frac{\overline{X}_{(n)} - \theta}{\sqrt{\frac{\sigma_0^2}{n}}} > \frac{c - \theta}{\sqrt{\frac{\sigma_0^2}{n}}}\right\} = 1 - \Phi\left(\frac{c - \theta}{\sqrt{\frac{\sigma_0^2}{n}}}\right)\,, \end{aligned}$$

wobei Φ die Verteilungsfunktion einer $N(0,1)$-verteilten Zufallsvariable bezeichne. Da der letzte Ausdruck als Funktion von θ streng monoton wachsend ist, gilt für

$H_0 = (0, \theta_0]$:

$$\sup_{\theta \in H_0} P_\theta\{X \in K\} = 1 - \Phi\left(\frac{c - \theta_0}{\sqrt{\frac{\sigma_0^2}{n}}}\right).$$

Dies ist gleich α genau dann, wenn

$$c = \theta_0 + \Phi^{-1}(1-\alpha)\sqrt{\frac{\sigma_0^2}{n}}.$$

Für $\theta_0 = 120,\ \sigma_0 = 20,\ n = 10$ und $\alpha = 0.05$ erhält man wegen $\Phi^{-1}(1-\alpha) = 1.648...$ den Wert $c = 130.4....$

(v) Somit erhält man folgende Antwort: Auf einem Signifikanzniveau von 5% kann die gewünschte Verbesserung des IQ auf einen Wert über 120 nicht nachgewiesen werden. (D.h., die Studenten sollen auch noch in die Fortsetzungsvorlesung kommen, um eine weitere Chance zu haben.)

Die in diesem Beispiel sichtbaren Schritte zur Konstruktion eines Signifikanztests werden im folgenden noch einmal zusammengefaßt:

12.5. Allgemeines Verfahren zur Konstruktion von Signifikanztests:

(i) Verteilungsannahmen.

(ii) Formulierung der Nullhypothese H_0.

(iii) Wahl der Testgröße T und Bestimmung ihrer Verteilung unter Annahme von H_0.

(iv) Bestimmung des kritischen Bereichs K zum Signifikanzniveau α.

(v) Entscheidung: Falls der beobachtete Wert in K liegt, wird H_0 abgelehnt, sonst wird gegen H_0 nichts eingewendet.

Dabei gehören die Punkte (i), (ii) zur Modellierung, und die eigentliche Rechnung steckt in den Punkten (iii) und (iv). Für Punkt (iii) definiert man sich folgende häufig auftretende Verteilungen:

Definition 12.6. Seien die Zufallsvariablen $(X_1, \dots, X_{n+1})$ i.i.d. $N(0,1)$-verteilt über einem W-Raum $(\Omega, \mathscr{A}, P)$.

a) Die Verteilung

$$\chi_n^2 := P \circ (X_1^2 + \dots + X_n^2)^{-1}$$

heißt die Chi-Quadrat-Verteilung mit n Freiheitsgraden.

b) Die Verteilung

$$\mathrm{t}_n := P \circ \left(\frac{X_{n+1}}{\sqrt{\frac{X_1^2 + \cdots + X_n^2}{n}}} \right)^{-1}$$

heißt die t-Verteilung mit n Freiheitsgraden.

c) Sei $r, s \in \{1, \ldots, n\}$ mit $r + s \leq n$. Dann heißt die Verteilung

$$F_{r,s} := P \circ \left(\frac{\frac{1}{r}(X_1^2 + \cdots + X_r^2)}{\frac{1}{s}(X_{r+1}^2 + \cdots + X_{r+s}^2)} \right)^{-1}$$

die F-Verteilung mit r und s Freiheitsgraden.

Satz 12.7 (Verteilung der wichtigsten Testgrößen). *Seien* $(X_1, \dots, X_n)$ *i.i.d.* $N(\mu, \sigma^2)$*-verteilt.*

a) *Das Stichprobenmittel* $\overline{X}_{(n)} = \frac{1}{n}\sum_{i=1}^n X_i$ *ist* $N(\mu, \sigma^2/n)$*-verteilt.*

b) *Sei* $s_{(n)}^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_{(n)})^2$ *die Stichprobenvarianz. Dann ist* $\frac{n-1}{\sigma^2} s_{(n)}^2$ χ_{n-1}^2*-verteilt.*

c) *Die Testgröße*

$$\frac{\overline{X}_{(n)} - \mu}{\frac{1}{\sqrt{n}} s_{(n)}}$$

ist t_{n-1}*-verteilt.*

(ohne Beweis)

Für die Bestimmung des kritischen Bereichs K im Punkt 12.5 (iv) braucht man „Umkehrfunktionen“ zu Verteilungsfunktionen. Der entsprechende Begriff dafür sind die Quantile:

Definition 12.8. Sei X eine Zufallsvariable über $(\Omega, \mathscr{A}, P)$ und $0 < \alpha < 1$. Jede Zahl τ_α mit

$$P\{X < \tau_\alpha\} \le \alpha \le P\{X \le \tau_\alpha\}$$

heißt α-Quantil von X. Ein $\frac{1}{2}$-Quantil heißt ein Median von X. Die Quantile von $N(0,1)$- / t_n- / $F_{r,s}$- / χ_n^2-verteilten Zufallsvariablen werden mit u_α- / $\mathrm{t}_{n;\alpha}$- / $F_{r,s;\alpha}$- / $\chi_{n;\alpha}^2$ bezeichnet.

Beispiel 12.9 (t-Test). Der (zweiseitige) t-Test ist gegeben durch folgende Daten:

(i) $P \circ X_0^{-1} = N(\mu, \sigma^2)$ mit μ und σ unbekannt.

(ii) $H_0 : \mu = \mu_0$.

(iii) $T(X_1, \dots, X_n) := \frac{\overline{X}_{(n)} - \mu}{\frac{1}{\sqrt{n}} s_{(n)}}$. Diese Testgröße ist nach 12.7 c) t_{n-1}-verteilt, falls H_0 zutrifft.

(iv) $K = \{x \in \mathbb{R}^n : |T(x)| > t_{n-1;1-\alpha/2}\}$.

Beispiel 12.10 (χ^2-Streuungstest). Dieser Test ist durch folgende Daten gegeben:

(i) $\mathscr{P}^{X_0} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R},\ \sigma > 0\}$. Damit ist der Parameter gleich $\theta = (\mu, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$.

(ii) Bei einem zweiseitigen Test ist $H_0 : \sigma^2 = \sigma_0^2$, d.h.

$$H_0 = \{(\mu, \sigma) : \sigma = \sigma_0\} \subset \Theta\,.$$

(iii) Die Stichprobenvarianz $s^2_{(n)}$ ist ein erwartungstreuer Schätzer für σ^2. Nach Satz 12.7 b) hat, falls H_0 zutrifft, die Statistik $T \circ X := \frac{n-1}{\sigma_0^2} s^2_{(n)}$ die Verteilung χ^2_{n-1}.

(iv) Wahl von K: Man wird die Nullhypothese ablehnen, falls die beobachtete Realisierung von s^2 weit von σ_0^2 abweicht, d.h. falls die Realisierung der (positiven) Größe $T \circ X$ sehr groß oder sehr klein wird. Daher macht man den Ansatz

$$K = \{x \in \mathbb{R}^n : T(x) < c_1 \text{ oder } T(x) > c_2\}$$

mit zwei Konstanten $0 < c_1 < c_2$. Dies ist typisch für zweiseitige Tests. Die Konstanten c_1 und c_2 werden so bestimmt, daß

$$\begin{aligned} \sup_{\theta \in H_0} P_\theta\{T \circ X < c_1\} &= \frac{\alpha}{2}\,, \\ \sup_{\theta \in H_0} P_\theta\{T \circ X > c_2\} &= \frac{\alpha}{2}\,. \end{aligned}$$

Für $\theta \in H_0$ ist

$$P_\theta\{T \circ X < c_1\} = \chi^2_{n-1}(-\infty, c_1) = \chi^2_{n-1}(-\infty, c_1]\,.$$

Dies ist gleich $\alpha/2$ für $c_1 = \chi^2_{n-1;\alpha/2}$. Genauso ist

$$P_\theta\{T \circ X > c_2\} = 1 - \chi^2_{n-1}(-\infty, c_2]$$

gleich $\alpha/2$, falls $c_2 = \chi^2_{n-1;1-\frac{\alpha}{2}}$. Damit erhält man

$$K = \{x \in \mathbb{R}^n : T(x) \notin [\chi^2_{n-1;\frac{\alpha}{2}}, \chi^2_{n-1;1-\frac{\alpha}{2}}]\}\,.$$

(v) Durchführung des Tests: Bei Beobachtung des Tupels $\hat{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ berechnet man $\overline{x} := \frac{1}{n}\sum_{i=1}^n x_i$ und $T(\hat{x}) := \frac{1}{\sigma_0^2}\sum_{i=1}^n (x_i - \overline{x})^2$. Dann wird die Nullhypothese H_0 genau dann verworfen, falls $T(\hat{x}) < \chi^2_{n-1;\frac{\alpha}{2}}$ oder $T(\hat{x}) > \chi^2_{n-1;1-\frac{\alpha}{2}}$ gilt.

Im Falle eines einseitigen Tests muß man den kritischen Bereich K entsprechend modifizieren. Für $H_0 : \sigma^2 \leq \sigma_0^2$ erhält man

$$K = \{x \in \mathbb{R}^n : T(x) > \chi^2_{n-1;1-\alpha}\}\,,$$

für $H_0 : \sigma^2 \geq \sigma_0^2$ erhält man

$$K = \{x \in \mathbb{R}^n : T(x) < \chi^2_{n-1;\alpha}\}\,.$$

Nun soll noch kurz auf die Intervallschätzung eingegangen werden. Hier ist der Begriff des Signifikanzintervalls von Bedeutung, der eng mit Signifikanztests zusammenhängt, wie das nachfolgende Beispiel 12.12 zeigen wird.

Definition 12.11. Sei $\Theta \subset \mathbb{R}$ und $0 < \alpha < 1$. Sei (T_1, T_2) ein Intervallschätzer für θ, d.h. $(T_1, T_2)\colon \mathbb{R}^n \to \mathbb{R}^2$ ist meßbar mit $T_1(x) \le T_2(x)$ für alle $x \in \mathbb{R}$. Dann heißt das vom Zufall abhängende Intervall $[T_1 \circ X, T_2 \circ X]$ ein Konfidenzintervall für θ zum Konfidenzniveau α, falls

$$P_\theta\{\theta \in [T_1 \circ X, T_2 \circ X]\} \ge 1 - \alpha \qquad \text{für alle } \theta \in \Theta\,.$$

Beispiel 12.12. (Fortsetzung von Beispiel 12.4.) Hier ist $\mathscr{P}^{X_0} = \{N(\mu, \sigma_0^2) : \mu \in \mathbb{R}$ und $\theta = \mu$, wobei σ_0 bekannt ist. Während in Beispiel 12.4 der einseitige Gaußtest behandelt wurde, sei hier $H_0 = \{\theta_0\}$. Dann ist der durch

$$K = \left\{ x \in \mathbb{R}^n : \left| \frac{\overline{x} - \theta_0}{\sqrt{\frac{\sigma_0^2}{n}}} \right| > u_{1-\frac{\alpha}{2}} \right\}$$

gegebene Test signifikant zum Niveau α (zweiseitiger Gaußtest). Daher gilt

$$P_{\theta_0}\left\{ \left| \frac{\overline{X} - \theta_0}{\sqrt{\frac{\sigma_0^2}{n}}} \right| \le u_{1-\frac{\alpha}{2}} \right\} \ge 1 - \alpha$$

für alle $\theta_0 \in \Theta$. Mit anderen Worten, es gilt für alle θ_0

$$P_{\theta_0}\Big\{ \theta_0 \in \Big[\overline{X} - u_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_0^2}{n}},\ \overline{X} + u_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_0^2}{n}} \Big] \Big\} \ge 1 - \alpha\,.$$

Das in dieser Formel auftretende Intervall ist somit ein Konfidenzintervall zum Niveau α für θ. Man sieht außerdem, daß die Nullhypothese H_0 genau dann verworfen wird, falls gilt

$$\theta_0 \notin \Big[\overline{X} - u_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_0^2}{n}},\ \overline{X} + u_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_0^2}{n}} \Big]\,.$$

Ausblick: Weitere statistische Fragestellungen.

In dieser Vorlesung konnten nur einige wenige statistische Ideen kurz diskutiert werden. In den Anwendungen wichtig sind noch eine Reihe weiterer Methoden, welche hier kurz erwähnt werden sollen.

- χ^2-Anpassungstest (Test auf bestimmte Verteilung): Während oben stets davon ausgegangen wurde, daß eine Normalverteilung vorliegt, ist jetzt die Frage, ob etwa X_0 $N(0,1)$-verteilt ist. Genauer gesagt, wählt man sich eine Partition $\mathbb{R} = I_1 \,\dot\cup \ldots \dot\cup\, I_k$ und testet dann die Nullhypothese

  $$H_0 : P \circ X_0^{-1}(I_j) = N(0,1)(I_j) \qquad \text{für alle } j = 1, \ldots, k\,.$$

  Falls nur die Frage ist, ob X_0 normalverteilt ist, wird man gewöhnlich zunächst eine Punktschätzung für $\mathrm{E}\,X_0$ und $\mathrm{Var}\,X_0$ vornehmen und dann auf Normalverteilung mit diesen Schätzungen als Parameter testen.

- Zweistichprobentests: Wir waren stets von einer Stichprobe ausgegangen. Einer der einfachsten Zweistichprobentests geht etwa von Zufallsvariablen X_i, Y_j aus, wobei $(X_1, \ldots, X_n)$ i.i.d. $N(\mu_1, \sigma_1^2)$-verteilt und $(Y_1, \ldots, Y_m)$ i.i.d. $N(\mu_2, \sigma_2^2)$-verteilt ist. Eine mögliche Nullhypothese ist dann $H_0 : \mu_1 = \mu_2$.

- χ^2-Unabhängigkeitstest: Dieser Test behandelt ebenfalls zwei Stichproben, er ist aber verteilungsfrei, d.h. es wird nicht von einer bestimmten Verteilung (wie oben der Normalverteilung) ausgegangen. Das Prinzip des Tests liegt darin, Partitionen $\mathbb{R} = I_1 \,\dot\cup \ldots \dot\cup\, I_k = J_1 \,\dot\cup \ldots \dot\cup\, J_l$ zu wählen und die Nullhypothese

$$H_0 : P \circ (X_0, Y_0)^{-1}(I_i \times J_j) = P \circ X_0^{-1}(I_i) \cdot P \circ Y_0^{-1}(J_j) \qquad (i = 1, \ldots, k,\ j = 1, \ldots, l)$$

  zu testen.

- Lineare Regression: Im einfachsten Fall nimmt man hier an, daß die beobachtete Zufallsgröße Y bis auf einen zufälligen Fehlerterm linear von einem Parameter $x \in \mathbb{R}$ abhängt, d.h. man hat die Form

$$Y = Y(x) = \alpha + \beta x + Z$$

  mit unbekannten Parametern $\alpha, \beta \in \mathbb{R}$ und einer von x unabhängigen Zufallsgröße Z. Beobachtet werden Messungen der Form (x_k, y_k), wobei x_k bekannt ist und y_k eine Realisierung der Zufallsgröße $Y(x_k)$ darstellt. Typische Fragestellungen hier sind etwa die der Punktschätzung der Parameter α und β (hier wird etwa die sog. Methode der kleinsten Quadrate verwendet) und die Prognose, bei welcher man z.B. Konfidenzintervalle für $Y(x)$ bei gegebenen x angibt.

- Varianzanalyse: Diese tritt in der Praxis etwa auf, wenn bei der medizinischen Behandlung von Patienten durch mehrere verschiedene Behandlungsmethoden die Frage untersucht werden soll, ob zwischen den Behandlungserfolgen bei den einzelnen Methoden signifikante Unterschiede festzustellen sind.

Anhang A. Endliche Produkte von Maßräumen

A.1. Produkte von Meßräumen, Produkt-σ-Algebren

Seien $(\Omega_i; \mathscr{A}_i)$, $1 \leq i \leq n$, endlich viele Meßräume. Dazu betrachte das kartesische Produkt

$$\Omega := \Omega_1 \times \ldots \times \Omega_n \equiv \prod_{i=1}^{n} \Omega_i.$$

Definition A.1. Die Menge

$$\mathscr{Z} := \{A_1 \times \ldots \times A_n : A_i \in \mathscr{A}_i, 1 \leq i \leq n\} \subset \mathbb{P}(\Omega)$$

heißt *Gesamtheit der Zylindermengen* (bzgl. der $\mathscr{A}_i$).

$$\mathscr{A}_1 \otimes \ldots \otimes \mathscr{A}_n := \bigotimes_{i=1}^{n} \mathscr{A}_i := \sigma(\mathscr{Z})$$

heißt die *Produkt-σ-Algebra der $\mathscr{A}_i$.*

Bemerkung A.2.

i) $\mathscr{Z}$ ist $\cap$-stabil.

   Denn. $(A_1 \times \ldots \times A_n) \cap (B_1 \times \ldots \times B_n) = (A_1 \cap B_1) \times \ldots \times (A_n \cap B_n)$. □

ii) $\otimes$ ist assoziativ: $\mathscr{A}_1 \otimes \mathscr{A}_2 \otimes \mathscr{A}_3 = (\mathscr{A}_1 \otimes \mathscr{A}_2) \otimes \mathscr{A}_3 = \mathscr{A}_1 \otimes (\mathscr{A}_2 \otimes \mathscr{A}_3)$.

   Denn. Zu zeigen ist wegen Symmetrie nur die erste Gleichheit und diese bedeutet definitionsgemäß $\sigma(\{A_1 \times A_2 \times A_3\}) = \sigma(\sigma(\{A_1 \times A_2\}) \times A_3)$. Dabei ist "$\subset$" klar und "$\supset$" folgt aus

$$\begin{aligned}\sigma(\{A_1 \times A_2 \times A_3 : A_1 \in \mathscr{A}_1, A_2 \in \mathscr{A}_2, A_3 \text{ fest}\}) &= \sigma(\{A_1 \times A_2\}) \times A_3 \\ &\subset \sigma(\{A_1 \times A_2 \times A_3' : A_1 \in \mathscr{A}_1, A_2 \in \mathscr{A}_2, A_3' \in \mathscr{A}_3\})\end{aligned}$$

   für jedes $A_3 \in \mathscr{A}_3$. □

iii) Sei $\mathrm{pr}_i : \Omega_1 \times \ldots \times \Omega_n \to (\Omega_i, \mathscr{A}_i)$ die i-te Koordinatenprojektion. Dann gilt

$$\bigotimes_{i=1}^{n} \mathscr{A}_i = \sigma(\mathrm{pr}_i : 1 \leq i \leq n).$$

   Die Produkt-σ-Algebra ist somit die kleinste σ-Algebra auf Ω, die alle Koordinatenprojektionen meßbar macht.

Denn. Für $A_i \in \mathscr{A}_i$ folgt $\mathrm{pr}_i^{-1}(A_i) = \Omega_1 \times \ldots \times A_i \times \ldots \times \Omega_n$, $1 \leq i \leq n$. Somit ist $A_1 \times \ldots \times A_n = \bigcap_{i=1}^n \mathrm{pr}_i^{-1}(A_i) \in \sigma(\mathrm{pr}_i : 1 \leq i \leq n)$. Damit gilt "$\subset$".
"$\supset$" folgt sofort aus $\mathrm{pr}_i^{-1}(A_i) \in \bigotimes_{i=1}^n \mathscr{A}_i$ wie eben gesehen. □

Satz A.3. *Zu $1 \leq i \leq n$ sei jeweils $\mathscr{E}_i \subset \mathscr{A}_i$ ein Erzeuger von $\mathscr{A}_i$ (also $\sigma(\mathscr{E}_i) = \mathscr{A}_i$) derart, daß es stets eine Folge $(E_{ik})_{k\in\mathbb{N}}$ in $\mathscr{E}_i$ gibt mit $E_{ik} \nearrow \Omega_i$ für $k \to \infty$. Dann gilt*

$$\sigma(\{E_1 \times \ldots \times E_n : E_i \in \mathscr{E}_i, 1 \leq i \leq n\}) = \bigotimes_{i=1}^n \mathscr{A}_i.$$

Beweis. "$\subset$" ist klar und für die umgekehrte Inklusion ist zu zeigen, daß $A_1 \times \ldots \times A_n \in \sigma(\{E_1 \times \ldots \times E_n\})$ gilt für alle $A_i \in \mathscr{A}_i$. Dazu wiederum reicht es (vgl. letzte Bemerkung), $\Omega_1 \times \ldots \times A_i \times \ldots \times \Omega_n \in \sigma(\{E_1 \times \ldots \times E_n\})$ zu verifizieren. Hierfür betrachte zu gegebenem $E_i \in \mathscr{E}_i$

$$F_k := E_{1k} \times \ldots \times E_{i-1,k} \times E_i \times E_{i+1,k} \times \ldots \times E_{nk} \in \sigma(\{E_1 \times \ldots \times E_n\}).$$

Dann ist auch $\Omega_1 \times \ldots \times E_i \times \ldots \times \Omega_n = \bigcup_{k\in\mathbb{N}} F_k$ in $\sigma(\{E_1 \times \ldots \times E_n\})$. Wegen

$$\begin{aligned}\Omega_1 \times \ldots \times A_i \times \ldots \times \Omega_n &\in \Omega_1 \times \ldots \times \sigma(\mathscr{E}_i) \times \ldots \times \Omega_n \\ &= \sigma(\{\Omega_1 \times \ldots \times E_i \times \ldots \times \Omega_n)\}\end{aligned}$$

folgt die Behauptung. □

Bemerkung A.4. Die Behauptung des Satzes wird falsch ohne die Forderung der aufsteigenden Folgen! Beispielsweise erzeugt $\{\emptyset\}$ die triviale σ-Algebra $\mathscr{A}_1 = \{\emptyset, \Omega_1\}$, aber wenn $\mathscr{A}_2$ mehr als zwei Mengen umfaßt, ist $\sigma(\{\emptyset\}) \neq \mathscr{A}_1 \otimes \mathscr{A}_2$.

Satz A.5. *Sind jeweils $\mathcal{U}_i$ abzählbare Basen von Topologien $\mathcal{T}_i$ auf Ω_i, $1 \leq i \leq n$, so ist $\{U_1 \times \ldots \times U_n : U_i \in \mathcal{U}_i\}$ abzählbare Basis der Produkttopologie $\prod_{i=1}^n \mathcal{T}_i =: \mathcal{T}$ auf Ω und es gilt*

$$\mathscr{B}(\Omega) \equiv \sigma(\mathcal{T}) = \bigotimes_{i=1}^n \sigma(\mathcal{T}_i) \equiv \bigotimes_{i=1}^n \mathscr{B}(\Omega_i).$$

Beweis. Die topologische Aussage ist bekannt und in der behaupteten Gleichung folgt die Inklusion "$\subset$" wegen $\{U_1 \times \ldots \times U_n : U_i \in \mathcal{U}_i\} \subset \mathcal{Z}$ und damit $\sigma(\mathcal{T}) = \sigma\{U_1 \times \ldots \times U_n : U_i \in \mathcal{U}_i\}) \subset \sigma(\mathcal{Z}) = \bigotimes_{i=1}^n \sigma(\mathcal{T}_i)$. Dabei ist zu beachten, daß jedes $U \in \mathcal{T}$ geschrieben werden kann als abzählbare Vereinigung $U = \bigcup_{k\in\mathbb{N}} U_{1k} \times \ldots \times U_{nk}$ mit Komponenten aus den $\mathcal{U}_i$.
Für "$\supset$" genügt zu zeigen, daß $\mathcal{Z} \subset \sigma\{U_1 \times \ldots \times U_n : U_i \in \mathcal{T}_i\})$. Hierfür wiederum müssen wir nur Mengen der Gestalt $\Omega_1 \times \ldots \times A_i \times \ldots \times \Omega_n$ betrachten, da ja – wie gesehen – beliebige Zylindermengen endliche Durchschnitte solcher Mengen sind. Ist jedoch $A_i \in \mathscr{B}(\Omega_i)$, so gilt $\Omega_1 \times \ldots \times A_i \times \ldots \times \Omega_n \in \Omega_1 \times \ldots \times \sigma(\mathcal{U}_i) \times \ldots \times \Omega_n = \sigma(\Omega_1 \times \ldots \times U_i \times \ldots \times \Omega_n : U_i \in \mathcal{U}_i)$. □

Korollar A.6. *Da die natürliche Topologie von $\mathbb{R}$ eine abzählbare Basis besitzt, gilt*

$$\mathscr{B}(\mathbb{R}^n) = \bigotimes_1^n \mathscr{B}(\mathbb{R}).$$

A.2. Produktmaße

Seien $(\Omega_i; \mathscr{A}_i; \mu_i)$, $1 \leq i \leq n$, nun endlich viele Maßräume.

Definition A.7. Ein Maß $\mu : \bigotimes_{i=1}^n \mathscr{A}_i \to \overline{\mathbb{R}}_+$ heißt *Produktmaß der* μ_i, wenn

$$\mu(A_1 \times \ldots \times A_n) = \prod_{i=1}^n \mu(A_i) \quad \text{für alle } A_i \in \mathscr{A}_i, 1 \leq i \leq r$$

(mit der Konvention $0 \cdot \infty = 0$).

Die Existenz und Eindeutigkeit des Produktmaßes ist nur für das Produkt σ-finiter Maßräume gewährleistet!

Seien also ab nun $(\Omega_i; \mathscr{A}_i; \mu_i)$, $1 \leq i \leq n$, σ-finit.
Wir konstruieren das Produktmaß zunächst für zwei Faktoren, Produkte höherer Ordnung definiert man dann induktiv in Verbindung mit geeigneten Assoziativitätsüberlegungen.

Lemma A.8. *Sei $f : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}_+$ $\mathscr{A}_1 \otimes \mathscr{A}_2$-meßbar. Dann gilt:*

i) Für jedes $x \in \Omega_1$ ist $f(x, \cdot\,) : \Omega_2 \to \overline{\mathbb{R}}_+$ $\mathscr{A}_2$-meßbar.

ii) $x \mapsto \int f(x, \cdot\,)\mathrm{d}\mu_2 =: \int f(x,y)\mu_2(\mathrm{d}y)$ ist $\mathscr{A}_1$-meßbar.

Beweis. Ohne Einschränkung betrachten wir nur endliche μ_2 (also $\mu_2(\Omega_2) < \infty$), denn sonst wählen wir wegen der σ-Finitheit $\mathscr{A}_2 \ni S_k \nearrow \Omega_2$ mit $\mu_2(S_k) < \infty$ und betrachten statt f nun $f_k := f \cdot 1_{\Omega_1 \times S_k}$, das bezüglich der zweiten Komponente auf dem endlichen Maßraum $(\Omega_2; \mathscr{A}_2; \mu_2(S_k \cap \cdot\,)$ lebt. Stimmt die Behauptung dann für die f_k, so wegen $f_k \nearrow f$ punktweise auch für f, wobei man jeweils den Permanenzsatz (punktweise Limiten meßbarer Funktionen sind meßbar) und für den zweiten Teil außerdem monotone Konvergenz anwenden muß.

Also sei nun μ_2 endlich. Dazu betrachten wir das Teilmengensystem der Produkt-σ-Algebra

$$\mathcal{D} := \{A \in \mathscr{A}_1 \otimes \mathscr{A}_2 : 1_A \text{ hat die Eigenschaften i) und ii)}\}.$$

Wir zeigen, daß $\mathcal{D}$ ein Dynkin-System ist (nach den Kriterien aus Hackenbroch, Integrationstheorie, bzw. Blatt 1, Aufgabe 2a):

- $\Omega = \Omega_1 \times \Omega_2 \in \mathcal{D}$, da $1_\Omega = 1$.
- $\mathcal{D} \ni A_l \nearrow A$, dann auch $A \in \mathcal{D}$: Es gilt $1_{A_l} \nearrow 1_A$ und damit folgen i) und ii) gemäß Permanenzsatz und monotoner Konvergenz.
- $A, B \in \mathcal{D}$, $A \subset B$, dann auch $B \setminus A \in \mathcal{D}$: Es ist $1_{B\setminus A} = 1_B - 1_A$ und i) ist klar. ii) folgt aus der Endlichkeit von μ_2, denn damit ist $\int 1_{B\setminus A}(x, \cdot)\mathrm{d}\mu_2 = \int 1_B(x, \cdot)\mathrm{d}\mu_2 - \int 1_A(x, \cdot)\mathrm{d}\mu_2$ ("$\infty - \infty$" kann nicht auftreten).

Außerdem umfaßt $\mathcal{D}$ den $\cap$-stabilen Erzeuger $\mathcal{Z}$ der Zylindermengen, denn $A = A_1 \times A_2 \in \mathcal{Z}$, dann ist $1_A(x,y) = 1_{A_1}(x)1_{A_2}(y)$, woraus sich i) ergibt und mit $\int 1_A(x, \cdot)\mathrm{d}\mu_2 = \int 1_{A_1}(x)1_{A_2}\mathrm{d}\mu_2 = 1_{A_1}(x)\mu_2(A_2)$ folgt ii).

Das Dynkin-Lemma besagt nun $\mathcal{D} = \mathscr{A}_1 \otimes \mathscr{A}_2$, was gleichbedeutend damit ist, daß die Behauptung für alle 1_A, $A \in \mathscr{A}_1 \otimes \mathscr{A}_2$ richtig ist, wegen Linearität also für alle Stufenfunktionen, mit Permanenz und monotoner Konvergenz daher für alle $\overline{\mathbb{R}}_+$-wertigen produktmeßbaren Funktionen. □

Lemma A.9. *Es existiert eindeutig das Produktmaß $\mu_1 \otimes \mu_2$ auf $\mathscr{A}_1 \otimes \mathscr{A}_2$. Explizit ist es gegeben durch*

$$\mu_1 \otimes \mu_2(A) = \int \left(\int 1_A(x, \cdot)\mathrm{d}\mu_2 \right) \mu_1(\mathrm{d}x) =: \int \mu_1(\mathrm{d}x) \int \mu_2(\mathrm{d}y) 1_A(x,y).$$

Beweis.

i) $\mu_1 \otimes \mu_2$ ist ein Maß: Klar ist $\mu_1 \otimes \mu_2(\emptyset) = \int \left(\int 1_\emptyset(x, \cdot)\mathrm{d}\mu_2\right) \mu_1(\mathrm{d}x) = \int 0 \,\mathrm{d}\mu_1 = 0$. $\mu_1 \otimes \mu_2$ ist additiv wegen $1_{A_1 \dot\cup A_2} = 1_{A_1} + 1_{A_2}$ und damit

$$\begin{aligned}\mu_1 \otimes \mu_2(A_1 \dot\cup A_2) &= \int \left(\int 1_{A_1}(x, \cdot) + 1_{A_2}(x, \cdot)\mathrm{d}\mu_2 \right) \mu_1(\mathrm{d}x) \\ &= \mu_1 \otimes \mu_2(A_1) + \mu_1 \otimes \mu_2(A_2)\end{aligned}$$

wegen Linearität der Integrale.

Für die σ-Additivität ist somit noch die σ-Stetigkeit von unten zu zeigen (Wahrscheinlichkeitstheorie, Satz 1.7). Sei daher $\mathscr{A}_1 \otimes \mathscr{A}_2 \ni A_l \nearrow A$. Dann folgt wegen $1_{A_l} \nearrow 1_A$

$$\begin{aligned}\mu_1 \otimes \mu_2(A_l) &= \int \left(\int 1_{A_l}(x, \cdot)\mathrm{d}\mu_2 \right) \mu_1(\mathrm{d}x) \\ &\nearrow \int \left(\int 1_A(x, \cdot)\mathrm{d}\mu_2 \right) \mu_1(\mathrm{d}x) = \mu_1 \otimes \mu_2(A)\end{aligned}$$

durch zweimalige Anwendung monotoner Konvergenz.

ii) $\mu_1 \otimes \mu_2$ ist Produktmaß von $\mu_1 \otimes \mu_2$:

$$\begin{aligned}\mu_1 \otimes \mu_2(A_1 \times A_2) &= \int \left(\int 1_{A_1}(x) 1_{A_2} \mathrm{d}\mu_2 \right) \mu_1(\mathrm{d}x) \\ &= \int 1_{A_1}(x)\mu_2(A_2)\mu_1(\mathrm{d}x) = \mu_2(A_2)\mu_1(A_1).\end{aligned}$$

iii) Zur Eindeutigkeit: $\mu_1 \otimes \mu_2$ liegt durch $\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ fest auf dem $\cap$-stabilen Erzeuger $\mathscr{Z}$ von $\mathscr{A}_1 \otimes \mathscr{A}_2$. Da die beiden Maße σ-finit sind, gibt es $\mathscr{A}_1 \ni S_k \nearrow \Omega_1$, $\mu_1(S_k) < \infty$, sowie $\mathscr{A}_2 \ni T_k \nearrow \Omega_2$, $\mu_2(T_k) < \infty$. Damit ergibt sich $S_k \times T_k \nearrow \Omega_1 \times \Omega_2$ mit $\mu_1 \otimes \mu_2(S_k \times T_k) = \mu_1(S_k)\mu_2(T_k) < \infty$ und die Eindeutigkeit folgt aus dem Eindeutigkeitssatz für σ-finite Maße, vgl. Hackenbroch, Integrationstheorie, oder Blatt 1, Aufgabe 2b.

□

Bemerkung A.10.

i) Es gilt auch $\mu_1 \otimes \mu_2(A) = \int \left(\int 1_A(\,\cdot\,, y) \mathrm{d}\mu_1 \right) \mu_2(\mathrm{d}y)$.

Denn. Dies ist richtig für $A = A_1 \times A_2$ und damit überhaupt analog zu eben. □

ii) $(\mu_1 \otimes \mu_2) \otimes \mu_3 = \mu_1 \otimes (\mu_2 \otimes \mu_3) =: \mu_1 \otimes \mu_2 \otimes \mu_3$ usw. für höhere Produkte.

Denn. Die Gleichung stimmt für beliebige $A = A_1 \times A_2 \times A_3 \in \mathscr{Z}$ (iteriertes Ausrechnen der Integrale) und gibt ein Produktmaß auf $\mathscr{A}_1 \otimes \mathscr{A}_2 \otimes \mathscr{A}_3$. Der Eindeutigkeitssatz für Maße liefert wie vorhin die Behauptung. □

iii) "Gegenbeispiel": Ohne σ-Finitheit wird alles falsch!
Seien $\Omega_1 = \Omega_2 = [0,1]$ beide versehen mit den Borelmengen und $\mu_1 = \lambda|_{[0,1]}$ sowie $\mu_2 = \zeta$ das Zählmaß, das auf dem überabzählbaren Intervall nicht σ-finit ist. Wir betrachten die Diagonale

$$\Delta := \{(x,x) : x \in [0,1]\} = f^{-1}\{0\} \in \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R}),$$

wo $f : [0,1]^2 \to \mathbb{R}$, $(x,y) \mapsto x - y$, stetig und damit $\mathscr{B}(\mathbb{R}^2) = \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R})$-meßbar ist. Aber nun berechnet sich (kurz λ für $\lambda|_{[0,1]}$)

$$\int \zeta(\mathrm{d}y) \int \lambda(\mathrm{d}x) 1_\Delta(x,y) = \int \zeta(\mathrm{d}y) \int_{\{y\}} 1 \,\mathrm{d}\lambda = \int 0 \,\mathrm{d}\zeta = 0$$

im Gegensatz zu

$$\int \lambda(\mathrm{d}x) \int \zeta(\mathrm{d}y) 1_\Delta(x,y) = \int \lambda(\mathrm{d}x) \int_{\{x\}} 1 \,\mathrm{d}\zeta = \int 1 \,\mathrm{d}\lambda = 1.$$

A.3. Der Satz von Fubini-Tonelli über Mehrfachintegrale

Nach wie vor seien $(\Omega_1;\mathscr{A}_1;\mu_1)$ und $(\Omega_2;\mathscr{A}_2;\mu_2)$ σ-finit sowie $\mu := \mu_1 \otimes \mu_2$ das Produktmaß auf der Produkt-σ-Algebra $\mathscr{A}_1 \otimes \mathscr{A}_2$.

Satz A.11 (Fubini-Tonelli).

i) (L. Tonelli, 1885-1946).
Sei $f : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}_+$ $\mathscr{A}_1 \otimes \mathscr{A}_2$-meßbar. Dann gilt (in $\overline{\mathbb{R}}_+$)

$$\int f \mathrm{d}\mu = \int \mu_1(\mathrm{d}x) \int \mu_2(\mathrm{d}y) f(x,y) = \int \mu_2(\mathrm{d}x) \int \mu_1(\mathrm{d}y) f(x,y). \qquad (*)$$

ii) (G. Fubini, 1879-1943).
Sei $f : \Omega_1 \times \Omega_2 \to \mathbb{C} \in L^1(\mu)$ (also: f ist integrierbar bzgl. des Produktmaßes). Dann gilt die Aussage von (*) sinngemäß, nämlich:

(a) $N := \{x \in \Omega_1 : \int \mu_2(\mathrm{d}y)|f(x,y)| = \infty\} \in \mathscr{A}_1$ mit $\mu_1(N) = 0$.

(b) $g : \Omega_1 \to \mathbb{C}$ definiert durch

$$g(x) := \begin{cases} \int \mu_2(\mathrm{d}y) f(x,y), & x \notin N, \\ 0, & x \in N, \end{cases}$$

ist in $L^1(\mu_1)$ und es gilt

$$\int f \mathrm{d}\mu = \int g \mathrm{d}\mu_1 \equiv \int \mu_1(\mathrm{d}x) \int \mu_2(\mathrm{d}y) f(x,y)$$

(und natürlich läßt sich die entsprechende Aussage völlig symmetrisch bzgl. vertauschter Rollen von μ_1 und μ_2 bilden).

Beweis. i) Nach Lemma A.8 existieren alle Integrale in $\overline{\mathbb{R}}_+$. Nach Lemma A.9 gilt (*) für Indikatorfunktionen 1_A, $A \in \mathscr{A}_1 \otimes \mathscr{A}_2$. Mittels Linearität folgt (*) für Stufenfunktionen, schließlich mit monotoner Konvergenz für alle produktmeßbaren $\overline{\mathbb{R}}_+$-wertigen Funktionen.

ii) Lemma A.8 für $|f|$ liefert die $\mathscr{A}_1$-Meßbarkeit von $x \mapsto \int \mu_2(\mathrm{d}y)|f(x,y)|$. Somit ist $N \in \mathscr{A}_1$ und nach der schon bewiesenen Tonelli-Aussage ergibt $f \in L^1(\mu)$, daß $\int \mu_1(\mathrm{d}x) \int \mu_2(\mathrm{d}y)|f(x,y)| = \int |f| \mathrm{d}\mu < \infty$ und damit $\mu_1(N) = 0$.

Also ist auch $\mu(N \times \Omega_2) = \mu_1(N)\mu_2(\Omega_2) = 0$. Betrachte nun $\tilde{f} := f \cdot 1_{(N \times \Omega_2)^c} = f \cdot 1_{N^c \times \Omega_2}$, dann ist $\tilde{f} = f$ μ-f.ü. Nun folgt für $f \geq 0$ (und damit auch $\tilde{f} \geq 0$)

$$\int g \mathrm{d}\mu_1 = \int \mu_1(\mathrm{d}x) \int \mathrm{d}\mu_2 \tilde{f}(x,\,\cdot\,) = \int \tilde{f} \mathrm{d}\mu = \int f \mathrm{d}\mu,$$

wobei die mittlere Gleichheit wieder nach Tonelli folgt. Für ein beliebiges komplexwertiges f folgt dann die Behauptung durch dessen Zerlegung in vier nichtnegative Komponenten: $f = (\operatorname{Re} f)_+ - (\operatorname{Re} f)_- + \mathrm{i}[(\operatorname{Im} f)_+ - (\operatorname{Im} f)_-]$.

$\square$

Anhang B. Übungsblätter

Mit einem Stern versehene Aufgaben sind freiwillig zu lösen und zählen nicht zum Punktesoll. Soweit nicht anders vermerkt, zählt eine Aufgabe vier Punkte.

— Blatt 1 —

1. (8 Punkte). Welche der folgenden maßtheoretischen Aussagen sind richtig, welche falsch (Beweis oder Gegenbeispiel)? Sei $(\Omega; \mathscr{A}; \mu)$ ein Maßraum.

i) Ist $\{\omega\} \in \mathscr{A}$ für jedes $\omega \in \Omega$, so ist $\mathscr{A}$ bereits die Potenzmenge $\mathbb{P}(\Omega)$ von Ω. (1 Punkt)

ii) Ist $\{\omega\} \in \mathscr{A}$ für jedes $\omega \in \Omega$ und $\{\omega \in \Omega : \mu(\{\omega\}) > 0\}$ meßbar, so hat diese Menge bereits Maß $\mu(\Omega)$. (1 Punkt)

iii) Ist $(A_n)_{n\in\mathbb{N}}$ eine Folge in $\mathscr{A}$ mit $\mu(A_n) = \mu(A_{n+1})$ und $A_{n+1} \subset A_n$ für alle $n \in \mathbb{N}$, so gilt $\mu\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \mu(A_1)$. (1 Punkt)

iv) Genau dann ist $A \subset \Omega$ in $\mathscr{A}$, wenn $1_A : \Omega \to \mathbb{R}$ meßbar. (1 Punkt)

v) Genau die abzählbaren Teilmengen von $\mathbb{R}$ sind die Borelmengen vom Lebesguemaß 0.
(2 Punkte)

vi) Die nichtnegativen meßbaren Funktionen auf Ω sind genau die punktweisen Limiten von wachsenden Folgen nichtnegativer Stufenfunktionen. (2 Punkte)

(Ohne Bewertung: Welche der falschen Aussagen werden richtig für abzählbares Ω bzw. endliches Maß μ?)

2. (Dynkin-Systeme).

i) Genau dann ist $\mathscr{A} \subset \mathbb{P}(\Omega)$ ein Dynkin-System, wenn es folgende drei Eigenschaften erfüllt:

(i*) $\Omega \in \mathscr{A}$.

(ii*) Aus $A, B \in \mathscr{A}$ mit $A \subset B$ folgt $B \setminus A \in \mathscr{A}$ ($\mathscr{A}$ ist *relativ komplementiert*).

(iii*) Ist $(A_n)_{n\in\mathbb{N}}$ eine aufsteigende Folge in $\mathscr{A}$ (also $A_n \subset A_{n+1}$ für alle n), so ist auch $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{A}$ ($\mathscr{A}$ ist *σ-stabil*).

ii) (Eindeutigkeitssatz). *Seien μ, ν zwei σ-finite Maße auf einem Meßraum $(\Omega; \mathscr{A})$ und $\mathscr{E} \subset \mathscr{A}$ $\cap$-stabiler Erzeuger von $\mathscr{A}$ (also $\sigma(\mathscr{E}) = \mathscr{A}$), auf dem μ und ν übereinstimmen. Außerdem existiere bereits eine aufsteigende Folge (A_n) in $\mathscr{E}$ mit $\mu(A_n) < \infty$ für alle n und $\bigcup_{n\in\mathbb{N}} A_n = \Omega$. Dann folgt $\mu = \nu$.*

Anleitung: Man zeige, daß sämtliche

$$\mathscr{A}_n := \{A \in \mathscr{A} : \mu(A_n \cap A) = \nu(A_n \cap A)\}$$

$\mathscr{E}$ umfassende Dynkin-Systeme sind, benutze das Dynkin-Lemma und Stetigkeit von unten.

3. (Äußere Maße). Eine Abbildung $\mu : \mathscr{A} \to \overline{\mathbb{R}}_+ (:= [0;\infty])$ auf einer Algebra $\mathscr{A}$ auf Ω heißt *äußeres Maß*, wenn μ monoton und σ-subadditiv ist (letzteres bedeutet für jede Folge (A_n) in $\mathscr{A}$ mit $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{A}$, daß $\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} \mu(A_n)$ gilt).

i) Jedes Maß auf einem Meßraum ist ein äußeres Maß.

ii) Das zu einem σ-additiven Inhalt auf einer Algebra gehörige äußere Maß ist ein äußeres Maß.

iii) Man gebe ein Beispiel eines äußeren Maßes auf einer σ-Algebra, das kein Maß ist. (Mit Begründung! Etwa ist das zum Lebesgue-Maß λ auf $\mathscr{B}(\mathbb{R})$ gehörige äußere Maß λ^* kein Maß auf der Potenzmenge von $\mathbb{R}$, aber es gibt wesentlich einfachere Möglichkeiten.)

***4.** (Mengenalgebren als kommutative Algebren mit 1).
Sei Ω eine Menge und $\mathscr{F} \subset \mathbb{P}(\Omega)$ eine Mengenalgebra auf Ω. Es bezeichne wie in der Algebra üblich $\mathbb{F}_2 \equiv \mathbb{Z}/2\mathbb{Z}$ den Körper mit zwei Elementen.

i) Bezüglich der Verknüpfungen Addition "$+$":$= \triangle$ (symmetrische Differenz) und Multiplikation "$\cdot$":$= \cap$ zusammen mit dem Strukturmorphismus $\varphi : \mathbb{F}_2 \to \mathscr{F}$, $0 \mapsto \emptyset$, $1 \mapsto \Omega$, wird $\mathscr{F}$ zu einer kommutativen $\mathbb{F}_2$-Algebra mit $1(=\Omega)$ (vgl. zu diesen Begriffen z.B. Kunz, Algebra, §6.VI).

ii) In diesem Kontext stellt die Bildung der charakteristischen Funktion

$$\chi : \mathscr{F} \to {\mathbb{F}_2}^{\Omega} \equiv \{f : \Omega \to \mathbb{F}_2\}, \quad A \mapsto 1_A,$$

einen Monomorphismus von kommutativen $\mathbb{F}_2$-Algebren mit 1 dar (man kann somit via χ die Mengenalgebren auf Ω, also die Mengenunteralgebren von $\mathbb{P}(\Omega)$, mit den Unteralgebren von ${\mathbb{F}_2}^{\Omega}$ identifizieren).

— Blatt 2 —

***5.** (σ-finite Maße). Sei $(\Omega;\mathscr{A};\mu)$ ein Maßraum. Genau dann ist μ σ-finit, wenn μ abzählbare positive Linearkombination von *zueinander singulären* Wahrscheinlichkeitsmaßen ist, d.h. es gibt eine höchstens abzählbare Familie ($N \subset \mathbb{N}$) $(\alpha_n)_{n\in N}$ positiver reeller Zahlen sowie eine Familie $(\mu_n)_{n\in N}$ von Wahrscheinlichkeitsmaßen auf $\mathscr{A}$, für die gilt: Zu jedem $n \in N$ existiert $A_n \in \mathscr{A}$ mit $\mu_m(A_n) = \delta_{mn}$ für alle $m, n \in N$ und

$$\mu = \sum_{n\in N} \alpha_n \mu_n.$$

6. Bestimme alle $\{0;1\}$-wertigen Maße auf $\mathbb{R}$ versehen mit der von sämtlichen Punkten aus $\mathbb{R}$ erzeugten σ-Algebra.

7. (σ-Ideale). Sei $(\Omega;\mathscr{A})$ ein Meßraum. Ein Mengensystem $\mathscr{N} \subset \mathbb{P}(\Omega)$ heißt σ-*Ideal*, wenn

(i) $\emptyset \in \mathscr{N}$.

(ii) $N \in \mathscr{N}$, $M \subset N \Rightarrow M \in \mathscr{N}$.

(iii) $(N_n)_{n\in\mathbb{N}} \subset \mathscr{N} \Rightarrow \bigcup_{n\in\mathbb{N}} N_n \in \mathscr{N}$.

Beweise nun:

i) Für jedes σ-Ideal $\mathscr{N}$ gilt

$$\begin{aligned}\sigma(\mathscr{A}\cup\mathscr{N}) &= \{A\triangle N : A\in\mathscr{A}, N\in\mathscr{N}\}\\ &= \{B\subset\Omega : \exists A\in\mathscr{A}, N\in\mathscr{N} \text{ mit } B\setminus N = A\setminus N\}\\ &= \{B\subset\Omega : \exists A\in\mathscr{A} \text{ mit } B\triangle A\in\mathscr{N}\}.\end{aligned}$$

ii) Ist μ ein Maß auf $(\Omega;\mathscr{A})$, so ist das System

$$\mathscr{N}_\mu := \{N\subset\Omega : \exists N\subset\widetilde{N}\in\mathscr{A} \text{ mit } \mu(\widetilde{N})=0\}$$

der μ-Nullmengen ein σ-Ideal von $\mathbb{P}(\Omega)$.

8. Sei $f:\Omega\to\mathbb{C}$ eine integrierbare Funktion auf dem Maßraum $(\Omega;\mathscr{A};\mu)$ und $(A_n)_{n\in\mathbb{N}}$ eine Folge in $\mathscr{A}$ mit $\mu(A_n)\to 0$ bei $n\to\infty$. Dann folgt

$$\int_{A_n} f\,\mathrm{d}\mu \to 0 \quad \text{für} \quad n\to\infty.$$

(Hinweis: Für ein geeignetes $m\in\mathbb{N}$ zerlege Ω in $\{|f|>m\}\cup\{|f|\le m\}$ und benutze majorierte Konvergenz.)

***9.** (Metrisierbarkeit der stochastischen Konvergenz).
Sei $(\Omega;\mathscr{A};P)$ ein Wahrscheinlichkeitsraum und $\mathscr{L} := \{X:\Omega\to\mathbb{C} \text{ meßbar}\}$ sowie $\mathscr{N} := \{X\in\mathscr{L} : X=0\ P\text{-fast überall}\}$.

i) $\mathscr{L}$ ist ein $\mathbb{C}$-Vektorraum, $\mathscr{N}$ ist Unterraum von $\mathscr{L}$ und auf dem Quotientenvektorraum $L := \mathscr{L}/\mathscr{N}$ erhält man eine Metrik $\rho(\,\cdot\,,\,\cdot\,) : L\times L\to\mathbb{R}$ durch

$$\rho(X,Y) := \int |X-Y|\wedge 1\,\mathrm{d}P \qquad (\text{mit } x\wedge y := \min(x,y) \text{ für } x,y\in\mathbb{R}).$$

ii) Eine Folge $(X_n)_{n\in\mathbb{N}}$ in L konvergiert bezüglich ρ gegen 0 genau dann, wenn für alle $\delta>0$

$$P\{|X_n|>\delta\}\to 0 \quad \text{mit} \quad n\to\infty.$$

— Blatt 3 —

10. (Endliche Wahrscheinlichkeitsräume, 8 Punkte).

i) Wie hoch ist die Wahrscheinlichkeit, daß in einer Übungsgruppe von 23 Studenten mindestens zwei Teilnehmer am gleichen Tag Geburtstag haben (jeder Tag ist gleich wahrscheinlich; keiner ist in einem Schaltjahr geboren).

ii) Wie hoch ist die Wahrscheinlichkeit, beim Lotto "6 aus 49" drei oder vier Richtige zu haben?

iii) Beim Schafkopfen (32 Karten, 4 Spieler erhalten je 8 Karten) bekommt der Ausspieler zufällig alle acht Eichel-Karten. Wie groß ist die Wahrscheinlichkeit hierfür? Wie wahrscheinlich ist es dann, daß jeder andere Mitspieler einen der übrigen drei Unter besitzt (und unser Ausspieler mit einem "Wenz-Tout" steinreich wird)? Ändert sich letztere Wahrscheinlichkeit, wenn man mit "kurzen" Karten spielt (also 24 Karten insgesamt, jeder Spieler erhält 6 Stück)? Wenn ja, wie?

11. (Münzwurf). Eine *Laplace*-Münze (also eine Münze mit je Wahrscheinlichkeit $\frac{1}{2}$ für "Kopf" und "Zahl") wird solange geworfen, bis zum ersten Mal "Zahl" fällt.

i) Zeichne ein entsprechendes Baumdiagramm und beschreibe das Zufallsexperiment als Wahrscheinlichkeitsraum mit $\Omega = \mathbb{N}$. Welche Annahme bezüglich der Münzwürfe steckt in der "natürlichen" Wahl des Wahrscheinlichkeitsmaßes?

ii) Schreibe das zum Experiment gehörige Wahrscheinlichkeitsmaß P als abzählbare Linearkombination von Diracmaßen, vgl. Satz 2.5.(iv).

iii) Beschreibe die beiden Ereignisse

$A :=$ "Man muß die Münze mindestens sechsmal werfen." und
$B :=$ "Man braucht eine ungerade Zahl von Würfen."

als Elemente der gewählten σ-Algebra und berechne ihre Wahrscheinlichkeiten.

12. (Zur Maßtheorie). Sei $(\Omega; \mathscr{A}; P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable (also X meßbar). Definiere $u := \inf X(\Omega)$, $v := \sup X(\Omega)$ (jeweils in $\overline{\mathbb{R}}$) und betrachte die Verteilungsfunkion

$$g : \mathbb{R} \to \mathbb{R}, \quad g(t) = P(\{\omega \in \Omega : X(\omega) \leq t\})$$

von P sowie deren kleinste Rechtsinverse

$$\varphi : \mathbb{R} \to \overline{\mathbb{R}}, \quad \varphi(t) = \sup\{s \in \mathbb{R} : g(s) < t\}.$$

Hierfür gilt

$$\varphi(t) \begin{cases} = -\infty & \text{für } t \leq 0, \\ \in [u, v] & \text{für } 0 < t \leq 1, \\ = \infty & \text{für } t > 1. \end{cases}$$

Bezeichnet dann noch $Q := \lambda|[0;1]$ die Einschränkung des Lebesguemaßes auf das Einheitsintervall, so gilt

$$P(\{X \leq t\}) = Q(\{\varphi \leq t\}).$$

***13.** Jede linksstetige Funktion auf $\mathbb{R}$ mit Werten in einem beliebigen metrischen Raum ist meßbar (bezüglich der Borelschen σ-Algebren).

— Blatt 4 —

14. (Zwei Würfel). Zwei Laplace-Würfel werden unabhängig geworfen.

i) Beschreibe das Zufallsexperiment als Wahrscheinlichkeitsraum über $\Omega := \{1, \ldots, 6\}^2$ und beschreibe darauf die Zufallsvariablen

$$\begin{aligned} X &:= \text{“geworfene Augenzahl des ersten Würfels“}, \\ Y &:= \text{“geworfene Augenzahl des zweiten Würfels“ und} \\ S &:= \text{“Summe der geworfenen Augenzahlen“}. \end{aligned}$$

ii) Zeichne die Verteilungsfunktionen von X und S und bestimme jeweils Erwartungswert und Varianz von X und S.

15. Nach durchzechter Nacht soll ein Student der Wahrscheinlichkeitstheorie n mit den Zahlen 1 bis n numerierte Übungsblätter in einen Ordner sortieren. Für jedes Blatt an der richtigen (also Blatt i an der i-ten) Position im Ordner erhält er eine Mark. Leider kann er aber nur noch völlig willkürlich arbeiten. Reicht der erwartete Gewinn für ein Mensaessen bei $n = 4$ bzw. $n = 100$?
Hinweis: Additivität des Erwartungswerts.

16. (Gaußmaß bzw. Normalverteilung).
Seien $\alpha \in \mathbb{R}$ und $\sigma > 0$. Betrachte dazu $g_{\alpha,\sigma^2} : \mathbb{R} \to]0;\infty[$,

$$g_{\alpha,\sigma^2}(t) := \frac{1}{\sqrt{2\pi}\sigma}\, \mathrm{e}^{-\frac{(t-\alpha)^2}{2\sigma^2}}.$$

Dann nennt man $N(\alpha, \sigma^2) := g_{\alpha,\sigma^2}\lambda$ (also das Maß mit Dichte g_{α,σ^2} bezüglich des Lebesguemaßes λ) auf $\mathscr{B}(\mathbb{R})$ das (eindimensionale) *Gaußmaß* oder die *Normalverteilung zum Erwartungswert α und zur Varianz σ^2*. Benutze ohne Beweis, daß die $N(\alpha, \sigma^2)$ Wahrscheinlichkeitsmaße sind, also $N(\alpha, \sigma^2)(\mathbb{R}) = 1$ gilt (Vorlesung, §6). Zeige nun:

i) Für $N(\alpha, \sigma^2)$-integrierbares $f : \mathbb{R} \to \mathbb{R}$ gilt

$$\int f \, \mathrm{d}N(\alpha, \sigma^2) = \int f(\sigma t + \alpha) N(0,1)(\mathrm{d}t).$$

ii) $\int t \, N(\alpha, \sigma^2)(\mathrm{d}t) = \alpha$.

iii) $\int (t-\alpha)^2 N(\alpha, \sigma^2)(\mathrm{d}t) = \sigma^2$.

— Blatt 5 —

17. (Deterministische Zufallsvariablen).
Sei $(\Omega; \mathscr{A}; P)$ ein Wahrscheinlichkeitsraum. Dann gilt

$$\begin{aligned} \operatorname{Var} X = 0 &\iff X \ P\text{-f.s. konstant} \\ &\iff \exists x \in \mathbb{R} \text{ mit } P \circ X^{-1} = \delta_x \iff F_X(\mathbb{R}) \subset \{0;1\} \end{aligned}$$

(mit der Konvention $\operatorname{Var} X = \infty$, falls $X \notin L^2(P)$).

18. Sei $(\Omega; \mathscr{A}; P)$ ein Wahrscheinlichkeitsraum und $(X_n)_{n\in\mathbb{N}}$ eine Folge nichtnegativer Zufallsvariablen, die punktweise auf Ω gegen eine Funktion X konvergiert. Dann gilt

$$\operatorname{E} X \leq \liminf_{n\to\infty} \operatorname{E} X_n,$$

aber die Ungleichung kann echt sein (sogar bei bzgl. n konstantem $\operatorname{E} X_n < \infty$).

19. (Stetige Verteilung, 8 Punkte).
Ein Spieler wirft einen Dartpfeil an eine 2m hohe Bretterwand. Die Zufallsvariable

$$Y := \text{“Höhe des Treffers über dem Boden“}$$

sei auf dem Intervall $[0;2]$ absolutstetig nach dem Lebesguemaß verteilt mit Dichte

$$h : [0;2] \to \mathbb{R}_+, \quad h(x) = 3\left(\sqrt{1 - \frac{x}{2}} + \frac{x}{2} - 1\right).$$

i) Wähle einen dem Experiment angepaßten Wahrscheinlichkeitsraum.

ii) Berechne die Wahrscheinlichkeiten, eine bestimmte Höhe über dem Boden bzw. die untere Hälfte der Bretterwand bzw. die obersten 20cm zu treffen.

iii) Berechne $\operatorname{E} Y$ und $\operatorname{Var} Y$.

iv) Vergleiche die tatsächliche Wahrscheinlichkeit, weiter als 20cm vom Erwartungswert weg zu treffen, mit der durch Chebyshev gegebenen oberen Abschätzung hierfür.

***20.** (Korrelationskoeffizienten).
Sei $(\Omega; \mathscr{A}; P)$ ein Wahrscheinlichkeitsraum und $X, Y \in L^2(P)$ mit $\operatorname{Var} X > 0$, $\operatorname{Var} Y > 0$.
Zeige: Stets ist $\rho(X,Y) \in [-1;1]$ und es gilt

$$|\rho(X,Y)| = 1 \iff \exists \alpha, \beta \in \mathbb{R} \text{ mit } Y = \alpha X + \beta \ P\text{-f.s.}$$

Gib außerdem ein konkretes Beispiel dafür, daß X und X^2 unkorreliert sein können, also $\rho(X, X^2) = 0$.

— Blatt 6 —

21. (Stochastische vs. fast sichere Konvergenz).
Gib auf dem Wahrscheinlichkeitsraum $([0,1]; \mathscr{B}([0,1]); P := \lambda|_{[0,1]})$ eine Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen mit Werten in $\{0,1\}$ an, die P-stochastisch, aber nirgends punktweise (also erst recht nicht P-f.s.) gegen 0 konvergiert.
Für welche p konvergiert die Folge in $L^p(P)$?

22. (Poisson-Verteilung).
Sei $(\Omega; \mathscr{A}; P)$ ein Wahrscheinlichkeitsraum und darauf $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen derart, daß X_n jeweils $\pi_{1/n}$-verteilt ist. Dann konvergiert $X_n \to 0$ in $L^p(P)$ für alle $p \geq 1$ (insbesondere stochastisch).
Zeige die stochastische Konvergenz auch direkt durch eine für alle $\varepsilon > 0$ simultane Abschätzung von $P\{|X_n| > \varepsilon\}$ nach oben.
Hinweis zum ersten Teil: Die Konvergenz von $\sum_k k^p/(n^k k!)$ gegen 0 mit $n \to \infty$ zeigt man mittels majorierter Konvergenz auf $(\mathbb{N}_0; \mathbb{P}(\mathbb{N}_0); \zeta)$. Die entsprechende Majorante erhält man durch $n = 1$ und ihre Integrierbarkeit aus einem Konvergenzkriterium für Reihen aus Analysis I.

23. (Konvergenz bzgl. absolutstetiger Maße, 8 Punkte).
Sei $(\Omega; \mathscr{A}; P)$ ein Wahrscheinlichkeitsraum und Q ein weiteres Wahrscheinlichkeitsmaß auf $\mathscr{A}$ mit $Q \ll P$. Betrachte eine Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen auf Ω und eine weitere Zufallsvariable X. Zeige:

i) Konvergiert $X_n \to X$ P-f.s., so auch Q-f.s. (1 Punkt).

ii) Konvergiert $X_n \to X$ P-stochastisch, so auch Q-stochastisch. Beweise dies:
i) mit Radon-Nikodym und Aufgabe 8 (3 Punkte).
ii) mittels Teilfolgen-Teilfolgen-Satz (2 Punkte).

iii) Gilt $Q = hP$ mit beschränkter Dichte $0 \leq h \leq c$, so folgt aus $X_n \to X$ in $L^p(P)$ auch die Konvergenz in $L^p(Q)$ (2 Punkte).

***24.** Ein Spieler würfelt solange (unabhängig), bis jede der Augenzahlen 1 bis 6 mindestens einmal gefallen ist. Bestimme den Erwartungswert der benötigten Anzahl von Würfen.

— Blatt 7 —

25. (Konvergenz bei diskreten Verteilungen, 8 Punkte).
Sei $(\Omega; \mathscr{A}; P)$ ein Wahrscheinlichkeitsraum und darauf $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen sowie X eine weitere Zufallsvariable derart, daß alle $P \circ X_n^{-1}$ und $P \circ X^{-1}$ auf $\mathbb{N}$ konzentriert sind.

i) Zeige die Äquivalenz folgender Aussagen, ohne den (unbewiesenen) Satz 6.11 zu benutzen:

  i) Für alle $k \in \mathbb{N}$ gilt $P \circ X_n^{-1}\{k\} \to P \circ X^{-1}\{k\}$.
  ii) $F_{X_n} \to F_X$ punktweise.
  iii) $X_n \to X$ in Verteilung.
  iv) $X_n \to X$ schwach bzgl. $C_b(\mathbb{R})$.
  v) $X_n \to X$ schwach bzgl. $C_c(\mathbb{R})$.
  vi) $X_n \to X$ schwach bzgl. $\mathscr{D}(\mathbb{R}) \equiv C_c^\infty(\mathbb{R})$.

  Hinweis: Benutze ohne Beweis: Sind $K \subset U \subset \mathbb{R}$ mit K kompakt, U offen, so gibt es $f \in \mathscr{D}(\mathbb{R})$ mit $f|_{U^c} = 0$, $f|_K = 1$.

ii) Dagegen fallen auch bei noch so spezieller Wahl des Wahrscheinlichkeitsraumes die Begriffe der stochastischen Konvergenz und der Konvergenz in Verteilung nicht zusammen. Zeige dazu:
Auf $(\{0,1\}; \mathbb{P}(\{0,1\}); \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1)$ gibt es Folgen von Zufallsvariablen, die in Verteilung, nicht aber stochastisch konvergieren.

26. Seien X, Y, Z Zufallsvariablen auf $(\Omega; \mathscr{A}; P)$.

i) Ist die Relation "stochastisch unabhängig" transitiv, d.h.: X, Y unabhängig, Y, Z unabhängig $\Longrightarrow$ X, Z unabhängig?

ii) Sind die beiden folgenden Aussagen äquivalent oder ist eine der beiden stärker?

  i) X, Y, Z unabhängig.
  ii) X, Y, Z paarweise unabhängig, also jedes der Paare (X, Y), (Y, Z), (X, Z) unabhängig.

27. In einer Kiste sind $n \geq 2$ Kugeln, fortlaufend numeriert mit $1, \ldots, n$. Zwei Kugeln werden nacheinander ohne Zurücklegen gezogen.

i) Beschreibe die Nummer der ersten bzw. zweiten gezogenen Kugel als Zufallsvariable X bzw. Y auf einem geeigneten Wahrscheinlichkeitsraum. Sind X und Y unabhängig?

ii) Berechne $\mathrm{Cov}(X, Y)$ und $\rho(X, Y)$.

***28.** i) (Komposition erhält Unabhängigkeit).
Sei $(\Omega; \mathscr{A}; P)$ ein Wahrscheinlichkeitsraum, $I \neq \emptyset$, sowie $X_i : (\Omega, \mathscr{A}) \to (S_i, \mathscr{S}_i)$ $\mathscr{A}$-$\mathscr{S}_i$-meßbar und $f_i : (S_i, \mathscr{S}_i) \to (R_i, \mathscr{R}_i)$ $\mathscr{S}_i$-$\mathscr{R}_i$-meßbar für alle $i \in I$.
Dann gilt: Ist die Familie $(X_i)_{i\in I}$ unabhängig, so auch die Familie $(f_i \circ X_i)_{i\in I}$.

ii) Sei $f : \mathbb{R} \to \mathbb{R}$ meßbar und X eine Zufallsvariable auf $(\Omega; \mathscr{A}; P)$. Genau dann sind X und $f \circ X$ unabhängig, wenn $f \circ X$ P-f.s. konstant ist.

— Blatt 8 —

29. (Reichhaltigkeit des Wahrscheinlichkeitsraums bei Unabhängigkeit).
Sei $(\Omega; \mathscr{A}; P)$ ein Wahrscheinlichkeitsraum und X, Y auf Ω Zufallsvariablen mit Verteilungen

$$P \circ X^{-1} = \sum_{k=1}^{m} p_k \delta_{x_k} \quad \text{und} \quad P \circ Y^{-1} = \sum_{l=1}^{n} q_l \delta_{y_l},$$

wobei $m, n \in \mathbb{N}$, $p_k > 0$ und $q_l > 0$ für sämtliche k, l sowie $x_1 < \ldots < x_m$ und $y_1 < \ldots < y_n$. Dann gilt:

i) Ist $|\Omega| = m \geq n$, so gibt es $f : \mathbb{R} \to \mathbb{R}$ meßbar mit $Y = f \circ X$.

ii) Sind dagegen X und Y stochastisch unabhängig, so folgt $|\Omega| \geq mn$ und $|\mathscr{A}| \geq 2^{mn}$.

iii) Will man eine Laplace-Münze r-mal unabhängig werfen, so ist (modulo meßbarer Bijektionen) $(\{0,1\}^r; \mathbb{P}(\{0,1\}^r))$ der kleinste Meßraum, auf dem man das Experiment modellieren kann.

30. (Faltung, 12 Punkte).

i) Das Faltungsprodukt zweier Wahrscheinlichkeitsmaße auf $(\mathbb{R}; \mathscr{B}(\mathbb{R}))$ ist ein Wahrscheinlichkeitsmaß (1 Punkt).

ii) Seien μ, ν endliche Maße auf $(\mathbb{R}; \mathscr{B}(\mathbb{R}))$. Dann gilt für jedes meßbare $f : \mathbb{R} \to \overline{\mathbb{R}}_+$ bzw. jedes $f \in L^1(\mu * \nu)$

$$\int f \mathrm{d}(\mu * \nu) = \int \mu(\mathrm{d}x) \int \nu(\mathrm{d}y) f(x+y) = \int \nu(\mathrm{d}y) \int \mu(\mathrm{d}x) f(x+y).$$

Insbesondere gilt $\mu * \nu(B) = \int \mu(B-y)\nu(\mathrm{d}y) = \int \nu(B-x)\mu(\mathrm{d}x)$ für jedes $B \in \mathscr{B}(\mathbb{R})$ (3 Punkte).

iii) Seien $a, b \in \mathbb{R}$. Berechne $\delta_a * \delta_b$ und $\delta_a * \lambda|_{[0,1]}$. Zeige $\lambda|_{[0,1]} * \lambda|_{[0,1]} = h\lambda$ mit

$$h(x) = \begin{cases} x, & 0 \leq x \leq 1, \\ 2 - x, & 1 < x \leq 2, \\ 0, & \text{sonst} \end{cases}$$

(5 Punkte).

iv) Als Beispiel sei $\mu := \frac{1}{6} \sum_{i=1}^{6} \delta_i$ die Verteilung der Augenzahl eines Laplace-Würfels. Berechne $\mu * \mu$ unter Verwendung der Rechenregeln aus Aufgabe 31. Welche Verteilung (vergleiche mit Aufgabe 14) gibt dieses Maß an? (3 Punkte).

***31.** (Kommutativität, Assoziativität und σ-Distributivität der Faltung von Maßen). Seien $(\mu_n)_{n \in \mathbb{N}_0}$ endliche Maße auf $(\mathbb{R}; \mathscr{B}(\mathbb{R}))$ mit $\sum_{n \in \mathbb{N}} \mu_n(\mathbb{R}) < \infty$. Dann gilt:

i) $\mu_1 * \mu_2 = \mu_2 * \mu_1$.

ii) $\mu_1 * \mu_2 * \mu_3 = (\mu_1 * \mu_2) * \mu_3 = \mu_1 * (\mu_2 * \mu_3)$.

iii) $\mu_1 * (\alpha\mu_2) = (\alpha\mu_1) * \mu_2 = \alpha(\mu_1 * \mu_2)$, $\alpha \geq 0$, und $\mu_0 * (\sum_{n=1}^{\infty} \mu_n) = \sum_{n=1}^{\infty} \mu_0 * \mu_n$.

— Blatt 9 —

32. Sei $(\Omega; \mathscr{A}; P)$ ein Wahrscheinlichkeitsraum.

i) (Terminale Funktionale).
Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen auf Ω. Sind folgende Abbildungsvorschriften terminale Funktionale?

  i) $(X_n)_{n \in \mathbb{N}} \mapsto \sum_{n=1}^{\infty} |X_n|$.

  ii) $(X_n)_{n \in \mathbb{N}} \mapsto 1_A$ mit $A := \{\sum_{n=1}^{\infty} |X_n| = \infty\}$.

ii) ("Fehlende Richtung" bei Borel-Cantelli).
Gib ein Beispiel für $(A_n)_{n \in \mathbb{N}} \subset \mathscr{A}$ derart, daß $\sum_{n \in \mathbb{N}} P(A_n) = \infty$, aber $P(\limsup A_n) = 0$.

33. Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von identisch verteilten Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega; \mathscr{A}; P)$ mit $X_1 \in L^1(P)$.
Bestimme $P\{|X_n| \leq n$ für fast alle $n \in \mathbb{N}\}$.

Hinweis: Satz 4.9.

34. Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge unabhängiger Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega; \mathscr{A}; P)$ mit $P\{X_n = 0\} = p > 0$ unabhängig von n.

Berechne $P\{X_n = X_{n+1} = 0$ für unendlich viele $n\}$.

35. (Schwaches Gesetz der großen Zahlen).
Sei $(\Omega; \mathscr{A}; P)$ ein Wahrscheinlichkeitsraum und $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Zufallsvariablen der Verteilung

$$P \circ X_n^{-1} = \frac{1}{2} f(n)\delta_{-n} + (1 - f(n))\delta_0 + \frac{1}{2} f(n)\delta_n,$$

wobei $f(n) := \frac{1}{n \ln(n+2)}$, $n \in \mathbb{N}$.

Dann genügt (X_n) dem schwachen Gesetz der großen Zahlen.

— Blatt 10 —

36. (Schwaches Gesetz der großen Zahlen).
Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von integrierbaren Zufallsvariablen auf $(\Omega; \mathscr{A}; P)$ mit $EX_n = 0$ für alle n.

i) Genügt (X_n) dem schwachen Gesetz der großen Zahlen (also $(S_n - ES_n)/n$ konvergiert gegen 0 stochastisch), so konvergiert $\frac{1}{n} X_n \to 0$ stochastisch.

ii) Bilden die X_n eine Orthonormalfolge in $L^2(P)$, so genügt (X_n) dem schwachen Gesetz der großen Zahlen.

37. (Starkes Gesetz der großen Zahlen).
Sei $(\Omega; \mathscr{A}; P)$ ein Wahrscheinlichkeitsraum.

i) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Zufallsvariablen auf Ω derart, daß X_n exponentialverteilt zum Parameter $\sqrt{n}$ ist. Dann gilt

$$\frac{1}{n} \sum_{i=1}^{n} X_i \longrightarrow 0 \quad P\text{-f.s.}$$

ii) Genügt eine Folge unabhängiger integrierbarer Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ auf Ω mit $EX_n = 0$, $n \in \mathbb{N}$, dem starken Gesetz der großen Zahlen (d.h. $(S_n - ES_n)/n \to 0$ fast sicher), so gilt für jedes $\varepsilon > 0$

$$\sum_{n=1}^{\infty} P\left\{ \frac{1}{n} |X_n| \geq \varepsilon \right\} < \infty.$$

38. (Fouriertransformierte der Normalverteilung).
Seien $\alpha \in \mathbb{R}$, $\sigma > 0$. Dann gilt für die Fouriertransformierte von $N(\alpha, \sigma^2)$

$$N(\alpha, \sigma^2)^\frown(x) := \int \mathrm{e}^{\mathrm{i}xt} N(\alpha, \sigma^2)(\mathrm{d}t) = \mathrm{e}^{\mathrm{i}\alpha x - \frac{\sigma^2 x^2}{2}}.$$

Anleitung: Reduktion auf $\alpha = 0$, $\sigma = 1$. Zeige dann: $N(0,1)^\frown$ ist die eindeutige Lösung der linearen DGL $h'(x) = -xh(x)$ mit $h(0) = 1$.

***39.** Seien X, Y unabhängige Zufallsvariablen auf $(\Omega; \mathscr{A}; P)$.

i) Ist Y integrierbar mit $EY = 0$, so folgt $E|X+Y| \geq E|X|$.

ii) Ist $X + Y$ integrierbar, so sind auch X und Y integrierbar.

— Blatt 11 —

40. Sei $(X_n)_{n\in\mathbb{N}}$ eine i.i.d.-Folge von $L^2(P)$-Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega; \mathscr{A}; P)$ mit $\sigma := \sqrt{\operatorname{Var} X_1} > 0$. Dann konvergiert die standardisierte Partialsummenfolge

$$S_n^* := \frac{S_n - ES_n}{\sqrt{n}\,\sigma}$$

nicht stochastisch.
Hinweis: Der Grenzwert wäre fast sicher konstant im Widerspruch zum zentralen Grenzwertsatz.

41. (Anwendung von Berry-Esséen). Gib mittels des Satzes von Berry-Esséen eine Näherung der Wahrscheinlichkeit, bei 600 Würfen mit einem Laplace-Würfel mindestens 90, aber höchstens 100 Sechsen zu erhalten. Benutze dabei eine Tabelle der Verteilungsfunktion der Standard-Normalverteilung, z.B. in [9], Seite 240. Gib auch den dabei maximal begangenen Fehler der Abschätzung an.

– Wiederholungsaufgaben zur Klausur –

(freiwillig abzugeben, ohne Bewertung)

42. (Integrieren nach diskreten Maßen).

i) Sei $(\Omega; \mathscr{A})$ ein Meßraum sowie $f : \Omega \to \mathbb{C}$ meßbar, $x \in \Omega$ beliebig. Dann gilt $\int f \mathrm{d}\delta_x = f(x)$.

ii) Seien $(\Omega; \mathscr{A})$ und f wie eben, weiter $(x_n)_{n\in\mathbb{N}}$ eine Folge in Ω sowie $(p_n)_{n\in\mathbb{N}}$ Folge in $\mathbb{R}_+$ mit $\sum_{n\in\mathbb{N}} p_n = 1$. Dann ist $\mu := \sum_{n\in\mathbb{N}} p_n \delta_{x_n}$ ein Wahrscheinlichkeitsmaß auf $\mathscr{A}$ und es gilt, falls $f \geq 0$ oder $f \in L^1(\mu)$:

$$\int f \mathrm{d}\mu = \sum_{n\in\mathbb{N}} f(x_n) p_n.$$

43. In einer Urne liegt eine rote Kugel. Ein Spieler zieht aus der Urne unendlich oft (und unabhängig) jeweils eine Kugel mit Zurücklegen. Nach jedem Zug werden zusätzlich (zum Zurücklegen) schwarze Kugeln in die Urne gegeben, und zwar

i) immer eine schwarze Kugel,

ii) nach dem n-ten Zug jeweils $2n + 1$ schwarze Kugeln.

Bestimme in beiden Fällen die Wahrscheinlichkeit von “Der Spieler zieht unendlich oft die rote Kugel“.

44. (Unabhängige Zufallsvariable zu gegebenen Verteilungen).
Seien $\mu_1, \ldots, \mu_n$ endlich viele Wahrscheinlichkeitsmaße auf $\mathscr{B}(\mathbb{R})$. Dann gibt es stets einen Wahrscheinlichkeitsraum $(\Omega; \mathscr{A}; P)$ und darauf unabhängige Zufallsvariable $X_1, \ldots, X_n$ derart, daß $P \circ X_i^{-1} = \mu_i$, $1 \leq i \leq n$.

Anhang C. Klausur

Es sind nur fünf der sechs vorgeschlagenen Aufgaben zu bearbeiten! Bei Abgabe ist dieses Angabenblatt mit beizufügen, worauf die nicht bearbeitete Aufgabe durchgestrichen sein muß.

Bearbeitungszeitraum: 180 Min.

1. Betrachte den Maßraum $(\mathbb{R}; \mathscr{B}(\mathbb{R}); \mu)$ mit $\mu := \frac{3}{8}\delta_{-1} + \frac{1}{2}\lambda|_{[0,1]} + \frac{1}{8}\delta_1$ (λ bezeichne das Lebesguemaß auf $\mathscr{B}(\mathbb{R})$). Zeige:

i) $(\mathbb{R}; \mathscr{B}(\mathbb{R}); \mu)$ ist ein Wahrscheinlichkeitsraum.

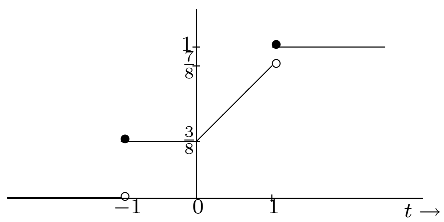
ii) Zeichne die Verteilungsfunktion von $\mathrm{id}_{\mathbb{R}}$.

iii) Berechne $E\,\mathrm{id}_{\mathbb{R}}$.

iv) Es gilt $\delta_1 \ll \mu$ (in Worten: δ_1 ist absolutstetig nach μ), aber nicht $\mu \ll \lambda$.

Lösung. i) Zu zeigen ist nur $\mu(\mathbb{R}) = 1$, aber $\mu(\mathbb{R}) = \frac{3}{8}\delta_{-1}(\mathbb{R}) + \frac{1}{2}\lambda|_{[0,1]}(\mathbb{R}) + \frac{1}{8}\delta_1(\mathbb{R}) = \frac{3}{8} + \frac{1}{2} + \frac{1}{8} = 1$.

ii)



iii) $E\,\mathrm{id}_{\mathbb{R}} = \int_{\mathbb{R}} x\mu(\mathrm{d}x) = \frac{3}{8}\int_{\mathbb{R}} x\delta_{-1}(\mathrm{d}x) + \frac{1}{2}\int_0^1 x\mathrm{d}x + \frac{1}{8}\int_{\mathbb{R}} x\delta_1(\mathrm{d}x) = \frac{3}{8}(-1) + \frac{1}{2}\frac{1}{2} + \frac{1}{8}1 = 0$.

iv) Aus $\mu(\{1\}) = \frac{1}{8} > 0$ folgt für $A \in \mathscr{B}(\mathbb{R})$ mit $\mu(A) = 0$, daß $A \subset \mathbb{R} \setminus \{1\}$, somit $\delta_1(A) = 0$. Nach Definition der Absolutstetigkeit von Maßen folgt $\delta_1 \ll \mu$. $\mu \not\ll \lambda$ ist direkte Konsequenz von $\mu(\{1\}) = \frac{1}{8} > 0 = \lambda(\{1\})$.

□

2. Um einen runden Tisch gibt es sechs Plätze mit Platzziffern 1-6. Sechs Gäste A-F ziehen per Los je eine Platzziffer und setzen sich nach dieser Ordnung. Mit welcher Wahrscheinlichkeit sitzen die Gäste A und B nebeneinander? (Zur Lösung gehört die Angabe des verwendeten Wahrscheinlichkeitsraums!)

Lösung. Beispielsweise wähle $\Omega = S_6$ die symmetrische Gruppe der Permutationen einer sechselementigen Menge, $\#\Omega = 6!$; dazu $\mathscr{A} = \mathbb{P}(\Omega)$ und $P = \frac{1}{6!}\sum_{\sigma\in S_6}\delta_\sigma$ (identifiziere Permutation σ mit Sitzordnung so: A-F $\cong$ 1-6, $\sigma(i)$ ist die Platzziffer von Gast i).

Günstige Fälle (A sitzt neben B): Es gibt sechs (gleichwahrscheinliche) Möglichkeiten, wo A und B nebeneinander sitzen können (der linke der beiden sitzt auf Platz i, $1 \le i \le 6$) und dann gibt es $2! = 2$ Möglichkeiten, wie beide auf diese Stühle verteilt sind, und 4! Möglichkeiten, wie die übrigen Gäste auf den anderen vier Plätzen sitzen. Somit

$$P\{\text{A sitzt neben B}\} = \frac{6\cdot 2!\cdot 4!}{6!} = \frac{1\cdot 2\cdot 3\cdot 4\cdot 2\cdot 6}{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6} = \frac{2}{5}.$$

□

3. Sei $f : [-1,1] \to \mathbb{R}$ gegeben durch $f(x) = 1-x^2$. Dazu betrachte $h := \frac{f}{\int_{-1}^1 f(x)\mathrm{d}x}$.

i) Gib einen Wahrscheinlichkeitsraum $(\Omega; \mathscr{A}; P)$ und darauf eine Zufallsvariable X an, so daß X gemäß $h\lambda|_{[-1,1]}$ verteilt ist.

ii) Berechne EX und $\operatorname{Var} X$.

iii) Schätze mit Chebychev die Wahrscheinlichkeit $P\{X \notin [-0.9, 0.9]\}$ ab.

Lösung. $\int_{-1}^1 f(x)\mathrm{d}x = x - \frac{1}{3}x^3|_{-1}^1 = \frac{4}{3}$.

i) Kanonische Wahl ist $\Omega = [-1,1]$, $\mathscr{A} = \mathscr{B}([-1,1])$, $P = h\lambda|_{[-1,1]}$ sowie $X = \mathrm{id}_{[-1,1]}$.

ii) $EX = \int_{-1}^1 x\frac{3}{4}(1-x^2)\mathrm{d}x = \frac{3}{4}\int_{-1}^1(x-x^3)\mathrm{d}x = 0$ wegen Punktsymmetrie des Integranden.

$\operatorname{Var} X = \int_{-1}^1(x-0)^2\frac{3}{4}(1-x^2)\mathrm{d}x = \frac{3}{4}\left(\frac{x^3}{3} - \frac{x^5}{5}\right)\Big|_{-1}^1 = \frac{3}{4}(\frac{2}{3} - \frac{2}{5}) = \frac{1}{5}$.

iii) $P\{|X - EX| > 0.9\} \le \frac{\operatorname{Var} X}{(0.9)^2} = \frac{1}{5}\frac{100}{81} = \frac{20}{81} \approx 0.2469$.

□

4. Sei $(\Omega; \mathscr{A}; P)$ ein Wahrscheinlichkeitsraum und darauf $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Zufallsvariablen mit Verteilung ($p \in [0,1]$)

i) $P \circ X_n^{-1} = B(1,p)$,

ii) $P \circ X_n^{-1} = B(n,p)$.

Berechne in beiden Fällen $P\{X_n = 0 \text{ für unendlich viele } n\}$. In welchem Fall kann man auf Unabhängigkeit verzichten?

Lösung. Borel-Cantelli-Schluß: $A_n := \{X_n = 0\} = X_n^{-1}(\{0\})$. Es ist $P(A_n) = (1-p)$ im Falle a), $P(A_n) = (1-p)^n$ im Falle b). Hieraus berechnet man

$$\text{a)}\ \sum_n P(A_n) = \sum_n (1-p) \begin{cases} < \infty, & p = 1, \\ = \infty, & p < 1, \end{cases} \quad \text{sowie}$$

$$\text{b)}\ \sum_n P(A_n) = \sum_n (1-p)^n \begin{cases} < \infty, & p > 0, \\ = \infty, & p = 0. \end{cases}$$

Folglich ist

$$\begin{aligned} &P\{X_n = 0 \text{ für unendlich viele } n\} \\ &\qquad = P(\limsup A_n) = \begin{cases} 0, & \text{a) mit } p = 1,\ \text{b) mit } p > 0, \\ 1, & \text{a) mit } p < 1,\ \text{b) mit } p = 0. \end{cases} \end{aligned}$$

Im Falle $p \in \{0,1\}$ ist das Experiment deterministisch und die X_n sowieso unabhängig. Im Fall b) mit $0 < p < 1$ gilt das Ergebnis auch für abhängige X_n.

□

5. Seien μ, ν Wahrscheinlichkeitsmaße auf den Borelmengen $\mathscr{B}(\mathbb{R})$ von $\mathbb{R}$ derart, daß $\mathrm{id}_{\mathbb{R}} \in L^2(\mu) \cap L^2(\nu)$.
Setze $\alpha := \int \mathrm{id}_{\mathbb{R}}\,\mathrm{d}\mu$, $\beta := \int \mathrm{id}_{\mathbb{R}}\,\mathrm{d}\nu$. Zeige

$$\int (x - (\alpha+\beta))^2 (\mu * \nu)(\mathrm{d}x) = \int (x-\alpha)^2 \mu(\mathrm{d}x) + \int (y-\beta)^2 \nu(\mathrm{d}y).$$

Hinweis: Entweder man rechnet dies direkt über die Definition der Faltung nach, dann ist es vorteilhaft, beide Seiten auf $\int x^2 \mu(\mathrm{d}x) - \alpha^2 + \int y^2 \nu(\mathrm{d}y) - \beta^2$ zu vereinfachen.
Oder man betrachtet μ und ν als die jeweilige Verteilung zweier unabhängiger (!) Zufallsvariablen X und Y auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega; \mathscr{A}; P)$ (welchem?) und übersetzt die Integrale der Behauptung in wahrscheinlichkeitstheoretische Größen bzgl. der X und Y.

Lösung. 1. Möglichkeit: Linke Seite: $\int (x-\alpha)^2 \mu(\mathrm{d}x) = \int x^2 - 2\alpha x + \alpha^2 \mu(\mathrm{d}x) = \int x^2 \mu(\mathrm{d}x) - 2\alpha \int x \mu(\mathrm{d}x) + \alpha^2 \int \mu(\mathrm{d}x) = \int x^2 \mu(\mathrm{d}x) - 2\alpha\alpha + \alpha^2 1 = \int x^2 \mu(\mathrm{d}x) - \alpha^2$.
Analog $\int (y-\beta)^2 \nu(\mathrm{d}y) = \int y^2 \nu(\mathrm{d}y) - \beta^2$.

Rechte Seite: $\int (x - (\alpha+\beta))^2 (\mu*\nu)(\mathrm{d}x) = \int ((x+y) - (\alpha+\beta))^2 \mu(\mathrm{d}x) \otimes \nu(\mathrm{d}y)$.
Nach Tonelli berechnet man dies zu

$$\begin{aligned} &\int \mu(\mathrm{d}x) \int \nu(\mathrm{d}y)((x+y)^2 - 2(\alpha+\beta)(x+y) + (\alpha+\beta)^2) \\ &= \int \mu(\mathrm{d}x) \int \nu(\mathrm{d}y)(x^2 + 2xy + y^2) - 2(\alpha+\beta) \int \mu(\mathrm{d}x) \int \nu(\mathrm{d}y)(x+y) \end{aligned}$$

$$
\begin{aligned}
&+ (\alpha+\beta)^2 \int \mu(\mathrm{d}x) \int \nu(\mathrm{d}y) \\
&= \int \mu(\mathrm{d}x) \left(x^2 + 2x\beta + \int y^2 \nu(\mathrm{d}y) \right) - 2(\alpha+\beta) \int \mu(\mathrm{d}x)(x+\beta) + (\alpha+\beta)^2 \\
&= \int x^2 \mu(\mathrm{d}x) + 2\alpha\beta + \int y^2 \nu(\mathrm{d}y) - 2(\alpha+\beta)(\alpha+\beta) + (\alpha+\beta)^2 \\
&= \int x^2 \mu(\mathrm{d}x) - \alpha^2 + \int y^2 \nu(\mathrm{d}y) - \beta^2.
\end{aligned}
$$

2. Möglichkeit: $\Omega = \mathbb{R}^2$, $\mathscr{A} = \mathscr{B}(\mathbb{R}^2)$, $P = \mu \otimes \nu$, $X = \mathrm{pr}_1$, $Y = \mathrm{pr}_2$ auf Ω. Dann sind X, Y unabhängig mit $P \circ X^{-1} = \mu$, $P \circ Y^{-1} = \nu$, $P \circ (X+Y)^{-1} = \mu * \nu$. Nach dem Transformationslemma gilt: $EX = \int \mathrm{id}_{\mathbb{R}} \,\mathrm{d}P \circ X^{-1} = \int \mathrm{id}_{\mathbb{R}} \,\mathrm{d}\mu = \alpha$, $EY = \beta$.
Weiterhin ist $\int (x-\alpha)^2 \mu(\mathrm{d}x) = \int (\mathrm{id}_{\mathbb{R}} - EX)^2 \mathrm{d}P \circ X^{-1} = \int (X - EX)^2 \mathrm{d}P = \operatorname{Var} X$ und völlig analog $\int (y-\beta)^2 \nu(\mathrm{d}y) = \operatorname{Var} Y$.
Die Behauptung folgt dann mittels $\int (x - (\alpha+\beta))^2 (\mu * \nu)(\mathrm{d}x) = \int (\mathrm{id}_{\mathbb{R}} - E(X+Y))^2 \mathrm{d}P \circ (X+Y)^{-1} = \int ((X+Y) - E(X+Y))^2 \mathrm{d}P = \operatorname{Var}(X+Y) = \operatorname{Var} X + \operatorname{Var} Y$, da X, Y unabhängig. □

6. Sei auf $(\Omega; \mathscr{A}; P)$ eine i.i.d.-Folge $(X_n)_{n \in \mathbb{N}}$ von Zufallsvariablen gegeben mit $P \circ X_1^{-1} = \pi_1$ (Poisson-Verteilung zum Parameter 1).
Zeige, daß alle folgenden Gesetze auf die Folge anwendbar sind und formuliere die resultierenden Aussagen für (X_n), dabei sind alle auftauchenden Größen bzgl. der X_n soweit als möglich konkret zu berechnen.

i) Schwaches Gesetz der großen Zahlen.
ii) Zweites starkes Gesetz der großen Zahlen.
iii) Zentraler Grenzwertsatz (de Moivre-Laplace).
iv) Satz von Berry-Esséen (die Konvergenz von $\sum k^3/k!$ folgt aus dem Quotientenkriterium, das dritte Moment muß aber nicht explizit berechnet werden).

Lösung. Aus i.i.d. und der Verteilung von X_1 folgt $EX_n = 1$, $\operatorname{Var} X_n = 1$, $ES_n = n$, $\operatorname{Var} S_n = n$, $n \in \mathbb{N}$. Somit:

i) SwGgZ: $\frac{1}{n^2} \sum_1^n \operatorname{Var} X_i = \frac{1}{n^2} n = \frac{1}{n} \to 0$, also sagt das SwGgZ: $\frac{S_n - n}{n} \to 0$ bzw. $\frac{S_n}{n} \to 1$ stochastisch.
ii) 2.StGgZ: $X_1 \in L^1(P)$, da $E|X_1| = EX_1 = 1 < \infty$ (X_n f.s. Werte in $\mathbb{N}_0$). Somit $\frac{S_n}{n} \to 1$ f.s.
iii) ZGws: $X_n \in L^2(P)$, da $\|X_n - 1\|_2^2 = \|X_n - EX_n\|_2^2 = \operatorname{Var} X_n = 1 < \infty$, also $X_n - 1 \in L^2(P)$, also auch X_n. Gleichzeitig folgt aus $\operatorname{Var} X_n > 0$, daß X_n nicht konstant ist. Folglich nach ZGws: $S_n^* = \frac{S_n - n}{\sqrt{n}} \to N(0,1)$ in Verteilung.

iv) Berry-Esséen: $X_1 \in L^3(P)$, da $\int |X_1^3| \mathrm{d}P = \sum_{k\in\mathbb{N}_0} k^3 \mathrm{e}^{-1}\frac{1}{k!} = \frac{1}{\mathrm{e}}\sum_0^\infty \frac{k^3}{k!} < \infty$ (Quotientenkriterium). Also gilt $\sup_{t\in\mathbb{R}} |F_{S_n^*}(t) - \Phi(t)| \leq \frac{6}{\sqrt{n}} E(|X_1 - 1|^3)$.

$\square$

Anhang D. Verwendete Maple-Befehle

Im folgenden werden die Maple-Befehlszeilen wiedergegeben, welche zur Erzeugung der Abbildungen verwendet wurden. Dies soll keine Einführung in die Anwendung von Maple sein, sondern ist für diejenigen gedacht, welche in Maple bereits Erfahrung besitzen. Die verwendete Version ist Maple V.4.

a) Beispiele von W-Verteilungen:

```
with(stats):
with(plots):

plot(statevalf[pdf,normald[0,1]], -4..4, view=[-4..4,0..0.6]);
plot(statevalf[cdf,normald[0,1]], -4..4, view=[-4..4,0..1.1]);

plot(statevalf[pdf,exponential[1,0]], -4..4);
plot(statevalf[cdf,exponential[1,0]], -4..5);

plot(statevalf[pdf,cauchy[0,1]], -4..4);
plot(statevalf[cdf,cauchy[0,1]], -4..4);

plot(statevalf[pdf,uniform[0,2]], -1..3);
plot(statevalf[cdf,uniform[0,2]], -1..3);

histogram([seq(Weight(n-0.5..n+0.5,
  statevalf[pf,poisson[3.5]](n)), n=0..15)]);
histogram([seq(Weight(n..n+1,
  statevalf[dcdf,poisson[3.5]](n)), n=0..15)]);

PLOT(POLYGONS(seq([[n,0],[n,statevalf[pf,poisson[3.5]](n)]],n=0..15)),
  THICKNESS(3)) ;
PLOT(POLYGONS(seq([[n,statevalf[dcdf,poisson[3.5]](n)],
  [n+1,statevalf[dcdf,poisson[3.5]](n)]],n=0..15)),
  POINTS((seq([n,statevalf[dcdf,poisson[3.5]](n)], n=0..15)),
   SYMBOL(POINT)) );

histogram([seq(Weight(n-0.5..n+0.5,
  statevalf[pf,binomiald[10,0.3]](n)), n=0..10)]);
histogram([seq(Weight(n..n+1,
  statevalf[dcdf,binomiald[10,0.3]](n)), n=0..10)]);

PLOT(POLYGONS(seq([[n,0],[n,statevalf[pf,binomiald[10,0.3]](n)]],
  n=0..10)), THICKNESS(3)) ;
```

```
PLOT(POLYGONS(seq([[n,statevalf[dcdf,binomiald[10,0.3]](n)],
  [n+1,statevalf[dcdf,binomiald[10,0.3]](n)]],n=0..10)),
   POINTS((seq([n,statevalf[dcdf,binomiald[10,0.3]](n)], n=0..10)),
   SYMBOL(POINT)) );

plot(statevalf[pdf,chisquare[5]], 0..15);
plot(statevalf[cdf,chisquare[5]], 0..15);

plot(statevalf[pdf,studentst[4]], -5..5);
plot(statevalf[cdf,studentst[4]], -5..5);
```

b) Normalverteilungen unkorreliert und korreliert:

```
rho := 0 ;
f1 := sqrt(1-rho^2)/(2*Pi) * exp(-0.5*(x^2-2*rho*x*y+y^2));
plot3d(f1,x=-2..2,y=-2..2,style=patch,shading=zgreyscale,
 orientation=[-68,77], axes=frame) ;

rho := 0.8 ;
f2 := sqrt(1-rho^2)/(2*Pi) * exp(-0.5*(x^2-2*rho*x*y+y^2));
plot3d(f2,x=-2..2,y=-2..2,style=patch,shading=zgreyscale,
 orientation=[-68,77], axes=frame) ;

rho := -0.8 ;
f3 := sqrt(1-rho^2)/(2*Pi) * exp(-0.5*(x^2-2*rho*x*y+y^2));
plot3d(f3,x=-2..2,y=-2..2,style=patch,shading=zgreyscale,
 orientation=[-68,77], axes=frame) ;
```

c) Beispiele für die Monte-Carlo-Methode:

```
with (stats):
f := x -> sqrt(1-x^2) ;
f := x -> 1/(1+x) ;

c := evalf(int(f,0..1)) ;

N := 100 :
S := array[1..N] :
summe := array[1..N] :
S := random[uniform[0,1]](N):
summe[1] := f(S[1]) :
for n from 2 to N do
```

```
  summe[n] := ( (n-1)*summe[n-1] + f(S[n])  ) / n :
od :

b := [seq([n,summe[n]], n=1..N)]:
PLOT(FONT(TIMES,ROMAN,9),CURVES(b),POLYGONS([[0,c],[N,c]]) );

for n from 1 to N do
  s := 0 :
  for j from 1 to n do
    s := s + evalf(f((j-1/2)/n)) :
  od :
  summe[n] := s / n :
 od:

b := [seq([n,summe[n]], n=1..N)]:
PLOT(FONT(TIMES,ROMAN,9),CURVES(b),POLYGONS([[0,c],[N,c]]) );

m := 1 :
n := 6 :
w := [41,216,27,272,27,216,41] :
summe := 0 :
for i from 0 to m-1 do
  h := evalf(1/(840*m)) :
  for j from 0 to n do
    summe := summe + evalf(h * w[j+1] * f((j + i*n)/(n*m)) ) :
  od :
od :
summe ; c;
```

d) Programme zum Zentralen Grenzwertsatz:

```
with(inttrans): with(stats): with(plots):
for n from 1 to 10 do
  a := sqrt(3/n) :
  f := x -> (sin(a*x)/(a*x))^n :
  p1 := plot(evalf(invfourier(f(x),x,w)) ,
      w=-5..5,style=line,numpoints=1000,thickness=3,color=black):
  p2 := plot(statevalf[pdf,normald[0,1]](w),
      w=-5..5,linestyle=2,numpoints=1000,thickness=3,color=black):
  print(display({p1,p2})) ;
od :

p := 0.6: q := 1-p:
for n from 1 to 10 do
```

```
  EX := n*p :
  VarX := sqrt(n*p*q):
  f := statevalf[pf,binomiald[n,p]] :
  p1 := PLOT(POLYGONS(seq([[(k-1/2-EX)/VarX, 0],
    [(k-1/2-EX)/VarX, VarX*f(k)]], k=0..n)),
    POLYGONS(seq([[(k+1/2-EX)/VarX, 0],
    [(k+1/2-EX)/VarX, VarX*f(k)]], k=0..n)),
    POLYGONS(seq([[(k-1/2-EX)/VarX, VarX*f(k)],
    [(k+1/2-EX)/VarX, VarX*f(k)]], k=0..n)), THICKNESS(3)):
  p2 := plot(statevalf[pdf,normald[0,1]](w), w=-5..5,
    linestyle=2,numpoints=1000,thickness=3,color=black):
  print(display({p1,p2})) :
od:
```

Abbildungsverzeichnis

Literaturverzeichnis

[1] H. Bauer: Maß- und Integrationstheorie. De Gruyter, Berlin 1990.

[2] H. Bauer: Wahrscheinlichkeitstheorie. 4. Auflage. De Gruyter, Berlin 1991.

[3] L. Breiman: Probability. Addison-Wesley, Reading 1968.

[4] K. L. Chung: A Course in Probability Theory. 2nd edition. Academic Press, New York 1974.

[5] P. Gänssler, W. Stute: Wahrscheinlichkeitstheorie. Springer, Berlin 1977.

[6] W. Hackenbroch: Integrationstheorie. Teubner, Stuttgart 1987.

[7] W. Hackenbroch, A. Thalmaier: Stochastische Analysis. Teubner, Stuttgart 1994.

[8] P. R. Halmos: Measure Theory. Van Nostrand Reinhold, New York 1969.

[9] U. Krengel: Einführung in die Wahrscheinlichkeitstheorie und Statistik. Vieweg, Braunschweig 1988.

[10] J. Lehn, H. Wegmann: Einführung in die Statistik. Teubner, Stuttgart 1985.

[11] J. C. Oxtoby: Maß und Kategorie. Springer, Berlin 1971.

[12] J. Pfanzagl: Elementare W-Theorie. 2. Auflage. De Gruyter, Berlin 1991.

[13] A. N. Shiryayev: Probability. Springer, New York 1984.