Stochastik I

Christian Bayer

12. Juli 2015

1 Wahrscheinlichkeitsräume

1.1 Mengensysteme

Sei Ω eine allgemeine, nichtleere Menge. Wir sprechen von der Menge aller möglichen Versuchsausgänge. Typisch in der Wahrscheinlichkeitstheorie sind auch folgende Sprechweisen, wobei A, B, A_n Ereignisse des Zufallsexperiment seien:

- $A \cup B$ bzw. $\bigcup_n A_n$ ist das Ereignis, dass A oder B bzw. eines der Ereignisse A_n eintritt;
- $A \cap B$, bzw. $\bigcap_n A_n$ ist das Ereignis, dass alle Ereignisse A, B bzw. A_n eintreten;
- $A^c \coloneqq \Omega \setminus A$ ist das Komplementärereignis von A, d.h., das Ereignis, dass A nicht eintritt;
- $A \setminus B = A \cap B^c$ ist das Ereignis, dass A aber nicht B eintritt;
- $A \Delta B \coloneqq (A \setminus B) \cup (B \setminus A)$ ist das Ereignis, dass entweder A oder B (aber nicht beide) eintreten;
- für $\omega \in \Omega$ heißt $\{\, \omega \,\}$ Elementarereignis.

Definition 1.1. Ein Mengensystem $\emptyset \neq \mathcal{F} \subset 2^\Omega$ heißt *Algebra* falls

(i) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$,

(ii) $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$.

Bemerkung 1.2. Für eine Algebra $\mathcal{F}$ folgt unmittelbar, dass $\emptyset \in \mathcal{F}$, $\Omega \in \mathcal{F}$, sowie $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$.

Definition 1.3. Ein Mengensystem $\emptyset \neq C \subset 2^\Omega$ heißt *monotone Klasse* falls

(i) für jede monoton wachsende Folge von Mengen $A_j \in C$ (d.h., $\forall j \in \mathbb{N}: \ A_j \subset A_{j+1}$, wir schreiben: $A_j \nearrow$) gilt $A \coloneqq \bigcup_{j\in\mathbb{N}} A_j \in C$ (wir schreiben: $A_j \nearrow A$);

(ii) für jede monoton fallende Folge von Mengen $A_j \in C$ (d.h., $\forall j \in \mathbb{N}: \ A_j \supset A_{j+1}$, wir schreiben: $A_j \searrow$) gilt $A \coloneqq \bigcap_{j\in\mathbb{N}} A_j \in C$ (wir schreiben: $A_j \searrow A$).

Definition 1.4. Ein Mengensystem $\emptyset \neq \mathcal{F} \subset 2^\Omega$ heißt *σ-Algebra* falls

(i) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$,

(ii) $A_j \in \mathcal{F}$, $j \in \mathbb{N}, \Rightarrow \bigcup_{j\in\mathbb{N}} A_j \in \mathcal{F}$.

Das Paar $(\Omega, \mathcal{F})$ heißt dann *Messraum*.

Satz 1.5. *Sei $\mathcal{F}$ eine Algebra. Dann ist $\mathcal{F}$ eine σ-Algebra genau dann wenn es eine monotone Klasse ist.*

Lemma 1.6. *Sei $\emptyset \neq \mathcal{F}_0 \subset 2^\Omega$ beliebig. Dann existieren eine kleinste (im Sinne der Inklusion von Mengen) σ-Algebra und eine kleinste monotone Klasse, die $\mathcal{F}_0$ enthalten. Wir bezeichnen diese σ-Algebra mit $\sigma(\mathcal{F}_0)$ und nennen sie „die von $\mathcal{F}_0$ erzeugte σ-Algebra“. Analog bezeichnen wir die monotone Klasse mit $MK(\mathcal{F}_0)$ und nennen sie „die von $\mathcal{F}_0$ erzeugte monotone Klasse“.*

Satz 1.7. *Sei $\mathcal{F}_0$ eine Algebra. Dann gilt $\sigma(\mathcal{F}_0) = MK(\mathcal{F}_0)$.*

1.2 Wahrscheinlichkeitsmaße

Definition 1.8. Sei $(\Omega, \mathcal{F})$ ein Messraum. Eine Abbildung $P : \mathcal{F} \to [0, 1]$ heißt *Wahrscheinlichkeitsmaß* (kurz: *W-Maß*), falls

(i) $P(\Omega) = 1$,

(ii) für jede Folge disjunkter Mengen $A_j \in \mathcal{F}$ (also $A_i \cap A_j = \emptyset$, $i \neq j$) gilt:

$$P\left(\bigcup_{j\in\mathbb{N}} A_j\right) = \sum_{j=1}^{\infty} P(A_j).$$

Die zweite Eigenschaft heißt *σ-Additivität*.

Falls $(\Omega, \mathcal{F})$ ein Messraum mit einem W-Maß P ist, dann nennen wir $(\Omega, \mathcal{F}, P)$ *Wahrscheinlichkeitsraum*.

Lemma 1.9. *Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum. Dann gelten für beliebige Mengen $A, B \in \mathcal{F}$ die folgenden Eigenschaften:*

a) $0 \leq P(A) \leq 1$;

b) $P(\emptyset) = 0$;

c) $P(A^c) = 1 - P(A)$;

d) $A \subset B \Rightarrow P(A) = P(B) - P(B \setminus A) \leq P(B)$;

e) $P(A\cup B)+P(A\cap B) = P(A)+P(B)$*, und allgemeiner für Ereignisse* $A_1, \ldots, A_n \in \mathcal{F}$ *die* Formel von Sylvester*:*

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{l=1}^{n} \left((-1)^{l-1} \sum_{\{k_1,\ldots,k_l\}\subset\{1,\ldots,n\}} P\left(\bigcap_{j=1}^{l} A_{k_j}\right)\right);$$

f) Stetigkeit von unten bzw. oben: *Geg. eine wachsende Folge $A_n \in \mathcal{F}$ bzw. eine fallende Folge $B_n \in \mathcal{F}$. Seien $A := \bigcup_{n=1}^\infty A_n$, $B := \bigcap_{n=1}^\infty B_n$. Dann gilt*

$$P(A) = \lim_{n\to\infty} P(A_n), \quad P(B) = \lim_{n\to\infty} P(B_n);$$

g) σ-Subadditivität bzw. Boole'sche Ungleichung: *für eine beliebige Folge $A_n \in \mathcal{F}$ gilt*

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

Bemerkung 1.10. Der Spezialfall der Stetigkeit von oben mit $B = \emptyset$ heißt auch *0-Stetigkeit*. Man kann zeigen: „Endliche Additivität + 0-Stetigkeit $\iff$ σ-Additivität".

Definition 1.11. Sei μ ein W-Maß auf $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Dann heißt die Funktion $F = F_\mu : \mathbb{R} \to \mathbb{R}$ definiert durch

$$F(x) := \mu\left(]-\infty, x]\right)$$

Verteilungsfunktion von μ.

Satz 1.12. *Die Verteilungsfunktion F eines W-Maßes μ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ hat die folgenden Eigenschaften:*

a) F ist monoton wachsend, $F(-\infty) := \lim_{x\to-\infty} F(x) = 0$, $F(+\infty) := \lim_{x\to\infty} F(x) = 1$;

b) F is rechts-stetig, d.h.,

$$\forall x \in \mathbb{R} : \lim_{\epsilon \searrow 0} F(x+\epsilon) = F(x);$$

c) für $a \leq b$ gilt $\mu(]a, b]) = F(b) - F(a)$;

d) F hat höchstens abzählbar unendlich viele Sprungstellen.

Satz 1.13. *Gegeben sei eine Funktion $F : \mathbb{R} \to \mathbb{R}$, die die Eigenschften a) und b) von Satz 1.12 erfüllt. Dann existiert ein eindeutiges W-Maß μ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, deren Verteilungsfunktion $F_\mu = F$ erfüllt.*

Definition 1.14. Sei $(\Omega, \mathcal{F})$ ein Messraum, μ ein (σ-endliches) Maßauf $\mathcal{F}$, $f : \Omega \to \mathbb{R}_{\geq 0}$ eine Borel-messbare Funktion mit

$$\int_\Omega f(\omega)\mu(d\omega) = 1.$$

Dann heißt f *Dichte* (bezüglich μ) des W-Maßes ν auf $\mathcal{F}$ definiert durch

$$\nu(A) := \int_\Omega \mathbf{1}_A(\omega) f(\omega)\mu(d\omega) = \int_A f(\omega)\mu(d\omega), \quad A \in \mathcal{F}.$$

2 Zufallsvariable und Momente

2.1 Zufallsvariable

Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$.

Definition 2.1. Sei $(E, \mathcal{E})$ ein Messraum. Eine $\mathcal{F}$-$\mathcal{E}$-messbare Funktion $X : \Omega \to E$ heißt *Zufallsvariable* (ZV). Wir betrachten fast ausschließlich die folgenden Fälle:

- $(E, \mathcal{E}) = \left(\mathbb{N}, 2^{\mathbb{N}}\right)$;
- $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (*reelle ZV*);
- $E = \overline{R} := \mathbb{R} \cup \{-\infty, +\infty\}$, $\mathcal{E} := \sigma(\mathcal{B}(\mathbb{R}) \cup \{\{-\infty\}, \{+\infty\}\})$ (*erweiterte reelle ZV*);
- $(E, \mathcal{E}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, $n \in \mathbb{N}$, (*Zufalllsvektor*);
- $(E, \mathcal{E}) = (\mathbb{C}, \mathcal{B}(\mathbb{C}))$ (*komplexe ZV*).

Bemerkung 2.2. $\{X \in A\} := X^{-1}(A) := \{\omega \mid X(\omega) \in A\}$. Dann gelten:

a) X ist $\mathcal{F}$-$\mathcal{E}$-messbar $\iff \forall A \in \mathcal{E} : X^{-1}(A) \in \mathcal{F}$;

b) X^{-1} vertauscht mit den Mengenoperationen der Komplementbildung sowie der Vereinigungs- und Durschnittsbildung;

c) $X : \Omega \to \mathbb{R}$ ist realle ZV genaud dann, wenn $\forall x \in \mathbb{Q} : \{X \in]-\infty x]\} \in \mathcal{F}$.

Schreibweise: $P(X \in A) := P(X^{-1}(A))$ und analog—z.B., $P(X \leq x) := P(X^{-1}(]-\infty, x]))$.

Definition 2.3. Geg. eine ZV $X : \Omega \to E$. Das Bildmaß

$$P \circ X^{-1}(A) := P_X(A) := P(X \in A), \quad A \in \mathcal{E},$$

heißt *Verteilung* von X. Für $E = \mathbb{R}$, nennen wir die Verteilungsfunktion F_{P_X} der Verteilung von X auch einfach *Verteilungsfunktion von* X und schreiben F_X.

Korollar 2.4. *$(E, \mathcal{E}, P_X)$ ist ein W-Raum.*

Definition 2.5. Sei $X : \Omega \to E$ eine Funktion. Wir bezeichnen die σ-Algebra

$$\sigma(X) := \sigma\left(\left\{ X^{-1}(A) \,\middle|\, A \in \mathcal{E} \right\}\right)$$

als die von X erzeugte σ-Algebra.

Satz 2.6. *Seien X, Y bzw. X_n, $n \in \mathbb{N}$, realle ZV.*

a) Ist $f : \mathbb{R} \to \mathbb{R}$ eine stetige (messbare) Funktion, dann ist $f \circ X$ eine reelle ZV.

b) Die folgenden Abbildungen sind reelle ZVen:

(i) $X \wedge Y := \min(X, Y)$ und $X \vee Y := \max(X, Y)$;

(ii) $X + Y$, $X - Y$, $X \cdot Y$, $\frac{X}{Y} \cdot \mathbf{1}_{Y \neq 0}$.

c) Die folgenden Abbildungen sind erweiterte reelle ZVen:

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

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

(v) $\lim_{n\to\infty} X_n \mathbf{1}_{\{\omega \mid \exists \lim_{n\to\infty} X_n \in \bar{R}\}}$.

Lemma 2.7 (Dichtetransformationsformel)**.** *Sei X eine n-dimensionale ZV mit stetiger Dichte f_X bezüglich des (n-dimensionalen) Lebesgue-Maßes (d.h., die Verteilung von X hat Dichte f_X). Sei $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ eine bijektive C^1-Funktion. Dann hat die ZV $Y := \varphi \circ X$ die Dichte*

$$f_Y(y) = \frac{f_X(\varphi^{-1}(y))}{\left|\det D\varphi(\varphi^{-1}(y))\right|}, \quad y \in \mathbb{R}^n.$$

Diese Formel ist die bekannte Subsititutionsregel aus der Analysis. Sie kann ausgedehnt werden auf Fälle, in denen φ lediglich bijektiv ist, wenn es auf eine offene Menge A mit $P(X \in A) = 1$ eingeschränkt wird.

2.2 Erwartungswert

Definition 2.8. Sei X eine reelle ZV.

a) Falls X P-integrierbar ist, also falls

$$\int_\Omega |X(\omega)|\, P(d\omega) < \infty,$$

definieren wir

$$E[X] := \int_\Omega X(\omega) P(d\omega).$$

$E[X]$ heißt *Erwartungswert von X*. Diese Definition kann in offensichtlicher Weise ausgedehnt werden auf nicht-integrierbare ZV X, wenn wenigstens

$$E[X_-] < \infty \text{ oder } E[X_+] < \infty.$$[1]

b) Falls X^k P-integrierbar ist für ein $k \in \mathbb{N}$, dann heißt $E\left[X^k\right]$ das $k$te Moment der ZV X.

c) Ist X^2 P-integrierbar, so definieren wir die *Varianz* von x durch

$$\mathrm{var}[X] := E[X^2] - E[X]^2 = E\left[(X - E[X])^2\right],$$

sowie die *Standardabweichung* $\sqrt{\mathrm{var}[X]}$.

Satz 2.9. *Alle im folgenden vorkommenden ZV seien integrierbar. Dann gelten die folgenden Rechenregeln für den Erwartungswert.*

a) Für $\alpha, \beta \in \mathbb{R}$ gilt $E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]$.

b) Ist X eine f.s. konstante ZV, also $P(X = \alpha) = 1$ für eine $\alpha \in \mathbb{R}$, dann gilt $E[X] = \alpha$.

c) Positivität: $X \geq 0$ f.s. impliziert $E[X] \geq 0$.

[1] D.h., wenn $E[X_-] < \infty$ und $E[X_+] = \infty$, setzen wir $E[X] = +\infty$, im umgekehrten Fall setzen wir $E[X] = -\infty$.

d) Monotonie: $X \leq Y$ f.s. impliziert $E[X] \leq E[Y]$.

e) Dreiecksungleichung: $|E[X]| \leq E[|X|]$.

f) Satz von Beppo-Levi: $X_n \nearrow X$, $E[(X_1)_-] < \infty$, dann gilt $E[X] = \lim_{n\to\infty} E[X_n]$.

g) Satz von Lebesgue: $X_n \to X$ f.s., und es gibt eine dominierende ZV Y mit $|X_n| \leq Y$ f.s., für alle $n \in \mathbb{N}$. Dann gilt $E[X] = \lim_{n\to\infty} E[X_n]$.

Satz 2.10. *Sei X eine ZV mit Werten in $\mathbb{N}$, so gilt*

$$E[X] = \sum_{n=1}^{\infty} P(X \geq n).$$

Bemerkung 2.11. Sei X eine reelle ZV. Ähnlich wie im obigen Satz gilt:

$$\sum_{n=0}^{\infty} P(|X| \geq n) \leq E[|X|] \leq 1 + \sum_{n=0}^{\infty} P(|X| \geq n),$$

in dem Sinn, dass X integrierbar ist genau dann wenn die obige Reihe konvergiert.

Bemerkung 2.12. Der Erwartungswert einer (reellen) ZV X hängt nur von der Verteilung P_X ab. Präziser gilt:

$$E[X] = \int_{\Omega} X(\omega)P(d\omega) = \int_{\mathbb{R}} xP_X(dx).$$

Ganz allgemein gilt in der Stochastik, dass man meist lediglich an der *Verteilung*, nicht an dem konkreten W-Raum $(\Omega, \mathcal{F}, P)$ und die konkrete ZV $X : \Omega \to \mathbb{R}$ interessiert ist.

Bemerkung 2.13. Sei X eine quadratintegrierbare reelle ZV. Dann gilt: $c = E[X]$ minimiert das Funktional

$$c \mapsto E\left[(X-c)^2\right].$$

2.3 Klassische Ungleichungen

Satz 2.14 (Markow'sche Ungleichung)**.** *Sei $\varphi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ eine monoton wachsende Funktion, $u \geq 0$ mit $\varphi(u) > 0$. Weiters sei X eine reelle ZV. Dann gilt*

$$P(|X| \geq u) \leq \frac{E[\varphi(|X|)]}{\varphi(u)}.$$

Insbesondere gilt für $0 < p < \infty$:

$$P(|X| \geq u) \leq \frac{E[|X|^p]}{u^p}.$$

Korollar 2.15 (Tschebyschow'sche Ungleichung)**.** *Sei X eine integrierbare reelle ZV, $\epsilon > 0$. Dann gilt*

$$P(|X - E[X]| \geq \epsilon) \leq \frac{\mathrm{var}[X]}{\epsilon^2}.$$

Satz 2.16 (Jensen'sche Ungleichung)**.** *Sei $\varphi : \mathbb{R} \to \mathbb{R}$ eine konvexe Funktion, X eine reelle ZV mit der Eigenschaft, dass X integrierbar ist. Dann gilt $E[\varphi(X)_-] < \infty$ und*

$$\varphi(E[X]) \leq E[\varphi(X)].$$

Bemerkung 2.17. Die Jensen'sche Ungleichung hat folgende Verallgemeinerungen:

a) Die Aussage von Satz 2.16 gilt wenn $\varphi : I \to \mathbb{R}$ konvex ist für ein Intervall I mit $P(X \in I) = 1$, d.h., φ muss nicht auf ganz $\mathbb{R}$ definiert sein.

b) Wenn X ein integrierbare n-dimensionaler Zufallsvektor ist (d.h., alle Komponenten sind integrierbar), $C \subset \mathbb{R}^n$ konvex mit $P(X \in C) = 1$, sowie $\varphi : C \to \mathbb{R}$ konvex, dann ist $E\left[\varphi(X)_-\right] < \infty$ und

$$\varphi\left(E[X_1], \ldots, E[X_n]\right) \leq E[\varphi(X)].$$

Lemma 2.18 (Eigenschaften konvexer Funktionen)**.** *Sei $\varphi : \mathbb{R} \to \mathbb{R}$ konvex. Dann gilt:*

(i) φ ist stetig (und damit Borel-messbar);

(ii) Für festes $x \in \mathbb{R}$ ist der Differenzenquotient

$$S_x(y) := \frac{\varphi(y) - \varphi(x)}{y - x}, \quad y \in \mathbb{R} \setminus \{ x \},$$

monoton wachsend in y und die links- bzw. rechts-seitigen Ableitungen

$$D^-\varphi(x) := \lim_{y \nearrow x} S_x(y) = \sup\{ S_x(y) \mid y < x \},$$
$$D^+\varphi(x) := \lim_{y \searrow x} S_x(y) = \inf\{ S_x(y) \mid y > x \},$$

existieren (in $\mathbb{R}$) und erfüllen $D^-\varphi(x) \leq D^+\varphi(x)$;

(iii) Für festes $x \in \mathbb{R}$ gilt:

$$(\forall y \in \mathbb{R} : \ \varphi(x) + (y - x)t \leq \varphi(y)) \iff t \in \left[D^-\varphi(x), D^+\varphi(x)\right].$$

2.4 Standardverteilungen

Wir geben zunächst einige wichtige Verteilungen für ZV X an, die Werte in $\mathbb{Z}$ annehmen.

(i) Sei A eine endliche Menge. Dann heißt eine ZV X mit Werten in A (diskret) *gleichverteilt* auf A, falls

$$\forall B \subset A : \ P(X \in B) = \frac{\#B}{\#A}.$$

(ii) Für $0 \leq p \leq 1$, heißt die Verteilung einer ZV mit $P(X = 1) = p$, $P(X = 0) = 1 - p$ *Bernoulliverteilung* mit Parameter p. Offenbar gilt $P_X = (1 - p)\delta_0 + p\delta_1$. (Unsymmetrischer Münzwurf)

(iii) Sei X eine ZV mit Werten in $\{ 0 \ldots, n \}$. Die Verteilung von X heißt *Binomialverteilung* mit Parameter n und p falls

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

(iv) Eine $\mathbb{N}_0$-wertige ZV X hat die *Poissonverteilung* mit Parameter $\lambda > 0$, falls

$$P(X = n) = e^{-\lambda}\frac{\lambda^n}{n!}.$$

Anschließend geben wir einige wichtige Verteilungen an, die durch eine Dichte gegeben sind.

Bemerkung 2.19. In der stochastischen Redeweise verstehen wir unter der *Dichte einer Zufallsvariable X* die Dichte der Verteilung P_X unter dem (ein- oder n-dimensionalen) Lebesguemaß.

(v) Die *Normalverteilung* mit *Mittelwert* $\mu \in \mathbb{R}$ und *Varianz* $\sigma^2 > 0$ hat die Dichte

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

Wir bezeichnen diese Verteilung mit $\mathcal{N}(\mu, \sigma^2)$.

(vi) Die n-dimensionale *Normalverteilung* ist gegeben durch die Parameter $\mu \in \mathbb{R}^n$ und $\Sigma \in \mathbb{R}^{n\times n}$, wobei wir verlangen, dass Σ positiv-definit und symmetrisch ist. Die Dichte der mit $\mathcal{N}(\mu, \Sigma)$ abgekürzten Verteilung (bezüglich des n-dimensionalen Lebesgue-Maßes) ist gegeben durch

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det\Sigma}} \exp\left(-\frac{1}{2}\left\langle x-\mu\,, \Sigma^{-1}(x-\mu)\right\rangle\right), \quad x \in \mathbb{R}^n,$$

wobei $\langle\cdot\,,\cdot\rangle$ das Euklid'sche innere Produkt auf $\mathbb{R}^n$ bezeichnet.

(vii) Eine $\mathbb{R}_{\geq 0}$-wertige ZV X hat die *Exponentialverteilung* mit Parameter $\lambda > 0$, wenn es die Dichte

$$f(x) = \lambda e^{-\lambda x}\mathbf{1}_{[0,\infty[}(x), \quad x \in \mathbb{R},$$

hat.

(viii) Die Verteilung einer $\mathbb{R}_{\geq 0}$-wertigen ZV mit Dichte

$$f(x) = \frac{\theta^r}{\Gamma(r)} x^{r-1} e^{-\theta x}\mathbf{1}_{\mathbb{R}_{\geq 0}}(x)$$

heißt *Gammaverteilung* mit Parametern $\theta > 0$ und $r > 0$. Hierbei verwenden wir die Gammafunktion definiert durch

$$\Gamma(r) := \int_0^\infty x^{r-1} e^{-x} dx, \quad r > 0.$$

(ix) Die *Cauchyverteilung* mit Parameter $a > 0$ ist die Verteilung einer $\mathbb{R}$-wertigen ZV mit Dichte

$$f(x) = \frac{1}{a\pi}\frac{1}{1+(x/a)^2}.$$

(x) Sei $A \in \mathcal{B}(\mathbb{R})$ mit $\lambda(A) > 0$. Eine ZV X mit Dichte

$$f(x) = \frac{1}{\lambda(A)}\mathbf{1}_A(x)$$

heißt *gleichverteilt* auf A. Wir schreiben auch $X \sim \mathcal{U}(A)$.

Bemerkung. Warum überhaupt Zufallsvariablen? An dieser Stelle könnte man sich die Frage stellen, warum man überhaupt Zufallsvariablen studiert in the Wahrscheinlichkeitstheorie, und nicht vielmehr nur W-Maße und Ereignisse. Zufallsvariablen haben die schöne Eigenschaft, dass sich viele mathematische Operation auf der Ebene von Zufallsvariablen sehr einfach ausdrücken lassen, die sich auf Ebene der entsprechenden Verteilungen nur sehr schwer ausdrücken lassen. Wenn beispielsweise X und Y zwei ZV sind, $g : \mathbb{R}^2 \to \mathbb{R}$ eine messbare Funktion, dann ist $g(X, Y)$ eine eindeutig bestimmte ZV — und in speziellen Situation kann man sogar Momente dieser ZV recht einfach bestimmen — aber die Verteilung von $g(X, Y)$ ist im Allgemeinen nur sehr schwer zu bestimmen.

Satz (Satz vom intuitiven Statistiker)**.** *Sei X ein n-dimensionaler Zufallsvektor, $g : \mathbb{R}^n \to \mathbb{R}$ messbar, so dass $g(X)$ eine P-integrierbare Zufallsvariable ist. Dann ist*

$$E\left[g(X)\right] = \int_{\mathbb{R}^n} g(x) P_X(dx).$$

Hat X eine Dichte f (bezüglich λ^n), dann gilt weiterhin

$$E[g(X)] = \int_{\mathbb{R}^n} g(x) f(x) dx.$$

3 Unabhängigkeit

3.1 Bedingte Wahrscheinlichkeiten

Definition 3.1. Sei $(\Omega, \mathcal{F}, P)$ ein W-Raum, $A, B \in \mathcal{F}$ mit $P(B) > 0$. Dann heißt

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

bedingte Wahrscheinlichkeit von A gegeben (unter) B.

Satz 3.2. *Es gelten die Bedingungen von Definition 3.1.*

a) *$Q(A) := P(A|B)$ definiert ein W-Maß auf $\mathcal{F}$.*

b) *(*Formel von der totalen Wahrscheinlichkeit*) Sei $I = \{1, \ldots, N\}$ für ein $N \in \mathbb{N}$ oder $I = \mathbb{N}$, $(B_i)_{i \in I}$ eine Folge disjunkter Ereignisse mit positiver W. Dann gilt für alle $A \in \mathcal{F}$:*

$$P\left(A \cap \left(\bigcup_{i \in I} B_i\right)\right) = \sum_{i \in I} P(A|B_i) P(B_i).$$

c) *(*Bayesformel*) Seien I und $(B_i)_{i \in I}$ wie oben, und weiters $\bigcup_{i \in I} B_i = \Omega$ und $A \in \mathcal{F}$ mit $P(A) > 0$. Dann gilt:*

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

3.2 Unabhängige Ereignisse

Im folgenden ist I eine allgemeine Indexmenge. Wir schreiben $\mathcal{E}(I)$ für die Menge aller *endlichen* Teilmengen von I. Natürlich arbeiten wir weiterhin auf einem W-Raum $(\Omega, \mathcal{F}, P)$.

Definition 3.3. Sei $(A_i)_{i\in I}$ ein System von Ereignissen. $(A_i)_{i\in I}$ heißt *unabhängig*, wenn $\forall J \in \mathcal{E}(I)$ gilt:

$$P\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} P(A_j). \tag{1}$$

Bemerkung 3.4. Unabhängigkeit ist eine Eigenschaft, die vom verwendeten W-Maß P abhängt.

Bemerkung 3.5. Aus *paarweiser* Unabhängigkeit—d.h. die Produktformel (1) gilt für alle $J \subset I$ mit $\#J = 2$—folgt i. A. nicht Unabhängigkeit!

Definition 3.6. Für alle $i \in I$ sei $\mathcal{G}_i \subset \mathcal{F}$. Dann heißt $(\mathcal{G}_i)_{i\in I}$ *unabhängig*, wenn die Produktformel (1) für alle $J \in \mathcal{E}(I)$ und für alle $A_j \in \mathcal{G}_j$, $j \in J$, gilt.

Lemma 3.7. *a)* $(\mathcal{G}_i)_{i\in I}$ *unabhängig genau dann wenn für alle* $J \in \mathcal{E}(I)$ *das System* $\left(\mathcal{G}_j\right)_{j\in J}$ *unabhängig ist.*
b) Sei K *eine beliebige Menge,* $I_k \subset I$ *paarweise disjunkt,* $k \in K$*. Wenn* $(\mathcal{G}_i)_{i\in I}$ *unabhängig ist, dann ist auch das System* $\left(\bigcap_{i\in I_k} \mathcal{G}_i\right)_{k\in K}$ *unabhängig.*
c) Sei $(\mathcal{G}_i)_{i\in I}$ *unabhängig und gelte weiters, dass* $\mathcal{G}_i \cup \{\varnothing\}$ *durchschnittsstabil ist für alle* $i \in I$*. Dann ist das System* $(\sigma(\mathcal{G}_i))_{i\in I}$ *unabhängig.*

Der Beweis verwendet

Lemma 3.8. *Sei* $(\Omega, \mathcal{F})$ *ein Messraum,* $\mathcal{G} \subset \mathcal{F}$ *ein durchschnittsstabiler Erzeuger von* $\mathcal{F}$*,* μ, ν *endliche Maße auf* $\mathcal{F}$ *mit* $\mu|_{\mathcal{G}} = \nu|_{\mathcal{G}}$*. Dann gilt* $\mu = \nu$ *auf ganz* $\mathcal{F}$*.*

Bemerkung 3.9. Sei $(A_i)_{i\in I}$ unabhängig. Dann sind auch die Systeme $(\{A_i\})_{i\in I}$, $\left(\{\varnothing, A_i, A_i^c, \Omega\}\right)_{i\in I}$, $\left(A_i^c\right)_{i\in I}$ unabhängig.

Definition 3.10. Sei $A_n \subset \Omega$, $n \in \mathbb{N}$. Dann definieren wir

$$\limsup_{n\to\infty} A_n := \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n = \{\omega \in \Omega \mid \omega \in A_n \text{ für unendl. viele } n\},$$

$$\liminf_{n\to\infty} A_n := \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n = \{\omega \in \Omega \mid \omega \in A_n \text{ für alle bis auf endl. viele } n\}.$$

Satz 3.11 (Lemma von Borel-Cantelli)**.** *Sei* $A_n \in \mathcal{F}$*,* $n \in \mathbb{N}$*. Dann gilt:*

a) falls $\sum_{n\in\mathbb{N}} P(A_n) < \infty$ *folgt* $P\left(\limsup_{n\to\infty} A_n\right) = 0$*;*

b) falls $(A_n)_{n\in\mathbb{N}}$ *unabhängig und* $\sum_{n\in\mathbb{N}} P(A_n) = \infty$*, dann folgt* $P\left(\limsup_{n\to\infty} A_n\right) = 1$.

3.3 Unabhängige Zufallsvariable

Es seien $(E_i, \mathcal{E}_i)$ Messräume und X_i seien ZV mit Werten in E_i, $i \in I$. (Meist: $(E_i, \mathcal{E}_i) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.)

Definition 3.12. $(X_i)_{i\in I}$ heißt *unabhängig*, falls $(\sigma(X_i))_{i\in I}$ unabhängig ist.

Lemma 3.13. *a) Sei $\mathcal{G}_i$ ein durschnittstabiler Erzeuger von $\mathcal{E}_i$, $i \in I$. Dann ist $(X_i)_{i\in I}$ unabhängig genau dann, wenn $\left(X_i^{-1}(\mathcal{G}_i)\right)_{i\in I}$ unabhängig ist.*
b) Insbesondere, wenn X_i reelle ZV sind, $i \in I$, dann ist $(X_i)_{i\in I}$ unabhängig genau dann, wenn $\forall J \in \mathcal{E}(I)$ und $\forall (x_j)_{j\in J} \in \mathbb{R}^J$ gilt:

$$P\left(\bigcap_{j\in J}\{\, X_j \le x_j \,\}\right) = \prod_{j\in J} P\left(X_j \le x_j\right).$$

Bemerkung 3.14. Sei $X = (X_1, \ldots, X_n)$ ein n-dimensionaler Zufallsvektor mit (gemeinsamer) Verteilungsfunktion $F_X(x) = P\left(X \in \prod_{i=1}^n]-\infty, x_i]\right)$, $x \in \mathbb{R}^n$. Dann gilt: die Komponenten von X sind unabhängig genau dann, wenn

$$\forall x \in \mathbb{R}^n : \; F_X(x) = \prod_{i=1}^n F_{X_i}(x_i),$$

wobei $F_{X_i}(x_i) = P(X_i \le x_i)$ die Verteilungsfunktion der $i$ten Komponente (Randverteilung) bezeichnet.

Bemerkung 3.15. Beachte: für $J \subset \{\, 1, \ldots, n \,\} = I$ erhält man die Verteilungsfunktion F_{X_J} von $X_J := (X_j)_{j\in J}$ aus F_X indem man die Koordinaten aus $I \setminus J$ gegen ∞ gehen lässt. Insbesonders erhält man $F_{X_1}(x_1) = \lim_{x_2\to\infty} \ldots \lim_{x_n\to\infty} F_X(x)$. D.h., die gemeinsame Verteilung von X impliziert die Randverteilung der Koordinaten. Im unabhängigen Fall implizieren auch die Randverteilungen (gegeben durch $F_{X_1}, \ldots, F_{X_n}$) die gemeinsame Verteilung durch die Produktformel.

Bemerkung 3.16. Angenommen, $X = (X_1, \ldots, X_n)$ hat eine Dichte f_X bezüglich des Lebesguemaßes λ^n. Dann haben analog zur obigen Bemerkung die einzelnen Komponenten X_i eine Dichte f_{X_i} bezüglich des eindimensionalen Lebesguemaßes gegeben durch ausintegrieren aller übrigen Koordinaten, etwa für $i = 1$:

$$f_{X_1}(x_1) = \int_{\mathbb{R}^{n-1}} f_X(x_1, z)dz, \quad x_1 \in \mathbb{R}.$$

Aus der Produktformel für Verteilungsfunktionen folgt unmittelbar (mit dem Satz von Fubini), dass die Komponenten von X unabhängig sind genau dann, wenn

$$f_X(x) = \prod_{i=1}^n f_{X_i}(x_i) \text{ für } \lambda^n\text{- fast alle } x \in \mathbb{R}^n.$$

Bemerkung 3.17. Seien $X_1, \ldots, X_n$ diskrete ZV (der Einfachheit halber: mit Werten in $\mathbb{N}$). Diese ZV sind unabhängig genau dann, wenn für alle $J \in \mathcal{E}(\{\, 1, \ldots, n \,\})$ und für alle $x_j \in \mathbb{N}$, $j \in J$, gilt:

$$P\left(\forall j \in J : \; X_j = x_j\right) = P\left(\bigcap_{j\in J}\{\, X_j = x_j \,\}\right) = \prod_{j\in J} P(X_j = x_j).$$

Auch in diesem Fall erhält man die *Randverteilung* einfach durch die gemeinsame Verteilung. Beispielsweise gilt:

$$P(X_1 = x) = \sum_{x_2 \in \mathbb{N}, \dots, x_n \in \mathbb{N}} P\left(X_1 = x, X_2 = x_2, \dots, X_n = x_n\right),$$

und analog für alle anderen Randverteilungen.

Satz 3.18. *Seien $(X_n)_{n \in \mathbb{N}}$ unabhängige ZV mit Werten in $(E_n, \mathcal{E}_n)$.*

a) Für Messräume $(H_n, \mathcal{H}_n)$ und messbare Funktionen $f_n : E_n \to H_n$, $n \in \mathbb{N}$, sind die ZV $(f_n(X_n))_{n \in \mathbb{N}}$ unabhängig.

b) Für paarweise disjunkte Teilmengen $J_k \subset \mathbb{N}$, $k \in K$, sind $\left(\sigma\left((X_j)_{j \in J_k}\right)\right)_{k \in K}$ unabhängig.

c) Sei weiters $f_k : \prod_{j \in J_k} E_j \to H_k$ messbar bezüglich $\bigotimes_{j \in J_k} \mathcal{E}_j$. Dann sind die ZV $\left(f_k\left((X_j)_{j \in J_k}\right)\right)_{k \in K}$ unabhängig.

3.4 Produktmaße

Definition 3.19. Seien $(\Omega_i, \mathcal{F}_i, P_i)$, $i = 1, \dots, n$, W-Räume. Setze

$$\Omega := \prod_{i=1}^n \Omega_i,$$

$$\mathcal{F} := \bigotimes_{i=1}^n \mathcal{F}_i := \sigma\left(\left\{ \prod_{i=1}^n A_i \,\middle|\, A_i \in \mathcal{F}_i, i = 1, \dots, n \right\}\right),$$

$$P\left(\prod_{i=1}^n A_i\right) := \bigotimes_{i=1}^n P_i\left(\prod_{i=1}^n A_i\right) = \prod_{i=1}^n P_i(A_i),$$

fortgesetzt auf $\mathcal{F}$. $\mathcal{F}$ heißt Produktsigmaalgebra, P Produktmaß, $(\Omega, \mathcal{F}, P)$ Produktraum.

Bemerkung 3.20. a) $\mathcal{F} = \sigma\left((\pi_i)_{i=1}^n\right)$ mit $\pi_i : \Omega \to \Omega_i$ definiert durch $\pi_i(\omega) = \omega_i$.
b) $\mathcal{B}(\mathbb{R}^n) = \bigotimes_{i=1}^n \mathcal{B}(\mathbb{R}) =: \mathcal{B}(\mathbb{R})^{\otimes n}$, $\lambda^n = \lambda^{\otimes n}$.

Satz 3.21 (Satz von Fubini)**.** *Sei X eine reelle ZV definiert auf dem Produktraum $(\Omega, \mathcal{F}, P)$.*

(i) $E\left[|X|\right] = \int_\Omega |X(\omega)|\, P(d\omega) < \infty \iff \int_{\Omega_1} \cdots \int_{\Omega_n} |X(\omega_1, \dots, \omega_n)|\, P_n(d\omega_n) \cdots P_1(d\omega_1) < \infty$ *(oder jede Permutation davon).*

(ii) In diesem Fall oder wenn $X \geq 0$ f.s. gilt

$$E[X] = \int_\Omega X(\omega) P(d\omega) = \int_{\Omega_1} \cdots \int_{\Omega_n} X(\omega_1, \dots, \omega_n) P_n(d\omega_n) \cdots P_1(d\omega_1)$$

(oder jede Umordnung der Integrale).

Definition 3.22. Sei $(\Omega, \mathcal{F}, P)$ ein W-Raum, X_i reelle ZV auf Ω, $i = 1, \dots, n$. Dann ist der Vektor $X = (X_1, \dots, X_n)$ ein Zufallsvektor und sein Bildmaß P_X heißt *gemeinsame Verteilung* von $X_1, \dots, X_n$.

Satz 3.23. *Sei X ein reeller Zufallsvektor.*

a) $X_1, \ldots, X_n$ *unabhängig genau dann, wenn* $P_X = \bigotimes_{i=}^n P_{X_i}$.

b) $X_1, \ldots, X_n$ *unabhängig,* $E[|X_i|] < \infty$, $i = 1, \ldots, n$. *Dann gilt*

$$E\left[|X_1 \cdots X_n|\right] < \infty \text{ und } E\left[\prod_{i=1}^n X_i\right] = \prod_{i=1}^n E[X_i].$$

c) Falls für alle beschränkten, messbaren Funktionen $f_i : \mathbb{R} \to \mathbb{R}$ *gilt*

$$E\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n E[f_i(X_i)],$$

dann sind $X_1, \ldots, X_n$ *unabhängig.*

3.5 Unendliche Produktmaße

Definition 3.24. Sei $(\Omega_n, \mathcal{F}_n)$, $n \in \mathbb{N}$, eine Folge von Messräumen. Setze $\Omega := \prod_{n\in\mathbb{N}} \Omega_n$ und definiere die Projektionen $\pi_J : \Omega \to \prod_{j\in J} \Omega_j$, $\omega \mapsto (\omega_j)_{j\in J}$, für $J \in \mathcal{E}(\mathbb{N})$. Ferner definiere für $J \in \mathcal{E}(\mathbb{N})$

$$\mathcal{Z}_J := \pi_J^{-1}\left(\bigotimes_{j\in J} \mathcal{F}_j\right) = \left\{ A \subset \Omega \;\middle|\; \exists A_J \in \bigotimes_{j\in J} \mathcal{F}_j : \ A = \pi_J^{-1}(A_J) \right\}$$

und $\mathcal{Z} := \bigcup_{J\in\mathcal{E}(\mathbb{N})} \mathcal{Z}_J$—Elemente von $\mathcal{Z}$ heißen *Zylindermengen*. Dann definieren wir die unendliche *Produktsigmaalgebra* als

$$\mathcal{F} := \bigotimes_{n\in\mathbb{N}} \mathcal{F}_n := \sigma(\mathcal{Z}),$$

$(\Omega, \mathcal{F}) =: \bigotimes_{n\in\mathbb{N}} (\Omega_n, \mathcal{F}_n)$ heißt unendlicher *Produktraum*.

Bemerkung 3.25. (i) Die obige Konstruktion funktioniert für allgemeine, auch überabzählbare Indexmengen. Für endliche Indexmengen entspricht die Definition der Definition 3.19. Analoges gilt für die Konstruktion des Produktmaßes, siehe unten.

(ii) Zylindermengen haben eine einfache Gestalt: sei zur Einfachheit $J = \{1, \ldots, n\}$ und $A \in \mathcal{Z}_J$. Dann gilt $A = A_J \times \prod_{i=n+1}^\infty \Omega_i$ für ein $A_J \in \bigotimes_{i=1}^n \mathcal{F}_i$.

(iii) $\mathcal{F} = \sigma\left((\pi_J)_{J\in\mathcal{E}(\mathbb{N})}\right) = \sigma\left((\pi_{\{n\}})_{n\in\mathbb{N}}\right)$.

Satz 3.26. *Sei* $(\Omega_n, \mathcal{F}_n, P_n)$, $n \in \mathbb{N}$, *eine Folge von Wahrscheinlichkeitsräumen. Dann existiert ein eindeutiges W-Maß P auf* $\bigotimes_{n\in\mathbb{N}} \mathcal{F}_n$ *mit der Eigenschaft, dass für alle* $J \in \mathcal{E}(\mathbb{N})$ *gilt:*

$$P \circ \pi_J^{-1} = \bigotimes_{j\in J} P_j.$$

$P =: \bigotimes_{n\in\mathbb{N}} P_n$ *heißt unendliches* Produktmaß *der* P_n *und* $(\Omega, \mathcal{F}, P) =: \bigotimes_{n\in\mathbb{N}} (\Omega_n, \mathcal{F}_n, P_n)$ *heißt unendlicher* Produktraum.

Bemerkung. Eine wesentliche Bedingung im obigen Satz ist, dass P_n W-Maße sind, nicht etwa endliche oder allgemeine Maße.

Korollar 3.27. *Gegeben sei eine Folge* $(\mu_n)_{n\in\mathbb{N}}$ *von W-Maßen auf* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. *Dann gibt es einen W-Raum* $(\Omega, \mathcal{F}, P)$ *und eine unabhängige Folge von ZV* $X_n : \Omega \to \mathbb{R}$ *mit* $\forall n : \ P_{X_n} = \mu_n$.

3.6 0-1-Gesetz von Kolmogorow

Definition 3.28. Sei $\mathcal{F}_n$, $n \in \mathbb{N}$, eine Folge von Teilsigmaalgebren von $\mathcal{F}$. Setze

$$\mathcal{T}_n := \sigma\left(\bigcup_{m=n}^{\infty} \mathcal{F}_m\right), \quad \mathcal{T}_\infty := \bigcap_{n=1}^{\infty} \mathcal{T}_n.$$

$\mathcal{T}_\infty$ heißt *terminale σ-Algebra* der Folge $(\mathcal{F}_n)_{n\in\mathbb{N}}$.

Satz 3.29 (0-1-Gesetz von Kolmogorow)**.** *Seien $(\mathcal{F}_n)_{n\in\mathbb{N}}$ unabhängige Teilsigmaalgebren von $\mathcal{F}$. Dann gilt für alle $A \in \mathcal{T}_\infty$: $P(A) \in \{0, 1\}$.*

Korollar 3.30. *a) Sei $(A_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Ereignisse. Dann gilt*

$$P\left(\limsup_{n\to\infty} A_n\right) \in \{0, 1\}, \quad P\left(\liminf_{n\to\infty} A_n\right) \in \{0, 1\}.$$

b) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger erweiter reellwertiger ZV. Dann sind die ZV $\limsup_{n\to\infty} X_n$ und $\liminf_{n\to\infty} X_n$ fast sicher konstant.

3.7 Kovarianz und Korrelation

In diesem Abschnitt betrachten wir ZV definiert auf einem festen W-Raum $(\Omega, \mathcal{F}, P)$.

Definition 3.31. Seien X, Y quadratintegrierbare reelle ZV.

(i) $\mathrm{cov}[X, Y] := E[(X - E[X])(Y - E[Y])]$ heißt *Kovarianz* von X und Y.

(ii) X und Y heißen *unkorreliert* falls $\mathrm{cov}[X, Y] = 0$.

(iii) Falls $\mathrm{var}[X] > 0$ und $\mathrm{var}[Y] > 0$, dann ist der *Korrelationskoeffizient* (kürzer: die Korrelation) von X und Y definiert durch

$$\mathrm{cor}[X, Y] := \frac{\mathrm{cov}[X, Y]}{\sqrt{\mathrm{var}[X]\,\mathrm{var}[Y]}}.$$

Satz 3.32. *Seien $X_1, \dots, X_n, Y_1, \dots, Y_m$ quadratintegrierbare reelle ZV, $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_m \in \mathbb{R}$.*

(i) cov *ist eine symmetrische, positiv semi-definite Bilinearform (am Raum der quadratintegrierbaren ZV). Insbesondere gilt also*

$$\mathrm{cov}\left[\sum_{i=1}^{n} \alpha_i X_i, \sum_{j=1}^{m} \beta_j Y_j\right] = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \,\mathrm{cov}[X_i, Y_j].$$

(ii) Als Spezialfall davon gilt die Gleichung von Bienaymé*:*

$$\mathrm{var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i,j=1}^{n} \mathrm{cov}[X_i, X_j],$$

mit $\mathrm{var}[X] = \mathrm{cov}[X, X]$. *Sind $X_1, \dots, X_n$ paarweise unkorreliert (d.h.: für alle $i \neq j$ sind X_i und X_j unkorreliert), dann gilt*

$$\mathrm{var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{var}[X_i].$$

Satz 3.33. *Seien X, Y quadratintegrierbare ZV.*

a) Es gilt die Cauchy-Schwarz'sche Ungleichung*:* $\mathrm{cov}[X,Y]^2 \le \mathrm{var}[X]\,\mathrm{var}[Y]$.

b) Falls $\mathrm{var}[X] > 0$, $\mathrm{var}[Y] > 0$, *dann ist* $-1 \le \mathrm{cor}[X,Y] \le 1$.

c) Aus Unabhängigkeit von X, Y folgt Unkorreliertheit.

d) Wenn $|\mathrm{cor}[X,Y]| = 1$, *dann gibt es $a, b \in \mathbb{R}$ mit $Y = aX + b$ fast sicher.*

Satz 3.34. *Sei $\mu \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^{n\times n}$ positiv definit und symmetrisch, $X \sim \mathcal{N}(\mu, \Sigma)$ — d.h., die Komponenten von X sind* gemeinsam *normalverteilt.*

(i) $E[X] := (E[X_1], \dots, E[X_n]) = \mu$, $\mathrm{cov}[X] := \left(\mathrm{cov}[X_i, X_j]\right)_{i,j=1}^n = \Sigma$.

(ii) $(X_1, \dots, X_n)$ *unabhängig genau dann, wenn $X_1, \dots, X_n$ paarweise unkorreliert.*

(iii) Sei $a \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$ regulär, $Y := a + AX$. Dann ist $Y \sim \mathcal{N}\left(a + A\mu, A\Sigma A^T\right)$.

Bemerkung 3.35. Die Aussage von Satz 3.34 (iii) gilt auch für $a \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times n}$, solange $A\Sigma A^T$ regulär ist – und noch allgemeiner.

Bemerkung. Aus der paarweisen Unkorreliertheit folgt nicht die (paarweise) Unabhängigkeit. (Beispiel: $X \sim \mathcal{N}(0,1)$ und X^2.)
Wenn $X_1, \dots, X_n$ jeweils für sich (eindimensional) normalverteilt und paarweise unkorreliert sind, dann sind sie nicht notwendigerweise unabhängig.

4 Gesetze der großen Zahlen

In diesem Kapitel sei $(\Omega, \mathcal{F}, P)$ ein W-Raum, $(X_n)_{n\in\mathbb{N}}$ eine darauf definierte Folge von reellen ZV, $S_n := \sum_{i=1}^n$. Falls die ZV X_n integrierbar sind, setzen wir auch $\widetilde{S}_n := \sum_{i=1}^n (X_i - E[X_i])$.

Wir führen die folgende Sprechweise ein: $(X_n)_{n\in\mathbb{N}}$ heißen unabhängig und identich verteilt (kurz: *u.i.v.*), wenn $(X_n)_{n\in\mathbb{N}}$ unabhängig und $\forall n:\ P_{X_n} = P_{X_1}$.

4.1 Summen von Zufallsvariablen

Satz 4.1 (Wald'sche Gleicheit)**.** *Sei $(X_n)_{n\in\mathbb{N}}$ u.i.v. mit $E[|X_1|] < \infty$. Sei weiters T eine ZV mit Werten in $\mathbb{N}$ (oder $\mathbb{N}_0$) mit $E[T] < \infty$, die unabhängig von $(X_n)_{n\in\mathbb{N}}$ ist. Dann ist die ZV S_T, genauer definiert durch*

$$S_T(\omega) = \sum_{i=1}^{T(\omega)} X_i(\omega),$$

P-integrierbar und es gilt

$$E\left[S_T\right] = E[T]E[X_1].$$

Satz 4.2 (Satz von Blackwell–Girshick)**.** *Es gelten die Voraussetzungen von Satz 4.1, aber die ZV T, $(X_n)_{n\in\mathbb{N}}$ seien sogar P-quadratintegrierbar. Dann ist S_T quadratintegrierbar und*

$$\mathrm{var}\left[S_T\right] = E[X_1]^2\,\mathrm{var}[T] + E[T]\,\mathrm{var}[X_1].$$

4.2 Schwaches Gesetz der großen Zahlen

Definition 4.3. Eine Folge $(X_n)_{n\in\mathbb{N}}$ integrierbarer reeller ZV erfüllt das *schwache Gesetz der großen Zahlen* (schw.G.d.g.Z.) genau dann, wenn

$$\forall \epsilon > 0 : \lim_{n\to 0} P\left(\left|\frac{1}{n}\widetilde{S}_n\right| > \epsilon\right) = 0.$$

Sie erfüllt das *starke Gesetz der großen Zahlen* (st.G.d.g.Z.) genau dann, wenn

$$P\left(\limsup_{n\to\infty}\left|\frac{1}{n}\widetilde{S}_n\right| = 0\right) = 1.$$

Bemerkung 4.4. Aus der Maßtheorie kennt man den Begriff der *Konvergenz im Maß*: ist μ ein Maß auf z.B. auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $f, f_n : \mathbb{R}^d \to \mathbb{R}$ messbar, dann konvergiert f_n zu f in μ-Maß, falls

$$\forall \epsilon > 0,\ \forall A \in \mathcal{B}(\mathbb{R}^d) \text{ mit } \mu(A) < \infty : \lim_{n\to\infty} \mu\left(\{\, x \mid |f(x) - f_n(x)| > \epsilon \,\}\right) = 0.$$

Im Falle eines W-Maßes P (auf einem allgemeinen W-Raum $(\Omega, \mathcal{F}, P)$), können wir $A = \Omega$ nehmen und nennen das Konzept auch *Konvergenz in Wahrscheinlichkeit*.

Wenn wir annehmen, dass der Cesaro-Grenzwert $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E[X_i]$ existiert, dann behauptet das schw.G.d.g.Z. also die Konvergenz von $\frac{1}{n}S_n$ zu $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E[X_i]$ in Wahrscheinlichkeit.

Konvergenz in Wahrscheinlichkeit legt den Limes f.s. fest, ist aber schwächer als Konvergenz fast sicher – wie im st.G.d.g.Z.

Satz 4.5. *Seien $(X_n)_{n\in\mathbb{N}}$ paarweise unkorrelierte quadratintegrierbare ZV und gelte weiters*

$$V := \sup_{n\in\mathbb{N}} \operatorname{var}[X_n] < \infty.$$

Dann erfüllt $(X_n)_{n\in\mathbb{N}}$ das schw.G.d.g.Z. und es gilt die Abschätzung

$$P\left(\left|\frac{1}{n}\widetilde{S}_n\right| \geq \epsilon\right) \leq \frac{V}{n\epsilon^2}.$$

Korollar 4.6. *Sei $(X_n)_{n\in\mathbb{N}}$ eine u.i.v. Folge mit $E[X_1^2] < \infty$. Dann gilt das schw.G.d.g.Z. in der Form*

$$\forall \epsilon > 0 : \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - E[X_1]\right| > \epsilon\right) = 0.$$

D.h. $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{n\to\infty} E[X_1]$ in Wahrscheinlichkeit.

4.3 Starkes Gesetz der großen Zahlen

Satz 4.7. *Sei $(X_n)_{n\in\mathbb{N}}$ u.i.v., $E[X_1^2] < \infty$. Dann erfüllt $(X_n)_{n\in\mathbb{N}}$ das st.G.d.G.Z., d.h.*

$$P\left(\lim_{n\to\infty} \frac{1}{n}S_n = E[X_1]\right) = 1.$$

Bemerkung 4.8. Unabhängigkeit wurde im Beweis nur für „$\operatorname{var}[S_n] = n \operatorname{var}[X_1]$“ verwendet, und kann daher durch paarweise Unkorreliertheit ersetzt werden.

Satz 4.9 (Etemadi 1981; ohne Beweis)**.** *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge paarweiser unabhängiger identisch verteilter ZV mit $E[|X_1|] < \infty$. Dann gilt das st.G.d.g.Z.*

Bemerkung 4.10. Zur Konvergenzgeschwindigkeit im st.G.d.g.Z. gibt der *Satz vom iterierten Logarithmus* eine Antwort: für eine u.i.v. Folge $(X_n)_{n\in\mathbb{N}}$ mit $E[X_1^2] < \infty$ gilt

$$\limsup_{n\to\infty} \frac{\left|\widetilde{S}_n\right|}{\sqrt{2\operatorname{var}[X_1]n\log(\log(n))}} = 1 \text{ f.s.}$$

Beispiel 4.11. Sei $(X_n)_{n\in\mathbb{N}}$ eine unabhängige Folge reeller ZV mit $X_1 = 0$ f.s. und

$$P(X_n = -n) = P(X_n = +n) = \frac{1}{2n\log n}, \quad P(X_n = 0) = 1 - \frac{1}{n\log n}.$$

Dann gilt das schw.G.d.g.Z., aber nicht das st.G.d.g.Z. Zum Beweis des schw.G.d.g.Z. in diesem Fall verwendet man die Bedingung von *Markow*: das schw.G.d.g.Z. gilt für Folgen quadratintegrierbarer ZV, wenn

$$\lim_{n\to\infty} \frac{\operatorname{var}\left[\sum_{i=1}^n X_i\right]}{n^2} = 0,$$

wie unmittelbar aus der Ungleichung von Tschebyschow folgt.

4.4 Anwendungen

Beispiel 4.12. Sei X eine reelle ZV, $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Kopien von X, $A \in \mathcal{B}(\mathbb{R})$, gesucht ist $P(X \in A)$. Betrachte die relative Häufigkeit des Eintretens von A basierend auf n Beobachtungen, d.h., die ZV

$$p_n(A) := \frac{1}{n}\sum_{i=1}^n \mathbf{1}_A(X_i).$$

Da die Folge $\mathbf{1}_A(X_i)$ u.i.v. und quadratintegrierbar ist, gilt mit dem st.G.d.g.Z.

$$\lim_{n\to\infty} p_n(A) = P(X \in A) = P_X(A) \text{ f.s.}$$

Definition 4.13. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge reller ZV. Dann heißt die zufällig Funktion $F_n : \mathbb{R} \to [0, 1]$ definiert durch

$$F_n(x) := \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{]-\infty,x]}(X_i), \quad x \in \mathbb{R},\ n \in \mathbb{N},$$

empirirsche Verteilungsfunktion von $X_1, \dots, X_n$.

Satz 4.14 (Satz von Glivenko-Cantelli; Hauptsatz der Statistik)**.** *Sei $(X_n)_{n\in\mathbb{N}}$ eine u.i.v. Folge reeller ZV, F die Verteilungsfunktion von X_1 und F_n die Folge der empirischen Verteilungsfunktionen. Dann gilt*

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0 \text{ f.s.},$$

d.h. $F_n \xrightarrow{n\to\infty} F$ f.s. bezüglich $\|\cdot\|_\infty$.

Satz 4.15 (Weierstraß'scher Approximationssatz)**.** *Sei $f : [0,1] \to \mathbb{R}$ stetig. Dann gibt es eine Folge von Polynomen h_n von Grad n oder kleiner mit*

$$\|f - h_n\|_\infty \xrightarrow{n\to\infty} 0.$$

Beispiel 4.16 (Monte Carlo Simulation)**.** Sei X eine reelle ZV mit $E[|X|] < \infty$, $(X_n)_{n\in\mathbb{N}}$ eine u.i.v. Folge von Kopien von X. Dann wissen wir, dass $\frac{1}{N}\sum_{i=1}^N X_i \to E[X]$ f.s. für $N \to \infty$. Außerdem können wir den Fehler im Sinn des MSE (mean-squared error) berechnen als

$$\mathrm{MSE} = E\left[\left(\frac{1}{N}\sum_{i=1}^N X_i - E[X]\right)^2\right] = \mathrm{var}\left[\frac{1}{N}\sum_{i=1}^N X_i\right] = \frac{1}{N^2}\,\mathrm{var}\left[\sum_{i=1}^N X_i\right] = \frac{\mathrm{var}[X]}{N},$$

falls $E[X^2] < \infty$.

Wenn X eine Dichte f hat, können wir das *Monte Carlo Verfahren* als Methode der numerischen Approximation des Integrals

$$\int_{\mathbb{R}} x f(x) dx$$

verstehen. Praktische Voraussetzung ist, dass wir in der Lage sind, Realisierungen von X_i (effizient) am Computer zu simulieren.

5 Zentraler Grenzwertsatz

Gegeben sei eine u.i.v. Folge reeller (oder $\mathbb{R}^d$-wertiger) ZV definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit $E[X_1^2] < \infty$. Wir wissen vom G.d.g.Z.:

$$\frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{n\to\infty} E[X_1] \text{ f.s.}$$

Es stellt sich die Frage nach der Verteilung des *Fehlers* $\frac{1}{n}S_n - E[X_1]$. Wir werden sehen, dass in einem geeigneten Sinn gilt:

$$\sqrt{n}\left[\frac{1}{n}S_n - E[X_1]\right] \to \mathcal{N}(0, \mathrm{var}[X_1]) \text{ für } n \to \infty.$$

Bemerkung 5.1. Die Normalverteilung $\mathcal{N}$ ist eine logische Vermutung als Grenzverteilung. Unterteilt man die obige skalierte Fehlerfolge in gerad- und ungeradzahlige Beiträge, so ergibt sich, dass die Grenzverteilung μ folgende Eigenschaft haben muss: wenn $U_\infty \sim \mu$ und $G_\infty \sim \mu$ unabhängig sind, dann ist auch $\frac{1}{\sqrt{2}}(U_\infty + G_\infty) \sim \mu$. Es ist einfach zu sehen, dass die Normalverteilung $\mathcal{N}(0,\sigma^2)$ diese Eigenschaft hat, es gilt aber auch die Umkerhung: wenn obige Eigenschaft gilt, dann gibt es ein $\sigma^2 \geq 0$ mit $\mu \sim \mathcal{N}(0,\sigma^2)$.

5.1 Charakteristische Funktionen

Definition 5.2. Sei μ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Die *charakteriscti-sche Funktion* oder *Fouriertransformierte* von μ ist definiert durch

$$\widehat{\mu}(u) := \int_{\mathbb{R}} e^{iux}\mu(dx) := \int_{\mathbb{R}} \cos(ux)\mu(dx) + i\int_{\mathbb{R}} \sin(ux)\mu(dx), \quad u \in \mathbb{R}.$$

Die *charakteristische Funktion* einer reellen ZV X is definiert durch

$$\varphi_X(u) := \widehat{P_X}(u) = E\left[e^{iuX}\right].$$

Ist μ hingegen ein W-Maß auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ bzw. X ein d-dimensionaler Zufallsvektor, dann definieren wir analog

$$\widehat{\mu}(u) := \int_{\mathbb{R}^d} e^{i\langle u, x\rangle} \mu(dx), \quad \varphi_X(u) := E\left[e^{i\langle u, X\rangle}\right], \quad u \in \mathbb{R}^d.$$

Bemerkung 5.3. a) Da $\mathcal{B}(\mathbb{C}) \cong \mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R})^{\otimes 2}$, ist eine $\mathbb{C}$-wertige Funktion f $\mathcal{B}(\mathbb{C})$-messbar genau dann, wenn $\Re f$ und $\Im f$ $\mathcal{B}(\mathbb{R})$-messbar sind.

b) Die obigen Integrale existieren immer, da mit $f_u(x) := e^{iux}$ $(u, x \in \mathbb{R})$ bzw. $f_u(x) := e^{i\langle u, x\rangle}$ $(u, x \in \mathbb{R}^d)$ gilt:

$$|f_u(x)|^2 = 1.$$

Damit gilt jedendfalls

$$\left|\widehat{\mu}(u)\right| \leq \int |f_u(x)| \, \mu(dx) = 1.$$

c) Eng verwandt mit der charakteristischen Funktion ist die *momentenerzeugende Funktion*

$$M_X(u) := E\left[e^{\langle u, X\rangle}\right], \quad u \in \mathbb{R}^d,$$

welche aber im Allgemeinen nur an der Stelle $u = 0$ definiert sein muss. Formal gilt:

$$\varphi_X(u) = M_X(iu), \quad M_X(u) = \varphi_X(-iu).$$

d) Wir konzentrieren uns auf den Fall $d = 1$.

Beispiel 5.4. Für $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$, erhalten wir

$$\varphi_X(u) = \exp\left(i\mu u - \frac{1}{2}\sigma^2 u^2\right), \quad u \in \mathbb{R}.$$

Bemerkung 5.5. Im d-dimensionalen Fall, gilt für $X \sim \mathcal{N}(\mu, \Sigma)$ und $\mu \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d\times d}$ symmetrisch und positiv-definit die analoge Formel

$$\varphi_X(u) = \exp\left(i\langle \mu, u\rangle - \frac{1}{2}\langle u, \Sigma u\rangle\right), \quad u \in \mathbb{R}^d.$$

Man beachte, dass diese Formeln auch für $\sigma^2 = 0$ bzw. Σ lediglich positiv-semidefinit Sinn ergeben und charakteristiche Funktionen von ZV sind. Damit kann man die Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ bzw. $\mathcal{N}(\mu, \Sigma)$ auf diese Fälle ausdehnen.

Satz 5.6. *Ein W-Maß auf* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *ist eindeutig bestimmt durch ihre Fouriertransformierte. Genauer: gegeben W-Maße μ und ν auf* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *mit $\widehat{\mu} = \widehat{\nu}$, dann ist $\mu = \nu$.*

Im Beweis verwenden wir das folgende Resultat:

Satz (Satz von Stone-Weierstrass)**.** *Sei* $C_0(\mathbb{R}, \mathbb{R})$ *der Raum aller stetigen Funktionen, die im Unendlichen verschwinden, d.h.*

$$C_0(\mathbb{R}, \mathbb{R}) := \left\{ f : \mathbb{R} \to \mathbb{R} \;\middle|\; f \text{ stetig}, \lim_{|x|\to\infty} f(x) = 0 \right\}.$$

Betrachten wir eine Teilmenge $C \subset C_0(\mathbb{R}, R)$ mit den Eigenschaften:

(i) C ist eine Algebra, d.h. für $f, g \in C$ und $\alpha \in \mathbb{K}$ sind $f \cdot g, f + g, \alpha f \in C$;

(ii) C ist punktetrennend, d.h., $\forall x \neq y \in E$ gibt es ein $f \in C$ mit $f(x) \neq f(y)$;

(iii) C verschwindet nirgends, d.h. für jedes $x \in \mathbb{R}$ existiert ein $f \in C$ mit $f(x) \neq 0$.

Dann ist C dicht in $C_0(\mathbb{R}, \mathbb{R})$ bezüglich $\|\cdot\|_\infty$.

Bemerkung 5.7. a) Sei $\widehat{\mu} : \mathbb{R} \to \mathbb{C}$ die Fouriertransformierte eines W-Maßes μ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und $a < b \in \mathbb{R}$. Dann gilt die Inversionsformel:

$$\mu\left(]a, b[\right) + \frac{1}{2}\mu\left(\{ a, b \}\right) = \lim_{U\to\infty} \frac{1}{2\pi} \int_{-U}^{U} \frac{e^{-iua} - e^{-iub}}{iu} \widehat{\mu}(u) du.$$

Ist $\widehat{\mu}$ integrierbar bezüglich dem Lebesguemaß, dann hat μ eine Dichte f, wleches für fast alle $x \in \mathbb{R}$ gegeben ist durch

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \widehat{\mu}(u) du.$$

b) Sowohl der Eindeutigkeitssatz als auch die Inversionsformel müssen annehmen, dass $\widehat{\mu}$ die Fouriertransformierte eines W-Maßes ist. Der *Satz von Bochner* liefert eine Charakterisierung aller Fouriertransformierten von W-Maßen: Eine Funktion $f : \mathbb{R} \to \mathbb{C}$ ist Fouriertransformierte eines W-Maßes genau dann, wenn die folgenden Bedingungen erfüllt sind:

(i) f ist stetig und $f(0) = 1$;

(ii) f ist positiv semidefinit, d.h. für alle $n \in \mathbb{N}$ und alle $u_1, \ldots, u_n \in \mathbb{R}, \lambda_1, \ldots, \lambda_n \in \mathbb{C}$ gilt

$$\sum_{i,j=1}^{n} f(u_i - u_j)\lambda_i \overline{\lambda_j} \geq 0.$$

Satz 5.8. *Sei μ ein reelles W-Maß bzw. X eine relle ZV.*

a) $\widehat{\mu}(0) = 1$, $\left\|\widehat{\mu}\right\|_\infty = 1$;

b) $\widehat{\mu}$ ist (gleichmäßig) stetig;

c) zwei reelle ZV X, Y sind unabhängig dann und nur dann, wenn für alle $u_1, u_2 \in \mathbb{R}$ gilt: $\varphi_{(X,Y)}(u_1, u_2) = \varphi_X(u_1) \cdot \varphi_Y(u_2)$;

d) dann gilt auch $\varphi_{X+Y}(u) = \varphi_X(u) \cdot \varphi_Y(u)$.

Satz 5.9. *Sei μ ein W-Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit endlichem n'ten Moment $M_n = \int_{\mathbb{R}} x^n \mu(dx)$. Dann ist die Fouriertransformierte $\widehat{\mu}$ n-fach stetig differenzierbar und*

$$\widehat{\mu}^{(k)}(0) = i^k M_k, \quad k = 0, \ldots, n.$$

5.2 Konvergenz in Verteilung

Definition 5.10. a) Sei $(\mu_n)_{n\in\mathbb{N}}$ eine Folge von W-Maßen auf $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$. μ_n konvergiert *schwach* gegen ein W-Maß μ auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ genau dann, wenn

$$\forall f \in C_b(\mathbb{R}^d, \mathbb{R}) : \int_{\mathbb{R}^d} f(x)\mu_n(dx) \xrightarrow{n\to\infty} \int_{\mathbb{R}^d} f(x)\mu(dx),$$

wobei $C_b(\mathbb{R}^d, \mathbb{R}) := \left\{ f : \mathbb{R}^d \to \mathbb{R} \;\middle|\; f \text{ stetig und beschränkt} \right\}$. Wir schreiben kurz $\mu_n \rightharpoonup \mu$ oder $\mu_n \xrightarrow[w]{n\to\infty} \mu$.

b) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge $\mathbb{R}^d$-wertiger ZV und μ ein W-Maß auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ bzw. X eine weitere $\mathbb{R}^d$-wertige ZV. X_n konvergiert *in Verteilung* bzw. *schwach* gegen μ bzw. X genau dann, wenn $P_{X_n} \rightharpoonup \mu$ bzw. $P_{X_n} \rightharpoonup P_X$.

Bemerkung 5.11. Die ZVen X_n in der Definition müssen nicht alle auf dem selben W-Raum definiert sein.

Satz 5.12. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge reller ZV mit Verteilungsfunktionen F_{X_n} sowie X eine relle ZV mit Verteilungsfunktion F_X. Die folgenden Aussagen sind äquivalent:*

(i) $X_n \rightharpoonup X$;

(ii) für alle $x \in S(F_X) := \{ y \in \mathbb{R} \mid F_X \text{ stetig in } y \}$ gilt: $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$.

Bemerkung 5.13. a) Es sei noch einmal betont, dass die Konvergenz in Verteilung nur von der Verteilung, nicht von der spezifischen Wahl der ZV abhängt.

b) Wenn X_n Dichten f_n und X eine Dichte f (bezüglich λ^d) haben, und $f_n \to f$ λ^d-fast überall, dann folgt $X_n \rightharpoonup X$.

c) Für die Richtung „(ii) $\Rightarrow$ (i)“ im Beweis wurde lediglich verwendet, dass $F_{X_n} \to F_X$ auf einer dichten Teilmenge von $\mathbb{R}$.

Satz 5.14 (Satz von Slutzky)**.** *Seien $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$ Folgen $\mathbb{R}^d$-wertiger ZVen. Gelte $X_n - Y_n \to 0$ in Wahrscheinlichkeit und $Y_n \rightharpoonup Y$. Dann gilt $X_n \rightharpoonup Y$. Insbesondere impliziert die Konvergenz in Wahrscheinlichkeit die Konvergenz in Verteilung.*

Bemerkung 5.15. Man beachte, dass im Allgemeinen die Aussage des Satzes von Slutzky *nicht* gilt, wenn man die Konvergenz in Wahrscheinlichkeit von $X_n - Y_n$ durch Konvergenz in Verteilung ersetzt. Bemerkenswerterweise impliziert i.A. nämlich die schwache Konvergenz zweier Folgen von ZV $(X_n)_{n\in\mathbb{N}}$ und $(Y_n)_{n\in\mathbb{N}}$ nicht die schwache Konvergenz der Folge $(X_n + Y_n)_{n\in\mathbb{N}}$.

Definition 5.16. Eine Folge von W-Maßen $(\mu_n)_{n\in\mathbb{N}}$ auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ heißt *straff*, falls

$$\lim_{K\to\infty} \sup_{n\in\mathbb{N}} \mu_n\left(\left\{ x \in \mathbb{R}^d \;\middle|\; \|x\| > K \right\}\right) = 0.$$

Satz 5.17 (Auswahlsatz von Helly)**.** *Sei $(\mu_n)_{n\in\mathbb{N}}$ eine straffe Folge von W-Maßen auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Dann existiert ein W-Maß μ auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ und eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ mit $\mu_{n_k} \rightharpoonup \mu$.*

Satz 5.18 (Stetigkeitssatz von Lévy)**.** *Gegeben sei eine Folge $(\mu_n)_{n\in\mathbb{N}}$ von W-Maßen auf $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$.*

(i) Falls es ein W-Maß μ gibt mit $\mu_n \rightharpoonup \mu$, dann konvergiert auch $\widehat{\mu_n}$ gegen $\widehat{\mu}$ punktweise.

(ii) Angenommen es gibt eine Funktion $f : \mathbb{R}^d \to \mathbb{C}$ die die folgenden Eigenschaften erfüllt:

- *f ist partiell stetig in 0, d.h. für $i = 1, \dots, d$ ist die Abbildung $\epsilon \mapsto f(\epsilon e_i)$ stetig in 0, wobei e_i den i-ten Einheitsvektor von $\mathbb{R}^d$ bezeichnet;*
- *$\widehat{\mu}_n \to f$ punktweise.*

Dann existiert ein W-Maß μ auf $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$ mit $f = \widehat{\mu}$ und $\mu_n \rightharpoonup \mu$.

Lemma. *Sei μ ein W-Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $K > 0$. Dann gilt*

$$\mu\left([-K, K]^c\right) \leq \frac{K}{2} \int_{-2/K}^{2/K} (1 - \widehat{\mu}(u))du.$$

5.3 Zentraler Grenzwertsatz

Satz 5.19 (Zentraler Grenzwertsatz)**.** *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge u.i.v. $\mathbb{R}^d$-wertiger, quadratintegrierbarer ZV. Sei $\mu = E[X_1] \in \mathbb{R}^d$ und $\Sigma = \operatorname{cov}[X] \in \mathbb{R}^{d\times d}$. Dann gilt*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \rightharpoonup \mathcal{N}(0, \Sigma).$$

Im Falle $d = 1$ bezeichne $\sigma^2 = \operatorname{var}[X_1]$. Dann können wir auch formulieren:

$$\frac{1}{\sqrt{n\sigma^2}} \sum_{i=1}^{n} (X_i - \mu) \rightharpoonup \mathcal{N}(0, 1).$$

Satz 5.20 (Cramér – Wold)**.** *Gegeben eine Folge $(X_n)_{n\in\mathbb{N}}$ $\mathbb{R}^d$-wertiger ZV. Dann konvergiert X_n in Verteilung gegen eine $\mathbb{R}^d$-wertige ZV X genau dann, wenn für jedes $u \in \mathbb{R}^d$ die Folge eindimensionaler ZV $(\langle u, X_n\rangle)_{n\in\mathbb{N}}$ in Verteilung gegen eine reelle ZV X^u konvergiert. In diesem Fall gilt dann weiters $P_{X^u} = P_{\langle u,X\rangle}$.*

Bemerkung 5.21. a) Der *Satz von Berry-Esseen* liefert Schranken für die Konvergenzgeschwindigkeit. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge u.i.v. reeller ZV mit $\gamma := E[|X_1|^3] < \infty$. Bezeichne Φ die Verteilungsfunktion der Standardnormalverteilung $\mathcal{N}(0, 1)$. Dann gilt mit $S_n^* := \frac{1}{\sqrt{n\sigma^2}} \sum_{i=1}^{n} (X_i - \mu)$

$$\sup_{x\in\mathbb{R}} \left| P\left(S_n^* \leq x\right) - \Phi(x) \right| \leq \frac{c\gamma}{\sigma^3 \sqrt{n}}$$

für eine universelle Konstante $0.4097 \leq c \leq 0.4785$.
b) Falls $(X_n)_{n\in\mathbb{N}}$ nicht identisch verteilt ist, dann wird unter dem zentralen Grenzwertsatz die folgende Aussage verstanden:

$$\frac{\sum_{j=1}^{n} (X_j - E[X_j])}{\sqrt{\sum_{j=1}^{n} \operatorname{var}[X_j]}} \rightharpoonup \mathcal{N}(0, 1).$$

Beispiel 5.22. Die Monte Carlo Methode (siehe Beispiel 4.16) besteht in der Approximation von $\mu := E[X]$ durch $\frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^n X_i$, wobei die Folge $(X_i)_{i\in\mathbb{N}}$ eine u.i.v. Folge von ZV mit $P_X = P_{X_i}$, $i \in \mathbb{N}$. Unter der Voraussetzung $\sigma^2 := \mathrm{var}[X] < \infty$ haben wir bereits gesehen:

$$\mathrm{MSE} = E\left[\left(\mu - \frac{1}{n}S_n\right)^2\right] = \frac{\sigma^2}{n}.$$

Allerdings ist der MSE oft kein praktisch relevantes Maß für den zufälligen Fehler der Monte Carlo Simulation. In der Praxis, werden Anforderungen der folgenden Art gestellt: gegeben seien Fehlertoleranzen $\epsilon, \delta > 0$, wähle n derart, dass

$$(\star) \qquad P\left(\left|\mu - \frac{1}{n}S_n\right| > \epsilon\right) \leq \delta.$$

Das schwache Gesetz der großen Zahlen zeigt, dass $(\star)$ erfüllt ist für n genügend groß. Nach dem zentralen Grenzwertsatz gilt

$$\frac{\sqrt{n}}{\sigma}\left(\mu - \frac{1}{n}S_n\right) \rightharpoonup \mathcal{N}(0,1).$$

Damit gilt

$$P\left(\left|\mu - \frac{1}{n}S_n\right| \leq \epsilon\right) = P\left(-\epsilon\frac{\sqrt{n}}{\sigma} \leq \frac{\sqrt{n}}{\sigma}\left(\mu - \frac{1}{n}S_n\right) \leq \epsilon\frac{\sqrt{n}}{\sigma}\right) \approx P\left(-\epsilon\frac{\sqrt{n}}{\sigma} \leq Z \leq \epsilon\frac{\sqrt{n}}{\sigma}\right)$$

für $Z \sim \mathcal{N}(0,1)$. (Wir nehmen also an, dass die Konvergenz schon stattgefunden hat!) Sei Φ die Verteilungsfunktion von $\mathcal{N}(0,1)$. Dann erhalten wir also die Approximation

$$P\left(\left|\mu - \frac{1}{n}S_n\right| \leq \epsilon\right) \approx \Phi\left(\epsilon\frac{\sqrt{n}}{\sigma}\right) - \Phi\left(-\epsilon\frac{\sqrt{n}}{\sigma}\right) = 2\Phi\left(\epsilon\frac{\sqrt{n}}{\sigma}\right) - 1.$$

Insgesamt bekommen wir also

$$P\left(\left|\mu - \frac{1}{n}S_n\right| > \epsilon\right) \approx 2 - 2\Phi\left(\epsilon\frac{\sqrt{n}}{\sigma}\right) \overset{!}{=} \delta.$$

Durch Umformen bekommen wir

$$\Phi^{-1}(1-\delta/2) \approx \epsilon\frac{\sqrt{n}}{\sigma} \iff n \approx \frac{\sigma^2}{\epsilon^2}\Phi^{-1}(1-\delta/2).$$

Bemerkung. a) $(\star)$ kann als (asymptotisches) Konfidenzintervall verstanden werden: $P\left(\frac{1}{n}S_n \in [\mu-\epsilon, \mu+\epsilon]\right) \geq 1-\delta$. Hier wird $\frac{1}{n}S_n$ als (statistischer) Schätzer für den (unbekannten) Parameter μ verstanden.
b) Interpretieren wir $X_i = \mu + \epsilon_i$, und normieren wir die Summe S_n derart, dass $\mathrm{var}[c_n S_n] = \mathcal{O}(1)$, d.h., $c_n \sim 1/\sqrt{n}$. Dann ist $c_n S_n$ approximativ normalverteilt.

5.4 Große Abweichungen

Satz 5.23 (Cramér). *Sei $(X_n)_{n\in\mathbb{N}}$ eine u.i.v. Folge reeller ZV mit endlicher Momentenerzeugendenfunktion $M_{X_1}(u) = E\left[e^{uX_1}\right] \in \mathbb{R}$, $u \in \mathbb{R}$, $S_n := \sum_{i=1}^n X_i$. Definiere die* Ratenfunktion *I als*

$$I(x) := \sup_{u\in\mathbb{R}}\left(ux - \log M_{X_1}(u)\right).$$

Dann gilt für jedes $x > E[X_1]$

$$\lim_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n} S_n \geq x\right) = -I(x).$$

Beispiel 5.24. Falls $X_i \sim \mathcal{N}(0,1)$, dann gilt $M_{X_1}(u) = e^{u^2/2}$, und weiters $I(x) = x^2/2$, und damit für jedes $x > 0$:

$$P\left(\frac{1}{n} S_n \geq x\right) \approx e^{-nx^2/2}, \quad n \to \infty.$$

A Notationen

Wir sammeln einige in diesem Skriptum verwendete Notationen.

- $A_n \nearrow A$: A_n ist eine wachsende Folge von Mengen mit „Grenzmenge“ $A = \bigcup_{n=1}^{\infty} A_n$;
- $A_n \searrow A$: A_n ist eine fallende Folge von Mengen mit „Grenzmenge“ $A = \bigcap_{n=1}^{\infty} A_n$;
- cov bzw. cor stehen für Kovarianz und Korrelationskoeffizienten von zwei ZV;
- f.s. steht für „fast sicher“, also „fast überall“ im Kontext eines W-Maßes;
- λ oder dx … Lebesgue-Maß;
- $MK(\mathcal{F}_0)$ … von $\mathcal{F}_0$ erzeugte monotone Klasse;
- 2^{Ω} … Potenzmenge der Menge Ω;
- $(\Omega, \mathcal{F}, P)$ Wahrscheinlichkeitsraum;
- $\mathbb{R}_{\geq 0} := [0, \infty[$;
- $\overline{R} = [-\infty, +\infty] := \mathbb{R} \cup \{-\infty, +\infty\}$
- $\sigma(\mathcal{F}_0)$ … von $\mathcal{F}_0$ erzeugte σ-Algebra;
- $\sigma(X)$ von einer ZV X erzeugte σ-Algebra;
- $\subset, \supset$ … Mengeninklusionen, die Gleichheit nicht ausschließen;
- $x_n \nearrow x$ … x_n strebt von unten gegen x (also: x_n is monoton steigend mit Grenzwert x);
- $x_n \searrow x$ … x_n strebt von oben gegen x (also: x_n is monoton fallend mit Grenzwert x);
- $X_+ = X \wedge 0$ Positivteil von X;
- $X_- = (-X) \wedge 0$ Negativteil von X;