

Mathematische Stochastik

im

Wintersemester 2013/14

am

Fachbereich Mathematik

der

Universität Hamburg

gelesen von

Holger Drees

Inhaltsverzeichnis

1	Mathematische Modelle von Zufallsexperimenten	1
2	Wahrscheinlichkeitsmaße auf den reellen Zahlen	7
3	Zufallsvariablen und ihre Verteilungen	10
4	Unabhängigkeit und bedingte Wahrscheinlichkeiten	15
5	Erwartungswerte und Varianzen	21
6	Grenzwertsätze	27

1 Mathematische Modelle von Zufallsexperimenten

Grundbegriffe

- **Grundraum:** nichtleere Menge $\Omega \neq \emptyset$
Der Grundraum enthält alle *möglichen* Ergebnisse des betrachteten Zufallsexperiments
- **Ereignis:** Teilmenge $A \subset \Omega$
Nicht immer werden alle Teilmengen von Ω als Ereignisse aufgefasst, sondern nur Mengen aus einem Mengensystem $\mathcal{A} \subset 2^\Omega$. (Dabei bezeichnet 2^Ω die Potenzmenge von Ω , also die Menge aller Untermengen von Ω .)
Sprechweise: Wenn das Ergebnis $\omega \in \Omega$ des Zufallsexperiments in $A \in \mathcal{A}$ liegt, so sagt man, dass A eingetreten sei. Insbesondere gilt stets, dass Ω immer eintritt und \emptyset nie eintritt.
- **Elementarereignis:** Element $\omega \in \Omega$ des Grundraums.
Beachte: Ein Elementarereignis $\omega \in \Omega$ ist *kein* Ereignis, wohl aber (bei geeigneter Wahl von \mathcal{A}) die zugehörige Einpunktmenge $\{\omega\}$.

Logische Zusammenhänge zwischen Ereignissen A, B lassen sich mengentheoretisch darstellen:

$A \cup B$ beschreibt das Ereignis, dass das Ereignis A oder das Ereignis B eintritt.

$A \cap B$: A und B treten ein.

$A \setminus B$: A tritt ein, aber B tritt nicht ein.

B^c : B tritt nicht ein.

$A \Delta B := (A \setminus B) \cup (B \setminus A)$: Entweder A oder B tritt ein.

$A \subset B$: Falls A eintritt, so auch B .

Jedem Ereignis soll jetzt eine Wahrscheinlichkeit zugeordnet werden. Dazu:

1.2 Definition Sei $\Omega \neq \emptyset$ ein Grundraum. Ein Mengensystem $\mathcal{A} \subset 2^\Omega$ heißt **σ -Algebra** auf Ω , falls

$$(i) \quad \Omega \in \mathcal{A}$$

$$(ii) \quad A \in \mathcal{A} \implies A^c \in \mathcal{A}$$

$$(iii) \quad A_n \in \mathcal{A} \quad \forall n \in \mathbb{N} \implies \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$$

Dann heißt (Ω, \mathcal{A}) **Messraum** (oder **messbarer Raum**) und die Mengen $A \in \mathcal{A}$ heißen **messbare Mengen** oder **Ereignisse**. □

1.3 Definition Sei (Ω, \mathcal{A}) ein Messraum. Eine Abbildung $P : \mathcal{A} \rightarrow [0, 1]$ heißt **Wahrscheinlichkeitsmaß**, falls

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

$$(ii) \quad A_n \in \mathcal{A}, n \in \mathbb{N}, \text{ disjunkt} \implies P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n) \quad (\text{“}\sigma\text{-Additivität”})$$

Dann heißt (Ω, \mathcal{A}, P) **Wahrscheinlichkeitsraum**. □

1.5 Satz Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Dann gilt

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

$$(ii) \quad \forall A_1, \dots, A_n \in \mathcal{A}: \quad \bigcup_{i=1}^n A_i \in \mathcal{A}$$

$$\forall A_1, \dots, A_n \in \mathcal{A} \text{ disjunkt: } P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

$$(iii) \quad \forall A \in \mathcal{A}: \quad P(A^c) = 1 - P(A)$$

$$(iv) \quad \forall A, B \in \mathcal{A}, A \subset B: \quad B \setminus A \in \mathcal{A}, \quad P(B \setminus A) = P(B) - P(A) \\ \text{insbesondere gilt } P(A) \leq P(B)$$

$$(v) \quad \forall A, B \in \mathcal{A}: \quad A \cap B \in \mathcal{A}, \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$(vi) \quad \forall A_n \in \mathcal{A}, n \in \mathbb{N}: \quad P\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} P(A_n) \quad (\text{“}\sigma\text{-Subadditivität”}) \quad \square$$

1.6 Satz (Siebformel von Sylvester-Poincaré) Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A_i \in \mathcal{A}$ für alle $1 \leq i \leq n$. Dann gilt

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{\emptyset \neq I \subset \{1, \dots, n\}} (-1)^{|I|-1} P\left(\bigcap_{i \in I} A_i\right)$$

Dabei bezeichnet $|I|$ die Mächtigkeit der Menge I , d.h. die Anzahl der Elemente von I . □

1.8 Satz Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A_n \in \mathcal{A}$ für alle $n \in \mathbb{N}$.

$$(i) \quad \text{Gilt } A_n \subset A_{n+1} \text{ für alle } n \in \mathbb{N}, \text{ so folgt } P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

(Diese Eigenschaft wird σ -Stetigkeit von unten genannt.)

$$(ii) \quad \text{Gilt } A_{n+1} \subset A_n \text{ für alle } n \in \mathbb{N}, \text{ so folgt } P\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

(Diese Eigenschaft wird σ -Stetigkeit von oben genannt. Man beachte, dass auch ohne die Monotonievoraussetzung stets $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{A}$ gilt.) □

1.9 Definition Ein Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) heißt **diskret**, falls $\{\omega\} \in \mathcal{A}$ für alle $\omega \in \Omega$ und eine abzählbare Menge Ω_0 existiert mit $P(\Omega_0) = 1$. Dann heißt auch P diskret und die Abbildung

$$f : \Omega \rightarrow [0, 1], \quad f(\omega) := P\{\omega\}$$

heißt **Zähldichte** (oder **Wahrscheinlichkeitsfunktion**).

(Man beachte, dass die Voraussetzung $\{\omega\} \in \mathcal{A}$ für alle $\omega \in \Omega$ sicher stellt, dass jede abzählbare Untermenge von Ω in \mathcal{A} liegt. Hier und im folgenden wird “abzählbar” immer im Sinne von “höchstens abzählbar”, also “endlich oder abzählbar” gebraucht.) \square

1.10 Bemerkung Mengen $A \in \mathcal{A}$ mit $P(A) = 1$ heißen **Einsmengen**; man sagt, dass A **P -fast sicher** (P -f.s.) **eintritt**.

Mengen $A \in \mathcal{A}$ mit $P(A) = 0$ heißen **Nullmengen**; man sagt, dass A **P -fast sicher nicht eintritt**.

Diskrete Wahrscheinlichkeitsräume sind also solche Wahrscheinlichkeitsräume, bei denen alle Einpunktmengen messbar sind und die eine abzählbare Einsmenge besitzen. \square

1.11 Satz Sei (Ω, \mathcal{A}) ein Messraum mit $\{\omega\} \in \mathcal{A}$ für alle $\omega \in \Omega$.

(i) Ist P ein diskretes Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}) mit Zähldichte f , so gilt

$$P(A) = \sum_{\omega \in A} f(\omega) \quad \forall A \in \mathcal{A}$$

und f besitzt die Eigenschaften

$$(a) \quad f(\omega) \geq 0 \quad \forall \omega \in \Omega$$

$$(b) \quad f(\omega) = 0 \text{ für alle bis auf abzählbar viele } \omega \in \Omega$$

$$(c) \quad \sum_{\omega \in \Omega} f(\omega) = 1$$

(ii) Ist $f : \Omega \rightarrow \mathbb{R}$ eine Abbildung mit den Eigenschaften (a)–(c), so existiert genau ein Wahrscheinlichkeitsmaß P auf (Ω, \mathcal{A}) mit $P\{\omega\} = f(\omega)$ für alle $\omega \in \Omega$. P ist dann diskret mit Zähldichte f . \square

1.12 Beispiel Ist $\Omega \neq \emptyset$ eine endliche Menge, so heißt das Wahrscheinlichkeitsmaß auf $(\Omega, 2^\Omega)$ mit Zähldichte

$$f(\omega) = \frac{1}{|\Omega|} \quad \forall \omega \in \Omega$$

die **Gleichverteilung** oder **Laplace-Verteilung** auf Ω .

Es gilt dann $P(A) = |A|/|\Omega|$ für alle $A \subset \Omega$. \square

Nachfolgend werden geeignete Grundräume zur Beschreibung von k Ziehungen aus einer Urne, die mit $1, \dots, n$ durchnummerierte Kugeln enthält, sowie deren Mächtigkeit angegeben. Dabei können wahlweise gezogene Kugeln wieder zurückgelegt werden oder nicht, und die Zugreihenfolge mitberücksichtigt oder ignoriert werden.

1.13 Ziehen mit Zurücklegen und mit Berücksichtigung der Reihenfolge

$$\Omega_I = \{1, \dots, n\}^k$$

Interpretation: $\omega = (\omega_1, \dots, \omega_k) \in \Omega$ bedeutet, dass für alle $1 \leq i \leq k$ im i -ten Zug die Kugel mit der Nummer ω_i gezogen worden ist.

Es gilt: $|\Omega_I| = n^k$ □

1.14 Ziehen ohne Zurücklegen und mit Berücksichtigung der Reihenfolge

$$\Omega_{II} = \{(\omega_1, \dots, \omega_k) \in \{1, \dots, n\}^k \mid \omega_i \neq \omega_j \forall 1 \leq i < j \leq k\}$$

Interpretation: $\omega = (\omega_1, \dots, \omega_k) \in \Omega$ bedeutet, dass für alle $1 \leq i \leq k$ im i -ten Zug die Kugel mit der Nummer ω_i gezogen worden ist.

Es gilt: $|\Omega_{II}| = \frac{n!}{(n-k)!}$ □

1.15 Ziehen ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge

$$\Omega_{III} = \{A \subset \{1, \dots, n\} \mid |A| = k\}$$

Interpretation: $A \in \Omega$ enthält genau die k gezogenen Kugelnummern.

Es gilt: $|\Omega_{III}| = \frac{n!}{k!(n-k)!} = \binom{n}{k}$

Alternativ kann man z.B. auch den folgenden Grundraum wählen, der den Grundräumen in den anderen betrachteten Fällen ähnlicher ist:

$$\tilde{\Omega}_{III} = \{(\omega_1, \dots, \omega_k) \in \{1, \dots, n\}^k \mid \omega_i < \omega_j \forall 1 \leq i < j \leq k\}$$

Interpretation: $\omega = (\omega_1, \dots, \omega_k) \in \Omega$ bedeutet, dass die der Größe nach geordneten gezogenen Kugelnummern gerade $\omega_1, \dots, \omega_k$ sind.

Es gilt natürlich: $|\tilde{\Omega}_{III}| = |\Omega_{III}| = \binom{n}{k}$ □

1.16 Ziehen mit Zurücklegen und ohne Berücksichtigung der Reihenfolge

$$\Omega_{IV} = \{(\omega_1, \dots, \omega_k) \in \{1, \dots, n\}^k \mid \omega_i \leq \omega_j \forall 1 \leq i < j \leq k\}$$

Interpretation: $\omega = (\omega_1, \dots, \omega_k) \in \Omega$ bedeutet, dass die der Größe nach geordneten gezogenen Kugelnummern gerade $\omega_1, \dots, \omega_k$ sind.

Es gilt: $|\Omega_{IV}| = \binom{n+k-1}{k}$ □

Nachfolgend sollen einige Wahrscheinlichkeitsmaße definiert werden, die ausgehend von den obigen Urnenmodellen motiviert werden können.

1.19 Binomialverteilung

Sei $p \in [0, 1]$. Das durch

$$\mathcal{B}_{(n,p)}\{k\} := \binom{n}{k} p^k (1-p)^{n-k} \quad \forall 0 \leq k \leq n$$

(gemäß Satz 1.11) eindeutig festgelegte Wahrscheinlichkeitsmaß $\mathcal{B}_{(n,p)}$ auf dem Messraum $(\{0, \dots, n\}, 2^{\{0, \dots, n\}})$ heißt **Binomialverteilung** zu den Parametern n und p .

$\mathcal{B}_{(n,p)}\{k\}$ gibt die Wahrscheinlichkeit an, in n unabhängigen Zufallsexperimenten, bei denen jeweils die “Erfolgswahrscheinlichkeit” p beträgt, insgesamt genau k “Erfolge” zu erzielen. \square

1.20 Multinomialverteilung

Seien $m \geq 2$ und $p_1, \dots, p_m \in [0, 1]$ so, dass $\sum_{j=1}^m p_j = 1$. Das auf dem Grundraum

$$\Omega := \left\{ (k_1, \dots, k_m) \in \{0, \dots, n\}^m \mid \sum_{j=1}^m k_j = n \right\}$$

versehen mit der Potenz- σ -Algebra $\mathcal{A} = 2^\Omega$ durch

$$\mathcal{M}_{(n;p_1, \dots, p_m)}\{(k_1, \dots, k_m)\} := \binom{n}{k_1, \dots, k_m} p_1^{k_1} \cdots p_m^{k_m} \quad \forall (k_1, \dots, k_m) \in \Omega$$

(gemäß Satz 1.11) eindeutig festgelegte Wahrscheinlichkeitsmaß $\mathcal{M}_{(n;p_1, \dots, p_m)}$ heißt **Multinomialverteilung**.

Hierbei ist der Multinomialkoeffizient definiert als

$$\binom{n}{k_1, \dots, k_m} := \frac{n!}{k_1! \cdots k_m!}.$$

$\mathcal{M}_{(n;p_1, \dots, p_m)}$ gibt die Wahrscheinlichkeit an, in n unabhängigen Zufallsexperimenten mit m möglichen Ausgängen, bei denen jeweils die Wahrscheinlichkeit für das j -te Ergebnis p_j beträgt, insgesamt genau k_j mal das j -te Ergebnis zu erzielen für alle $1 \leq j \leq m$. \square

1.21 Hypergeometrische Verteilung

Seien $N, K, n \in \mathbb{N}$ gegeben, so dass $\max(n, K) \leq N$. Das auf dem Grundraum

$$G_{(N,K,n)} := \{ \max(0, n + K - N), \dots, \min(n, K) \}$$

versehen mit der Potenz- σ -Algebra $\mathcal{A} = 2^{G_{(N,K,n)}}$ durch

$$\mathcal{H}_{(N,K,n)}\{k\} := \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} =: h_{(N,K,n)}(k) \quad \forall k \in G_{(N,K,n)}$$

(gemäß Satz 1.11) eindeutig festgelegte Wahrscheinlichkeitsmaß $\mathcal{H}_{(N,K,n)}$ heißt **hypergeometrische Verteilung**.

$\mathcal{H}_{(N,K,n)}\{k\}$ gibt die Wahrscheinlichkeit an, in n Ziehungen ohne Zurücklegen aus einer Gesamtheit von N Gegenständen, von denen K markiert sind, genau k markierte Gegenstände zu ziehen, wenn alle Gegenstände jeweils die gleiche Ziehungswahrscheinlichkeit besitzen. \square

Die oben genannten Wahrscheinlichkeitsmaße werden mitunter auch auf größeren abzählbaren Grundräumen definiert (z.B. die hypergeometrische Verteilung auf $\{0, \dots, n\}$ statt auf $G_{(N,K,n)}$). Die Zähldichte wird dann für die neu hinzugekommenen Elementarereignisse jeweils als 0 definiert.

1.22 Definition *Das durch*

$$\mathcal{P}_\lambda\{k\} = e^{-\lambda} \frac{\lambda^k}{k!} \quad \forall k \in \mathbb{N}_0$$

*eindeutig festgelegte Wahrscheinlichkeitsmaß \mathcal{P}_λ auf $(\mathbb{N}_0, 2^{\mathbb{N}_0})$ heißt **Poisson-Verteilung** zum Parameter $\lambda > 0$. □*

1.23 Satz (Poisson'scher Grenzwertsatz) *Ist $p_n \in [0, 1]$, $n \in \mathbb{N}$, eine Folge, so dass $\lim_{n \rightarrow \infty} np_n = \lambda > 0$, so gilt für alle $k \in \mathbb{N}_0$*

$$\lim_{n \rightarrow \infty} \mathcal{B}_{(n, p_n)}\{k\} = \mathcal{P}_\lambda\{k\}.$$

□

2 Wahrscheinlichkeitsmaße auf den reellen Zahlen

Viele auf natürliche Weise auftretende Zufallsexperimente, die ein reellwertiges Ergebnis liefern, können nicht durch Wahrscheinlichkeitsmaße auf dem Messraum $(\mathbb{R}, 2^{\mathbb{R}})$ modelliert werden, da die Potenzmenge als σ -Algebra zu komplex ist. Insbesondere kann man keine Gleichverteilung auf einem Intervall versehen mit der Potenzmenge als σ -Algebra definieren. Man sucht daher eine kleinere σ -Algebra auf \mathbb{R} , die dennoch alle “interessanten Mengen”, z.B. alle Intervalle enthält.

2.2 Definition Sei $\Omega \neq \emptyset$ ein Grundraum und $\mathcal{C} \subset 2^{\Omega}$ ein Mengensystem von Untermengen von Ω . Die **von \mathcal{C} erzeugte σ -Algebra** (i.Z. $\sigma(\mathcal{C})$) ist die kleinste σ -Algebra \mathcal{A} , die \mathcal{C} umfasst, d.h.

(i) \mathcal{A} ist eine σ -Algebra auf Ω

(ii) $\mathcal{C} \subset \mathcal{A}$

(iii) Ist $\tilde{\mathcal{A}}$ eine σ -Algebra auf Ω mit $\mathcal{C} \subset \tilde{\mathcal{A}}$, so gilt $\mathcal{A} \subset \tilde{\mathcal{A}}$.

\mathcal{C} heißt dann **Erzeugendensystem** oder kurz **Erzeuger** von \mathcal{A} . □

2.3 Lemma In der Situation von Definition 2.2 gilt

$$\sigma(\mathcal{C}) = \bigcap_{\substack{\mathcal{A} \subset 2^{\Omega} \text{ } \sigma\text{-Algebra, } \mathcal{C} \subset \mathcal{A}}} \mathcal{A}.$$

□

2.4 Definition Sei $I \subset \mathbb{R}$ ein (nicht notwendigerweise beschränktes) Intervall. Die von den links halboffenen Intervallen in I erzeugte σ -Algebra $\sigma(\{(a, b] \mid a, b \in I, a < b\})$ heißt **Borel- σ -Algebra** auf I ; i.Z. $\mathcal{B}(I)$.

Speziell für $I = \mathbb{R}$ schreiben wir $\mathbb{B} = \mathcal{B}(\mathbb{R})$. □

2.5 Bemerkung (i)

$$\begin{aligned} \mathbb{B} &= \sigma(\{(a, b) \mid a, b \in \mathbb{R}, a < b\}) \\ &= \sigma(\{[a, b) \mid a, b \in \mathbb{R}, a < b\}) \\ &= \sigma(\{[a, b] \mid a, b \in \mathbb{R}, a < b\}) \\ &= \sigma(\{(-\infty, x] \mid x \in \mathbb{R}\}) \end{aligned}$$

(ii) $\mathbb{B} = \sigma(\{O \subset \mathbb{R} \mid O \text{ offen}\})$

(iii) $\mathbb{B} = \sigma(\{A \subset \mathbb{R} \mid A \text{ abgeschlossen}\})$

(iv) $\mathbb{B} = \sigma(\{K \subset \mathbb{R} \mid K \text{ kompakt}\})$

Es gelten auch die analogen Aussagen über $\mathcal{B}(I)$ für Intervalle $I \subset \mathbb{R}$. □

2.6 Satz (Maßfortsetzungssatz spezialisiert auf Wahrscheinlichkeitsmaße) Sei (Ω, \mathcal{A}) ein Messraum und $\mathcal{A} = \sigma(\mathcal{C})$.

(i) Gilt

- $\Omega \in \mathcal{C}$
- $A, B \in \mathcal{C} \implies A \setminus B \in \mathcal{C}, A \cup B \in \mathcal{C}$

und ist $\tilde{P} : \mathcal{C} \rightarrow [0, 1]$ eine σ -additive Abbildung mit $\tilde{P}(\Omega) = 1$, d.h. für alle disjunkten Mengen $A_n \in \mathcal{C}$, $n \in \mathbb{N}$, für die $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{C}$, gilt die Gleichung $\tilde{P}(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \tilde{P}(A_n)$, so existiert ein Wahrscheinlichkeitsmaß P auf (Ω, \mathcal{A}) mit

$$P(C) = \tilde{P}(C) \quad \forall C \in \mathcal{C}.$$

(Es existiert also mindestens eine Fortsetzung P von \tilde{P} auf die von \mathcal{C} erzeugte σ -Algebra \mathcal{A} .)

(ii) Sind P_1, P_2 Wahrscheinlichkeitsmaße auf (Ω, \mathcal{A}) so, dass

$$P_1(C) = P_2(C) \quad \forall C \in \mathcal{C},$$

ist \mathcal{C} \cap -stabil (d.h. für alle $A, B \in \mathcal{C}$ liegt auch $A \cap B$ in \mathcal{C}) und existieren $C_n \in \mathcal{C}$, $n \in \mathbb{N}$, mit $\bigcup_{n \in \mathbb{N}} C_n = \Omega$, so gilt

$$P_1(A) = P_2(A) \quad \forall A \in \mathcal{A}.$$

(Es kann also maximal eine Fortsetzung einer σ -additiven Abbildung auf einem \cap -stabilen Erzeugendensystem \mathcal{C} zu einem Wahrscheinlichkeitsmaß P auf (Ω, \mathcal{A}) geben.) \square

2.7 Korollar und Definition (i) Ist P ein Wahrscheinlichkeitsmaß auf (\mathbb{R}, \mathbb{B}) , so gilt für die Funktion

$$F : \mathbb{R} \rightarrow [0, 1], \quad x \mapsto P(-\infty, x]$$

(a) F ist monoton steigend

(b) F ist rechtsseitig stetig

(c) $F(-\infty) := \lim_{x \downarrow -\infty} F(x) = 0$, $F(\infty) := \lim_{x \uparrow \infty} F(x) = 1$

F heißt **Verteilungsfunktion** (Vf.) von P .

(ii) Besitzt $F : \mathbb{R} \rightarrow [0, 1]$ die Eigenschaften (a)–(c), so existiert genau ein Wahrscheinlichkeitsmaß P auf (\mathbb{R}, \mathbb{B}) mit Vf. F . \square

2.8 Definition Eine Funktion $f : \mathbb{R} \rightarrow [0, \infty)$, die Riemann-integrierbar ist mit $\int_{-\infty}^{\infty} f(t) dt = 1$, heißt **(Riemann-)Dichte**. \square

2.9 Satz und Definition Ist f eine Riemann-Dichte, so definiert

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

eine Verteilungsfunktion. Die Funktion f heißt auch **Riemann-Dichte von F** bzw. des gemäß Korollar 2.7 zugehörigen Wahrscheinlichkeitsmaßes P . \square

2.10 Bemerkung (ii) Eine Dichte eines Wahrscheinlichkeitsmaßes auf (\mathbb{R}, \mathbb{B}) ist (im Falle der Existenz) nicht eindeutig bestimmt. Beispielsweise kann die Dichte auf einer endlichen Menge gleich beliebigen nicht-negativen Werten gesetzt werden.

(iii) Ist $I \subset \mathbb{R}$ ein Intervall und P ein Wahrscheinlichkeitsmaß auf $(I, \mathcal{B}(I))$, so kann dieses “ergänzt” werden zu einem Wahrscheinlichkeitsmaß \tilde{P} auf (\mathbb{R}, \mathbb{B}) durch $\tilde{P}(B) := P(B \cap I)$ für alle $B \in \mathbb{B}$. Die Vf. (bzw. ggf. eine Dichte) von \tilde{P} wird dann auch als Vf. (bzw. ggf. Dichte) von P bezeichnet. Im Falle der Existenz kann die Dichte so gewählt werden, dass sie überall außerhalb von I den Wert 0 annimmt. \square

2.13 Satz Ist F eine stetige Vf., die überall bis auf endlich viele Ausnahmestellen $x_i, 1 \leq i \leq n$, stetig differenzierbar ist, so ist $f = F' \cdot 1_{\mathbb{R} \setminus \{x_1, \dots, x_n\}}$ eine Dichte von F . \square

3 Zufallsvariablen und ihre Verteilungen

3.2 Definition Seien (Ω, \mathcal{A}) und $(\tilde{\Omega}, \tilde{\mathcal{A}})$ Messräume. Eine Abbildung $X : \Omega \rightarrow \tilde{\Omega}$ heißt $\mathcal{A}, \tilde{\mathcal{A}}$ -**messbar** (oder kurz: **messbar**), falls $X^{-1}(\tilde{A}) = \{\omega \in \Omega \mid X(\omega) \in \tilde{A}\} \in \mathcal{A}$ für alle $\tilde{A} \in \tilde{\mathcal{A}}$. In diesem Fall heißt X auch eine $\tilde{\Omega}$ -wertige **Zufallsvariable** (Zv.) auf Ω und man schreibt

$$X : (\Omega, \mathcal{A}) \rightarrow (\tilde{\Omega}, \tilde{\mathcal{A}}).$$

$X(\omega)$ wird eine **Realisation von X** genannt ($\omega \in \Omega$). □

Beachte: $X^{-1}(\tilde{A})$ bezeichnet hier das **Urbild** der Menge \tilde{A} unter der Abbildung X und **nicht** das Bild der Menge unter einer Umkehrabbildung (die im Regelfall auch gar nicht existiert)!

3.3 Satz und Definition Ist in der Situation von Definition 3.2 P ein Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}) , so definiert

$$\tilde{P}(\tilde{A}) := P(X^{-1}(\tilde{A})) = P\{\omega \in \Omega \mid X(\omega) \in \tilde{A}\}, \quad \tilde{A} \in \tilde{\mathcal{A}},$$

ein Wahrscheinlichkeitsmaß auf $(\tilde{\Omega}, \tilde{\mathcal{A}})$, die **Verteilung von X unter P** (oder das **Bildmaß von P unter X**); i.Z. P^X (oder $\mathcal{L}(X|P)$). □

3.4 (Notationen) Statt $X^{-1}(\tilde{A}) = \{\omega \in \Omega \mid X(\omega) \in \tilde{A}\}$ schreibt man auch kurz $\{X \in \tilde{A}\}$.

Ebenso: $\{X \text{ erfüllt die Eigenschaft } E\} := \{\omega \in \Omega \mid X(\omega) \text{ erfüllt die Eigenschaft } E\}$, also z.B. für $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathbb{B})$, $t \in \mathbb{R}$

$$\begin{aligned} \{X \leq t\} &:= \{\omega \in \Omega \mid X(\omega) \leq t\} = X^{-1}((-\infty, t]) \\ \{X = t\} &:= \{\omega \in \Omega \mid X(\omega) = t\} = X^{-1}(\{t\}) \end{aligned}$$

usw. □

3.5 Satz Ist $\tilde{\mathcal{C}}$ ein Erzeugendensystem von $\tilde{\mathcal{A}}$, so ist $X : \Omega \rightarrow \tilde{\Omega}$ genau dann $\mathcal{A}, \tilde{\mathcal{A}}$ -messbar, wenn $X^{-1}(\tilde{C}) \in \mathcal{A}$ für alle $\tilde{C} \in \tilde{\mathcal{C}}$. □

3.6 Lemma Ist (Ω, \mathcal{A}) ein Messraum und $X : \Omega \rightarrow \tilde{\Omega}$, so ist

$$\mathcal{M} := \{\tilde{A} \subset \tilde{\Omega} \mid X^{-1}(\tilde{A}) \in \mathcal{A}\}$$

die größte σ -Algebra $\tilde{\mathcal{A}}$ auf $\tilde{\Omega}$, für die X $\mathcal{A}, \tilde{\mathcal{A}}$ -messbar ist. □

3.7 Korollar Sei (Ω, \mathcal{A}) ein Messraum und $X : \Omega \rightarrow \mathbb{R}$.

(i) X ist \mathcal{A}, \mathbb{B} -messbar $\iff \{X \leq t\} \in \mathcal{A}$ für alle $t \in \mathbb{R}$
Insbesondere sind monotone Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ \mathbb{B}, \mathbb{B} -messbar.

(ii) X ist \mathcal{A}, \mathbb{B} -messbar $\iff \{X \in O\} \in \mathcal{A}$ für alle offene Mengen $O \subset \mathbb{R}$
Insbesondere sind stetige Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ \mathbb{B}, \mathbb{B} -messbar. □

3.8 Definition Sei $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathbb{B})$ eine \mathbb{R} -wertige Zufallsvariable. Die durch

$$F_X(t) := P^X(-\infty, t] = P\{X \leq t\}, \quad t \in \mathbb{R},$$

definierte Verteilungsfunktion F_X der Verteilung P^X von X heißt auch **Verteilungsfunktion von X** . Besitzt F_X eine Dichte, so wird diese auch **Dichte von X** genannt. \square

3.10 Bemerkung Verteilungen und Wahrscheinlichkeitsmaße sind zueinander äquivalente Konzepte in dem Sinne, dass

- jede Verteilung ein Wahrscheinlichkeitsmaß ist, und umgekehrt
- jedes Wahrscheinlichkeitsmaß als Verteilung einer geeigneten Zufallsvariable darstellbar ist. (Ist (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, so ist $P = P^{Id}$, wobei $Id : \Omega \rightarrow \Omega$, $Id(\omega) = \omega$, die Identitätsabbildung auf Ω bezeichne.)

Man kann Zufallsexperimente daher durch Wahrscheinlichkeitsräume oder durch Verteilungen von Zufallsvariablen beschreiben, wobei man sich z.B. für die Darstellung entscheiden kann, die die Intuition besser unterstützt.

In der Regel interessiert bei der Darstellung mit Hilfe von Zufallsvariablen jedoch nicht die genaue Form der Zufallsvariablen *als Abbildung*, sondern nur deren Verteilung. Man lässt daher den Messraum, auf dem die Zufallsvariable definiert ist, oft unbestimmt. Übliche Sprechweisen sind z.B. “Sei X eine $\mathcal{B}_{(n,p)}$ -verteilte Zufallsvariable” oder “Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $X : (\Omega, \mathcal{A}) \rightarrow (\{0, \dots, n\}, 2^{\{0, \dots, n\}})$ mit Verteilung $P^X = \mathcal{B}_{(n,p)}$ ” (ohne genaue Angabe von (Ω, \mathcal{A}, P) oder von X als Abbildung). \square

3.11 Satz Sei $(\Omega_1, \mathcal{A}_1, P_1)$ ein Wahrscheinlichkeitsraum, $X : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$ eine Zufallsvariable und $f : (\Omega_2, \mathcal{A}_2) \rightarrow (\Omega_3, \mathcal{A}_3)$ eine weitere messbare Abbildung. Dann ist die Hintereinanderausführung $f \circ X : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_3, \mathcal{A}_3)$ eine Zufallsvariable und es gilt $P^{f \circ X} = (P^X)^f$. \square

Oft möchte man den aus Zufallsvariablen $X_i : (\Omega, \mathcal{A}) \rightarrow (\Omega_i, \mathcal{A}_i)$, $1 \leq i \leq n$, gebildeten Vektor $X = (X_1, \dots, X_n) : \Omega \rightarrow \times_{i=1}^n \Omega_i$ betrachten. Welche σ -Algebra sollte in dem Fall sinnvollerweise auf dem Bildraum $\times_{i=1}^n \Omega_i$ gewählt werden?

3.12 Definition Sind $(\Omega_i, \mathcal{A}_i)$, $1 \leq i \leq n$, Messräume, so wird die **Produkt- σ -Algebra** $\bigotimes_{i=1}^n \mathcal{A}_i$ auf $\times_{i=1}^n \Omega_i$ definiert als die von $\{\times_{i=1}^n A_i \mid A_i \in \mathcal{A}_i \forall 1 \leq i \leq n\}$ erzeugte σ -Algebra. Der Messraum $(\times_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{A}_i)$ heißt **Produktmessraum**.

Speziell im Fall identischer Messräume $(\Omega_i, \mathcal{A}_i) = (\Omega_1, \mathcal{A}_1)$ für alle $1 \leq i \leq n$ schreibt man kurz $(\Omega_1^n, \mathcal{A}_1^n) = (\times_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{A}_i)$. \square

3.13 Satz Seien (Ω, \mathcal{A}) , $(\Omega_i, \mathcal{A}_i)$ Messräume und $X_i : \Omega \rightarrow \Omega_i$ Abbildungen ($1 \leq i \leq n$). Dann sind äquivalent:

- (i) $X = (X_1, \dots, X_n) : \Omega \rightarrow \times_{i=1}^n \Omega_i$ ist $\mathcal{A}, \bigotimes_{i=1}^n \mathcal{A}_i$ -messbar.
- (ii) X_i ist $\mathcal{A}, \mathcal{A}_i$ -messbar für alle $1 \leq i \leq n$. \square

3.14 Bemerkung Bezeichnet $pr_j : \times_{i=1}^n \Omega_i \rightarrow \Omega_j$, $(\omega_1, \dots, \omega_n) \mapsto \omega_j$ die Projektion auf die j -te Komponente, so ist $\bigotimes_{i=1}^n \mathcal{A}_i$ gerade die kleinste σ -Algebra auf $\times_{i=1}^n \Omega_i$, die alle Projektionen pr_j , $1 \leq j \leq n$, messbar macht. \square

3.16 Bemerkung und Definition Ist Q ein Wahrscheinlichkeitsmaß auf dem Produktmessraum $(\times_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{A}_i)$ und bezeichnet pr_j wiederum die Projektion auf die j -te Komponente, so heißt das Wahrscheinlichkeitsmaß $Q_j := Q^{pr_j}$ die j -te **Randverteilung** oder **Marginalverteilung** von Q .

Ist $Q = P^X$ für eine Zufallsvariable $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}) \rightarrow (\times_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{A}_i)$, so heißt $Q_j = P^{pr_j \circ X} = P^{X_j}$ auch die j -te Randverteilung von X .

Die Randverteilungen Q_j , $1 \leq j \leq n$, sind durch Q natürlich eindeutig festgelegt. Die Umkehrung gilt jedoch nicht, d.h. i.allg. gibt es zu vorgegebenen Verteilungen Q_j , $1 \leq j \leq n$, mehrere Wahrscheinlichkeitsmaße auf dem Produktraum, die diese Verteilungen als Randverteilungen besitzen! \square

3.17 Bemerkung und Definition Speziell im Fall $(\Omega_i, \mathcal{A}_i) = (\mathbb{R}, \mathbb{B})$ für alle $1 \leq i \leq n$ wird $\mathbb{B}^n = \bigotimes_{i=1}^n \mathbb{B}$ als **Borel- σ -Algebra** auf \mathbb{R}^n bezeichnet. Man kann zeigen:

$$\begin{aligned} \mathbb{B}^n &= \sigma \left\{ \bigtimes_{i=1}^n (-\infty, x_i] \mid x_i \in \mathbb{R} \forall 1 \leq i \leq n \right\} \\ &= \sigma \left\{ \bigtimes_{i=1}^n (a_i, b_i] \mid -\infty < a_i \leq b_i < \infty \forall 1 \leq i \leq n \right\} \\ &= \sigma \{ O \subset \mathbb{R} \mid O \text{ offen} \}. \end{aligned}$$

Wegen der Eindeutigkeitsaussage des Maßfortsetzungssatzes ist daher ein Wahrscheinlichkeitsmaß Q auf $(\mathbb{R}^n, \mathbb{B}^n)$ (bzw. P^X für ein $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathbb{B}^n)$) eindeutig bestimmt durch die zugehörige **multivariate Verteilungsfunktion**

$$F : \mathbb{R}^n \rightarrow [0, 1]; \quad (x_1, \dots, x_n) \mapsto Q \left(\bigtimes_{i=1}^n (-\infty, x_i] \right)$$

(bzw.

$$F : \mathbb{R}^n \rightarrow [0, 1]; \quad (x_1, \dots, x_n) \mapsto P^X \left(\bigtimes_{i=1}^n (-\infty, x_i] \right) = P \{ X_i = pr_i \circ X \leq x_i \forall 1 \leq i \leq n \}$$

). Kann man F darstellen als

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \left(\int_{-\infty}^{x_2} \cdots \left(\int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_n \right) \cdots dt_2 \right) dt_1 \quad \forall x_i \in \mathbb{R}, 1 \leq i \leq n,$$

für eine n -fach Riemann-integrierbare (oder Lebesgue-integrierbare) Funktion f , so heißt diese (**multivariate Riemann-)**Dichte (bzw. Lebesgue-Dichte) von F bzw. Q (bzw. P^X , bzw. X).

Es gilt dann für alle Mengen $A \in \mathbb{B}^n$, für die $f \cdot 1_A$ Riemann-integrierbar ist, (bzw. im Fall von Lebesgue-Dichten ohne weitere Voraussetzung)

$$Q(A) = \int_{-\infty}^{x_1} \left(\int_{-\infty}^{x_2} \cdots \left(\int_{-\infty}^{x_n} f(t_1, \dots, t_n) 1_A(t_1, \dots, t_n) dt_n \right) \cdots dt_2 \right) dt_1.$$

Die Verteilungsfunktion der j -ten Randverteilung von Q ($1 \leq j \leq n$) ist dann gegeben durch

$$\begin{aligned} F_j(x_j) &= Q\left(\mathbb{R}^{j-1} \times (-\infty, x_j] \times \mathbb{R}^{n-j}\right) \\ &= F\left(\underbrace{\infty, \dots, \infty}_{(j-1) \text{ mal}}, x_j, \underbrace{\infty, \dots, \infty}_{(n-j) \text{ mal}}\right) \\ &= \lim_{x_i \uparrow \infty \forall i \neq j} F(x_1, \dots, x_n), \quad x_j \in \mathbb{R}. \end{aligned}$$

Besitzt Q eine Dichte f , so besitzt die j -te Randverteilung Q_j die Dichte

$$f_j(x_j) = \int_{-\infty}^{\infty} \cdots \left(\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \cdots \left(\int_{-\infty}^{\infty} f(t_1, \dots, t_{j-1}, x_j, t_{j+1}, \dots, t_n) dt_n \right) \cdots dt_{j+1} \right) dt_{j-1} \right) \cdots dt_1, \quad x_j \in \mathbb{R}.$$

Die Umkehrung gilt nicht, d.h. wenn alle Q_j , $1 \leq j \leq n$, jeweils Dichten besitzen, so muss Q deshalb nicht notwendig eine Dichte besitzen. \square

3.18 Beispiel Die **multivariate Standardnormalverteilung** $\mathcal{N}_{(0, E_n)}$ auf $(\mathbb{R}^n, \mathbb{B}^n)$ ist definiert als Wahrscheinlichkeitsmaß mit der Dichte

$$f(x_1, \dots, x_n) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n x_i^2\right), \quad x_1, \dots, x_n \in \mathbb{R}.$$

Hierbei bezeichnet E_n die $n \times n$ -Einheitsmatrix. \square

3.19 Satz (i) Jede stetige Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ist auch $\mathbb{B}^n, \mathbb{B}^m$ -messbar.

(ii) Sind $X_i : (\Omega, \mathcal{A}) \rightarrow (\Omega_i, \mathcal{A}_i)$ Zufallsvariablen und ist $f : (\times_{i=1}^n \Omega_i, \otimes_{i=1}^n \mathcal{A}_i) \rightarrow (\tilde{\Omega}, \tilde{\mathcal{A}})$ eine messbare Abbildung, so ist auch $f(X_1, \dots, X_n) : (\Omega, \mathcal{A}) \rightarrow (\tilde{\Omega}, \tilde{\mathcal{A}})$ messbar.

(iii) Sind $X, Y : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathbb{B})$ Zufallsvariablen, so auch $-X, X + Y, XY, X^+ = \max(X, 0), X^- = \min(X, 0)$ und $|X|$. Gilt $Y(\omega) \neq 0$ für alle $\omega \in \Omega$, so sind auch $1/Y$ und X/Y Zufallsvariablen. \square

3.20 Satz (Transformationssatz für Dichten) Ist (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, besitzt $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^n, \mathbb{B}^n)$ die Dichte f_X und gilt $P\{X \in M\} = 1$ für eine offene, zusammenhängende Menge $M \subset \mathbb{R}^n$, ist schließlich $g : M \rightarrow \mathbb{R}^n$ eine injektive, stetig differenzierbare Abbildung mit Umkehrabbildung $g^{-1} : g(M) \rightarrow M$ und invertierbarer Jacobi-Matrix Dg , dann besitzt $g(x) : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^n, \mathbb{B}^n)$ die Dichte

$$f_{g(X)}(y) = \frac{f_X(g^{-1}(y))}{|\det Dg(g^{-1}(y))|} 1_{g(M)}(y), \quad y \in \mathbb{R}^n.$$

\square

3.22 Satz Seien $X_n : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathbb{B})$, $n \in \mathbb{N}$, Zufallsvariablen. Dann sind auch $\sup_{n \in \mathbb{N}} X_n, \inf_{n \in \mathbb{N}} X_n, \limsup_{n \in \mathbb{N}} X_n, \liminf_{n \in \mathbb{N}} X_n, \lim_{n \in \mathbb{N}} X_n$ jeweils \mathcal{A}, \mathbb{B} -messbar (also Zufallsvariablen), wenn diese Abbildungen jeweils wohldefiniert sind mit Werten in \mathbb{R} (also z.B. $\lim_{n \in \mathbb{N}} X_n(\omega)$ für alle $\omega \in \Omega$ in \mathbb{R} existiert). \square

3.23 Bemerkung Versieht man $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ mit der σ -Algebra $\bar{\mathbb{B}} = \{B, B \cup \{-\infty\}, B \cup \{\infty\}, B \cup \{-\infty, \infty\} \mid B \in \mathbb{B}\}$, so sind für beliebige $\mathcal{A}, \bar{\mathbb{B}}$ -messbare Abbildungen $X_n : \Omega \rightarrow \bar{\mathbb{R}}$ auch die Abbildungen $\sup_{n \in \mathbb{N}} X_n, \inf_{n \in \mathbb{N}} X_n, \limsup_{n \in \mathbb{N}} X_n, \liminf_{n \in \mathbb{N}} X_n$ $\mathcal{A}, \bar{\mathbb{B}}$ -messbar und $\lim_{n \in \mathbb{N}} X_n$ ist $\mathcal{A}, \bar{\mathbb{B}}$ -messbar, wenn der Limes immer existiert. \square

4 Unabhängigkeit und bedingte Wahrscheinlichkeiten

4.1 Definition Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und I eine beliebige Indexmenge. Ereignisse A_i , $i \in I$, heißen **(P)-stochastisch unabhängig**, falls

$$P\left(\bigcap_{j=1}^m A_{i_j}\right) = \prod_{j=1}^m P(A_{i_j}) \quad (\text{“Produktregel”})$$

für jede beliebige Auswahl von paarweise verschiedenen Indizes $i_j \in I$, $1 \leq j \leq m$, $m \in \mathbb{N}$, mit $i_j \neq i_k$ für $1 \leq j < k \leq m$. \square

Warnung: Für den Nachweis der Unabhängigkeit von A_1, \dots, A_n reicht es **nicht** aus zu zeigen, dass

- $P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot \dots \cdot P(A_n)$ oder
- $P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$ für alle $1 \leq i < j \leq n$ (falls $n > 2$).

(Im zweiten Fall spricht man von paarweiser Unabhängigkeit.)

Gegeben seien n Zufallsexperimente, die durch Wahrscheinlichkeitsräume $(\Omega_i, \mathcal{A}_i, P_i)$, $1 \leq i \leq n$, beschrieben werden.

Frage: *Welcher Wahrscheinlichkeitsraum beschreibt dann die unabhängige Hintereinanderausführung dieser Experimente?*

Gemäß Definition 3.12 und Satz 3.13 sollte man als Grundraum

$$(\Omega, \mathcal{A}) = \left(\bigtimes_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{A}_i \right)$$

wählen. Zur Konstruktion eines geeigneten Wahrscheinlichkeitsmaßes P auf diesem Produktraum beachte man, dass einem Ereignis $A_i \in \mathcal{A}_i$ im i -ten Telexperiment das Ereignis

$$B_i = \bigtimes_{j=1}^{i-1} \Omega_j \times A_i \times \bigtimes_{j=i+1}^n \Omega_j$$

im Gesamtexperiment entspricht. Daher sollte gelten:

$$P(B_i) = P_i(A_i).$$

Ferner sollten die Ereignisse B_i für unterschiedliche i unabhängig sein, da sie Ereignissen in verschiedenen, als unabhängig angenommenen Telexperimenten entsprechen. Damit ergibt sich insbesondere

$$P\left(\bigtimes_{i=1}^n A_i\right) = P\left(\bigcap_{i=1}^n B_i\right) = \prod_{i=1}^n P(B_i) = \prod_{i=1}^n P_i(A_i).$$

4.3 Satz und Definition *Es gibt in der oben beschriebenen Situation genau ein Wahrscheinlichkeitsmaß P auf (Ω, \mathcal{A}) , so dass*

$$P\left(\bigtimes_{i=1}^n A_i\right) = \prod_{i=1}^n P_i(A_i) \quad \forall A_i \in \mathcal{A}_i, 1 \leq i \leq n.$$

P heißt **Produktwahrscheinlichkeitsmaß** der P_i , $1 \leq i \leq n$, und wird mit $\bigotimes_{i=1}^n P_i$ bezeichnet. (Ω, \mathcal{A}, P) heißt **Produktwahrscheinlichkeitsraum** der Wahrscheinlichkeitsräume $(\Omega_i, \mathcal{A}_i, P_i)$, $1 \leq i \leq n$. \square

4.4 Beispiel (i) Sind alle P_i diskret mit Zähldichten f_i , so ist $\bigotimes_{i=1}^n P_i$ diskret mit Zähldichte $f(\omega_1, \dots, \omega_n) = \prod_{i=1}^n f_i(\omega_i)$ für alle $\omega_i \in \Omega_i$, $1 \leq i \leq n$. Ein diskretes Wahrscheinlichkeitsmaß auf $(\bigtimes_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{A}_i)$ ist ein Produktwahrscheinlichkeitsmaß genau dann, wenn seine Zähldichte von obiger Form ist.

(ii) Ist $(\Omega_i, \mathcal{A}_i) = (\mathbb{R}, \mathbb{B})$ und bezeichnet F_i die Verteilungsfunktion von P_i für alle $1 \leq i \leq n$, so ist die Verteilungsfunktion von $\bigotimes_{i=1}^n P_i$ gegeben durch

$$F(x_1, \dots, x_n) = \prod_{i=1}^n F_i(x_i)$$

für alle $(x_1, \dots, x_n) \in \mathbb{R}^n$. Ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^n, \mathbb{B}^n)$ ist ein Produktwahrscheinlichkeitsmaß genau dann, wenn seine Verteilungsfunktion obige Produktgestalt hat.

(iii) Besitzen in der Situation von (ii) die Wahrscheinlichkeitsmaße P_i (Lebesgue- oder Riemann-)Dichten f_i , $1 \leq i \leq n$, so besitzt $\bigotimes_{i=1}^n P_i$ die multivariate Dichte

$$f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$$

für alle $(x_1, \dots, x_n) \in \mathbb{R}^n$. Ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^n, \mathbb{B}^n)$, das eine Dichte besitzt, ist ein Produktwahrscheinlichkeitsmaß genau dann, wenn seine Dichte von obiger Gestalt gewählt werden kann. \square

4.5 Definition (i) Zufallsvariablen $X_i : (\Omega, \mathcal{A}) \rightarrow (\Omega_i, \mathcal{A}_i)$, $1 \leq i \leq n$, auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) heißen **(P-)stochastisch unabhängig**, falls $P^{(X_1, \dots, X_n)} = \bigotimes_{i=1}^n P^{X_i}$, d.h.

$$P\{X_i \in A_i \forall 1 \leq i \leq n\} = \prod_{i=1}^n P\{X_i \in A_i\} \quad \forall A_i \in \mathcal{A}_i, 1 \leq i \leq n.$$

(ii) Sei I eine beliebige Indexmenge. Zufallsvariablen $X_i : (\Omega, \mathcal{A}) \rightarrow (\Omega_i, \mathcal{A}_i)$, $i \in I$, heißen **(P-)stochastisch unabhängig**, falls für alle $n \in \mathbb{N}$ und alle $i_1, \dots, i_n \in I$ mit $i_j \neq i_k$ für $1 \leq j < k \leq n$ die Zufallsvariablen X_{i_1}, \dots, X_{i_n} stochastisch unabhängig im Sinne von (i) sind. \square

4.6 Satz Zufallsvariablen $X_i : (\Omega, \mathcal{A}, P) \rightarrow (\Omega_i, \mathcal{A}_i)$, $1 \leq i \leq n$, sind genau dann unabhängig, wenn es \cap -stabile Erzeugendensysteme $\mathcal{C}_i \subset \mathcal{A}_i$ mit $\sigma(\mathcal{C}_i) = \mathcal{A}_i$, $1 \leq i \leq n$, gibt, so dass

$$P\{X_i \in C_i \forall 1 \leq i \leq n\} = \prod_{i=1}^n P\{X_i \in C_i\} \quad \forall C_i \in \mathcal{C}_i, 1 \leq i \leq n. \quad \square$$

4.8 Definition Sind X, Y unabhängige \mathbb{R} -wertige Zufallsvariablen mit Verteilungsfunktion F bzw. G , so heißt die Verteilung P^{X+Y} von $X+Y$ die **Faltung** von P^X und P^Y und wird mit $P^X * P^Y$ bezeichnet. Die Verteilungsfunktion von $X+Y$ heißt die **Faltung** von F und G und wird mit $F * G$ bezeichnet. \square

4.9 Satz (i) Ist in der Situation von Definition 4.8 P^X ein diskretes Wahrscheinlichkeitsmaß mit Zähldichte f , so gilt

$$F * G(x) = \sum_{t \in \mathbb{R}} f(t) \cdot G(x - t) \quad \forall x \in \mathbb{R}.$$

Ist auch P^Y diskret mit Zähldichte g , so ist $P^X * P^Y$ diskret mit Zähldichte

$$f * g(x) := \sum_{t \in \mathbb{R}} f(t) \cdot g(x - t), \quad x \in \mathbb{R}.$$

(ii) Besitzt F eine (Lebesgue- oder Riemann-)Dichte f , so gilt

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

Besitzt auch G eine Dichte g , so definiert

$$f * g(x) := \int_{-\infty}^{\infty} f(t) \cdot g(x - t) dt \quad x \in \mathbb{R},$$

eine Dichte von $F * G$. \square

4.11 Definition Ist (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{A}$ mit $P(B) > 0$. Dann heißt

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

die **bedingte Wahrscheinlichkeit von A gegeben B** . \square

4.15 Satz Sei $I \neq \emptyset$ abzählbar, (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $B_i \in \mathcal{A}$, $i \in I$, eine Zerlegung von Ω (d.h. die B_i sind disjunkt und $\bigcup_{i \in I} B_i = \Omega$) mit $P(B_i) > 0$ für alle $i \in I$. Dann gilt:

$$(i) P(A) = \sum_{i \in I} P(A|B_i) \cdot P(B_i) \quad \forall A \in \mathcal{A} \quad (\text{Satz von der totalen Wahrscheinlichkeit})$$

$$(ii) \quad P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{i \in I} P(A|B_i) \cdot P(B_i)} \quad \forall A \in \mathcal{A} \text{ mit } P(A) > 0 \quad (\text{Satz von Bayes}) \quad \square$$

Die gemeinsame Verteilung zweier diskreter Zufallsvariablen kann man dadurch festlegen, dass man zunächst die Verteilung der ersten Zufallsvariablen bestimmt und dann für jeden möglichen Wert dieser ersten Zufallsvariablen die Wahrscheinlichkeit angibt, dass die zweite Zufallsvariable einen bestimmten (aber beliebigen) Wert annimmt. Der 2. Schritt bei diesem Zugang führt auf den Begriff der bedingten Verteilung.

4.18 Satz und Definition Sei $X : (\Omega, \mathcal{A}) \rightarrow (\Omega_X, \mathcal{A}_X)$ eine diskrete Zufallsvariable mit Zähldichte f_X und $Y : (\Omega, \mathcal{A}) \rightarrow (\Omega_Y, \mathcal{A}_Y)$ eine beliebige weitere Zufallsvariable auf demselben Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Für jedes $x \in \Omega_X$ mit $f_X(x) = P\{X = x\} > 0$ heißt das durch

$$P^{Y|X=x}(A) := P(Y \in A \mid X = x) := \frac{P\{Y \in A, X = x\}}{P\{X = x\}}, \quad A \in \mathcal{A}_Y,$$

definierte Wahrscheinlichkeitsmaß auf $(\Omega_Y, \mathcal{A}_Y)$ die **bedingte Verteilung von Y gegeben $X = x$** .

Ist Y ebenfalls diskret, so auch $(X, Y) : (\Omega, \mathcal{A}) \rightarrow (\Omega_X \times \Omega_Y, \mathcal{A}_X \otimes \mathcal{A}_Y)$. Bezeichnet $f_{(X,Y)}$ die zugehörige Zähldichte, so ist $P^{Y|X=x}$ diskret mit Zähldichte

$$f_{Y|X=x}(y) := \frac{f_{(X,Y)}(x, y)}{f_X(x)} = \frac{f_{(X,Y)}(x, y)}{\sum_{\tilde{y} \in \Omega_Y} f_{(X,Y)}(x, \tilde{y})} \quad (4.1)$$

für alle $x \in \Omega_X$ mit $f_X(x) > 0$. $f_{Y|X=x}$ heißt **bedingte Zähldichte von Y gegeben $X = x$** . \square

Gibt man sich umgekehrt eine diskrete Verteilung P^X vor sowie für alle $x \in \Omega_X$ mit $P\{x\} > 0$ eine bedingte Verteilung von Y gegeben $X = x$, so ist damit die gemeinsame Verteilung von (X, Y) eindeutig festgelegt. Diese Idee führt auf den Begriff der Koppelung.

4.19 Satz und Definition Sei $(\Omega_1, \mathcal{A}_1, P_1)$ ein diskreter Wahrscheinlichkeitsraum mit Zähldichte f_1 und $(\Omega_2, \mathcal{A}_2)$ ein weiterer Messraum. Zu jedem $\omega_1 \in \Omega_1$ sei $P_2^1(\cdot \mid \omega_1)$ ein diskretes Wahrscheinlichkeitsmaß auf $(\Omega_2, \mathcal{A}_2)$ mit Zähldichte $f_2^1(\cdot \mid \omega_1)$. Dann ist

$$f(\omega_1, \omega_2) := f_1(\omega_1) \cdot f_2^1(\omega_2 \mid \omega_1), \quad \omega_1 \in \Omega_1, \omega_2 \in \Omega_2,$$

die Zähldichte eines diskreten Wahrscheinlichkeitsmaßes auf $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, der sog. **Koppelung von P_1 und P_2^1** , i.Z. $P_1 \otimes P_2^1$. \square

4.20 Bemerkung In der Situation von Satz und Definition 4.19 ist P_1 die 1. Randverteilung von $P_1 \otimes P_2^1$.

Für die Definition der Koppelung ist die genaue Wahl von $P_2^1(\cdot \mid \omega_1)$ im Fall $P_1\{\omega_1\} = 0$ irrelevant. Andernfalls ist $P_2^1(\cdot \mid \omega_1)$ die bedingte Verteilung der Projektion pr_2 auf die 2. Koordinate gegeben, dass die Projektion pr_1 auf die 1. Koordinate den Wert ω_1 annimmt. Mit $P := P_1 \otimes P_2^1$ gilt daher

$$P^{(pr_1, pr_2)} = P = P^{pr_1} \otimes P^{pr_2|pr_1=\cdot},$$

wobei $P^{pr_2|pr_1=\cdot}$ als Schreibweise für P_2^1 , also für die Familie der bedingten Verteilungen $P^{pr_2|pr_1=\omega_1}$ für $\omega_1 \in \Omega_1$ verwendet wird.

Ebenso erhält man in der Situation von Satz und Definition 4.18

$$P^{(X,Y)} = P^X \otimes P^{Y|X=\cdot}. \quad \square$$

4.21 Beispiel Ist P_2 ein diskretes Wahrscheinlichkeitsmaß mit Zähldichte f_2 und gilt $f_2^1(\cdot | \omega_1) = f_2(\cdot)$ für alle $\omega_1 \in \Omega_1$, so ist $P_1 \otimes P_2^1 = P_1 \otimes P_2$ das Produktwahrscheinlichkeitsmaß von P_1 und P_2 . \square

4.22 Bemerkung Um die gemeinsame Verteilung eines Vektors (X_1, \dots, X_n) von diskreten Zufallsvariablen zu beschreiben, kann man wie folgt vorgehen:

- Lege P^{X_1} durch Wahl einer Zähldichte f_1 fest.
- Für alle $2 \leq i \leq n$ und alle $(\omega_1, \dots, \omega_{i-1}) \in \times_{j=1}^{i-1} \Omega_j$ wähle Zähldichten $f_i^{i-1}(\cdot | \omega_1, \dots, \omega_{i-1})$ auf Ω_i mit zugehörigem Wahrscheinlichkeitsmaß $P_i^{i-1}(\cdot | \omega_1, \dots, \omega_{i-1})$.
- Definiere rekursiv

$$P^{(X_1, \dots, X_i)} := P^{(X_1, \dots, X_{i-1})} \otimes P_i^{i-1}, \quad 2 \leq i \leq n.$$

Dann ist $P^{(X_1, \dots, X_n)}$ diskret mit Zähldichte

$$f(\omega_1, \dots, \omega_n) = f_1(\omega_1) \cdot \prod_{i=2}^n f_i^{i-1}(\omega_i | \omega_1, \dots, \omega_{i-1}).$$

Ferner ist $P_i^{i-1}(\cdot | \omega_1, \dots, \omega_{i-1})$ gerade die bedingte Verteilung von X_i gegeben $(X_1, \dots, X_{i-1}) = (\omega_1, \dots, \omega_{i-1})$. \square

4.24 Bemerkung (i) Besonders einfach wird die Beschreibung solcher mehrstufigen Zufallsexperimente, wenn für alle $2 \leq i \leq n$ die bedingte Verteilung von X_i gegeben $(X_1, \dots, X_{i-1}) = (\omega_1, \dots, \omega_{i-1})$ nur vom letzten Ergebnis ω_{i-1} abhängt. Man sagt in dem Fall, dass (X_1, \dots, X_n) die *Markov-Eigenschaft* besitzt.

(ii) Bei einer stetigen Zufallsvariablen X kann man die bedingte Verteilung von Y gegeben $X = x$ nicht mehr einfach durch $P^{Y|X=x}(A) = P\{Y \in A, X = x\} / P\{X = x\}$ definieren, da der Nenner gleich 0 ist. Sind jedoch $X, Y \in \mathbb{R}^k$ bzw. \mathbb{R}^m -wertig mit gemeinsamer (Riemann- oder Lebesgue-)Dichte $f_{(X,Y)}$ von (X, Y) , so kann man in Analogie zu (4.1) eine *bedingte Dichte von Y gegeben X = x* definieren durch

$$f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x, y)}{f_X(x)}, \quad y \in \mathbb{R}^m,$$

wobei $f_X(x) := \int_{\mathbb{R}^m} f_{(X,Y)}(x, \tilde{y}) d\tilde{y}$ eine Dichte von X bezeichnet. Das Wahrscheinlichkeitsmaß auf $(\mathbb{R}^m, \mathbb{B}^m)$ mit Dichte $f_{Y|X=x}$ heißt dann die *bedingte Verteilung*

von Y gegeben $X = x$. (Diese ist allerdings nur außerhalb einer P^X -Nullmenge von Werten für x eindeutig bestimmt.)

Im allgemeineren Kontext werden bedingte Verteilungen in der Vorlesung “Maßtheoretische Konzepte der Stochastik” behandelt. \square

5 Erwartungswerte und Varianzen

5.2 Definition Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Eine Zufallsvariable $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathbb{B})$ heißt **diskret**, falls $X(\Omega)$ abzählbar ist.

Gilt $\sum_{x \in X(\Omega)} |x| P\{X = x\} < \infty$, so heißt X (**endlich**) (P) -**integrierbar** und

$$E(X) := E_P(X) := \sum_{x \in X(\Omega)} x P\{X = x\} = \sum_{x \in X(\Omega)} x P^X\{x\} \quad (5.1)$$

wird **Erwartungswert von X (unter P)** oder **Mittelwert von P^X** genannt.

$\mathcal{L}_1^d := \mathcal{L}_1^d(P) := \{X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathbb{B}) \mid X \text{ diskret mit } \sum_{x \in X(\Omega)} |x| P\{X = x\} < \infty\}$ \square

5.3 Bemerkung (i) Wegen $\sum_{x \in X(\Omega)} |x| P\{X = x\} < \infty$ spielt die Summationsreihenfolge in der Definition von $E(X)$ keine Rolle. Außerdem ist X genau dann endlich integrierbar, wenn $|X|$ endlich integrierbar ist, und es gilt dann $E(|X|) = \sum_{x \in X(\Omega)} |x| P\{X = x\}$.

(ii) Ist $X \geq 0$ (d.h. ist X $[0, \infty)$ -wertig), so kann man $E(X)$ stets wie in (5.1) definieren, also auch dann, wenn $\sum_{x \in X(\Omega)} |x| P\{X = x\} = \infty$. In dem Fall ist dann allerdings $E(X) = \infty$.

Ist X beliebig \mathbb{R} -wertig, so definiert man den Positivteil $X^+ = \max(X, 0)$ und den Negativteil $X^- := -\min(X, 0) = \max(-X, 0)$ von X (also z.B. $X^+(\omega) = X(\omega)$, falls $X(\omega) \geq 0$, und $X^+(\omega) = 0$ sonst), so dass $X = X^+ - X^-$. Für die nicht-negativen Zufallsvariablen X^+ und X^- kann man wie oben stets einen Erwartungswert definieren. Man setzt daher

$$E(X) := E(X^+) - E(X^-), \quad \text{falls } E(X^+) < \infty \text{ oder } E(X^-) < \infty.$$

Hierbei wird wie üblich für alle $x \in \mathbb{R}$ definiert $\infty + x := \infty$ und $x - \infty := -\infty$. Der so definierte Erwartungswert von X kann somit also Werte in $[-\infty, \infty]$ annehmen. \square

5.5 Satz Für Zufallsvariablen $X, Y \in \mathcal{L}_1^d(P)$ gilt

- (i) Gilt $X \leq Y$ P -f.s., so folgt $E(X) \leq E(Y)$ (Monotonie des Erwartungswerts)
- (ii) Für alle $a \in \mathbb{R}$ ist $aX + Y \in \mathcal{L}_1^d(P)$ und es gilt $E(aX + Y) = aE(X) + E(Y)$ (Linearität des Erwartungswerts)
- (iii) Sind X, Y stochastisch unabhängig, so ist $XY \in \mathcal{L}_1^d(P)$ und es gilt $E(XY) = E(X) \cdot E(Y)$. \square

5.6 Satz (Transformationssatz für diskrete Zufallsvariablen) Sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\tilde{\Omega}, \tilde{\mathcal{A}})$ eine Zufallsvariable und $f : (\tilde{\Omega}, \tilde{\mathcal{A}}) \rightarrow (\mathbb{R}, \mathbb{B})$ diskret. Dann ist $f \circ X$ eine diskrete

Zufallsvariable, die genau dann endlich P -integrierbar ist, wenn f endlich P^X -integrierbar ist. In dem Fall gilt

$$E_P(f \circ X) = E_{P^X}(f) = \sum_{x \in X(\Omega)} f(x) P\{X = x\}.$$

□

Für nicht diskrete \mathbb{R} -wertige Zufallsvariablen X macht es keinen Sinn, den Erwartungswert wie in (5.1) zu definieren. Hat z.B. X eine stetige Verteilungsfunktion, so ist $P\{X = x\} = 0$ für alle $x \in \mathbb{R}$ und die rechte Seite von (5.1) ist stets gleich 0. Um einen Erwartungswert für solche allgemeineren \mathbb{R} -wertigen Zufallsvariablen zu definieren approximiert man diese durch eine Folge diskreter Zufallsvariablen.

Im folgenden bezeichne $[\cdot]$ die Gauß-Klammer, d.h. $[x] := \sup\{k \in \mathbb{Z} | k \leq x\}$ für $x \in \mathbb{R}$.

5.7 Satz und Definition Zu $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathbb{B})$ definiere diskrete Zufallsvariablen

$$X^{(n)}(\omega) := 2^{-n} [2^n X(\omega)] = \sum_{i \in \mathbb{Z}} i 2^{-n} 1_{[i 2^{-n}, (i+1) 2^{-n})}(X(\omega)), \quad \omega \in \Omega, n \in \mathbb{N}_0,$$

so dass $X^{(n)}(\omega)$ von unten monoton gegen $X(\omega)$ konvergiert für alle $\omega \in \Omega$.

Gilt $X^{(n)} \in \mathcal{L}_1^d(P)$ für ein $n \in \mathbb{N}_0$, so auch für alle $n \in \mathbb{N}_0$ und

$$E(X) := E_P(X) := \lim_{n \rightarrow \infty} E_P(X^{(n)})$$

heißt **Erwartungswert von X (unter P)** oder **Mittelwert von P^X** .

Weitere geläufige Schreibweisen für $E_P(X)$ sind $\int X dP$ und $\int X(\omega) P(d\omega)$.

$$\mathcal{L}_1 := \mathcal{L}_1(P) := \{X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathbb{B}) \mid X^{(0)} \in \mathcal{L}_1^d(P)\}$$

□

5.8 Bemerkung (i) $E_P(X)$ hängt nur von P^X ab.

- (ii) Analog zum Vorgehen in Bemerkung 5.3(ii) kann man für beliebige $X \geq 0$ einen Erwartungswert $E(X) = \lim_{n \rightarrow \infty} E(X^{(n)}) \in [0, \infty]$ definieren. Für \mathbb{R} -wertige Zufallsvariablen definiert man dann oft $E(X) = E(X^+) - E(X^-) \in [-\infty, \infty]$, falls $E(X^+) < \infty$ oder $E(X^-) < \infty$.

Falls $E(X^+) < \infty$ oder $E(X^-) < \infty$, so nennt man X **(quasi) P -integrierbar**, im Fall $\max(E(X^+), E(X^-)) < \infty$ **endlich P -integrierbar**. Im zweiten Fall, der genau dann vorliegt, wenn $E(|X|) < \infty$, stimmt diese Definition mit der aus Satz und Definition 5.7 überein.

- (iii) Für $X \in \mathcal{L}_1^d$ stimmen die Erwartungswerte aus den Definitionen 5.2 und 5.7 überein. □

5.9 Satz Die Aussagen aus Satz 5.5 gelten auch allgemeiner für $X, Y \in \mathcal{L}_1(P)$. □

5.10 Satz (Transformationssatz) Sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\tilde{\Omega}, \tilde{\mathcal{A}})$ und $f : (\tilde{\Omega}, \tilde{\mathcal{A}}) \rightarrow (\mathbb{R}, \mathbb{B})$. Dann gilt

$$f \circ X \in \mathcal{L}_1(P) \iff f \in \mathcal{L}_1(P^X)$$

und in dem Fall gilt

$$E_P(f \circ X) = \int f \circ X dP = \int f dP^X = E_{P^X}(f).$$

□

5.11 Satz Besitzt X eine Lebesgue-Dichte f , so gilt

$$f \in \mathcal{L}_1(P) \iff \int_{\mathbb{R}} |x| f(x) dx < \infty$$

und in dem Fall gilt

$$E_P(X) = \int_{\mathbb{R}} x f(x) dx.$$

□

5.13 Bemerkung (Zusammenhang zwischen Erwartungswert und Lebesgue-Integral)

Bezeichnet $U_{(i-1, i]}$ die Gleichverteilung auf $(i-1, i]$, so kann man das **Lebesgue-Maß** \mathbb{X}^1 definieren durch

$$\mathbb{X}^1(B) := \sum_{i \in \mathbb{Z}} U_{(i-1, i]}(B \cap (i-1, i]), \quad B \in \mathcal{B}.$$

Eine allgemeinere Version des Maßfortsetzungssatzes (Satz 2.6) zeigt dann, dass $\mathbb{X}^1 : \mathbb{B} \rightarrow [0, \infty]$ die einzige σ -additive Abbildung ist mit $\mathbb{X}^1((a, b]) = b - a$ für alle $-\infty < a \leq b < \infty$.

Man kann nun analog zum Erwartungswert schrittweise das **Lebesgue-Integral** definieren:

- Ist $X : (\mathbb{R}, \mathbb{B}) \rightarrow ([0, \infty), \mathcal{B}[0, \infty))$ eine *diskrete* Zufallsvariable, so definiere

$$\int X d\mathbb{X}^1 := \sum_{x \in X(\mathbb{R})} x \mathbb{X}^1\{X = x\} \in [0, \infty]$$

- Ist $X : (\mathbb{R}, \mathbb{B}) \rightarrow ([0, \infty), \mathcal{B}[0, \infty))$ nicht notwendig diskret, so definiere

$$\int X d\mathbb{X}^1 := \lim_{n \rightarrow \infty} \int X^{(n)} d\mathbb{X}^1 \in [0, \infty]$$

- Ist $X : (\mathbb{R}, \mathbb{B}) \rightarrow (\mathbb{R}, \mathbb{B})$, so definiere

$$\int X d\mathbb{X}^1 := \int X^+ d\mathbb{X}^1 - \int X^- d\mathbb{X}^1,$$

falls $\min(\int X^+ d\mathbb{X}^1, \int X^- d\mathbb{X}^1) < \infty$.

(Zu dem gleichen Ergebnis kommt man im Übrigen, wenn man das Integral als $\sum_{i \in \mathbb{Z}} E_{U_{(i-1,i]}}(X 1_{(i-1,i]})$ definiert, wie es die obige Darstellung des Lebesgue-Maßes nahe legt.)

Ein Vergleich mit dem Riemann-Integral zeigt:

- Das Riemann-Integral $\int_a^b X(t) dt$ wird als Grenzwert von Summen definiert, die man erhält, wenn man den *Integrationsbereich* $(a, b]$ *diskretisiert*, d.h. X auf kleinen Intervallen $(t_k, t_{k+1}]$ durch eine Konstante ersetzt.
- Das Lebesgue-Integral wird als Grenzwert von Summen definiert, die man durch *Diskretisierung des Wertebereichs* erhält, d.h. X wird auf Mengen der Form $\{X \in [x_k, x_{k+1})\}$ für kurze Intervalle $[x_k, x_{k+1})$ durch eine Konstante approximiert.

Ist X Riemann-integrierbar auf $(a, b]$, so gilt

$$\int_a^b X(t) dt = \int_{(a,b]} X d\mathbb{X}^1 := \int X 1_{(a,b]} d\mathbb{X}^1.$$

Daher schreiben wir meistens (etwas unpräzise) auch das Lebesgue-Integral in der Form $\int_a^b X(t) dt$. □

5.14 Satz Ist Q ein Wahrscheinlichkeitsmaß auf (\mathbb{R}, \mathbb{B}) mit Lebesgue-Dichte f , so gilt für alle $B \in \mathbb{B}$

$$Q(B) = \int_B f d\mathbb{X}^1 := \int f 1_B d\mathbb{X}^1.$$

□

In Verallgemeinerung von Satz 5.11 bzw. als Analogon zu Satz 5.6 gilt

5.15 Satz Besitzt P^X die Lebesgue-Dichte f und ist $g : (\mathbb{R}, \mathbb{B}) \rightarrow (\mathbb{R}, \mathbb{B})$ eine messbare Abbildung, so gilt $g \circ X \in \mathcal{L}_1(P)$ genau dann, wenn $\int |g| \cdot f d\mathbb{X}^1 < \infty$. In diesem Fall gilt

$$E_P(g(X)) = \int g f d\mathbb{X}^1.$$

□

$E(X)$ gibt einen “mittleren Wert” von X an. Oft sind noch weitere Kennzahlen der Verteilung von X von Interesse, so z.B. Größen, die die mittlere quadratische Abweichung zwischen X und $E(X)$ als Maß für die Variabilität von X angeben.

5.16 Definition Zu einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) definiert man die Menge der **quadratintegrierbaren Zufallsvariablen** als

$$\mathcal{L}_2 := \mathcal{L}_2(P) := \{X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathbb{B}) \mid X^2 \in \mathcal{L}_1(P)\}.$$

Für $X \in \mathcal{L}_2(P)$ definiert man die **Varianz von X** als

$$\text{Var}(X) := E((X - E(X))^2).$$

□

5.17 Bemerkung (i) $Var(aX + b) = a^2 Var(X)$ für alle $a, b \in \mathbb{R}$ und $X \in \mathcal{L}_2$.

(ii) Im Sinne von Bemerkung 5.8(ii) kann man die Varianz von X für alle $X \in \mathcal{L}_1$ definieren. Für $X \in \mathcal{L}_1 \setminus \mathcal{L}_2$ gilt dann allerdings $Var(X) = \infty$.

Für $X \in \mathcal{L}_1$ erhält man $Var(X) = E(X^2) - (E(X))^2$. Insbesondere gilt $E(X^2) \geq (E(X))^2$ und daher

$$X \in \mathcal{L}_2 \iff E(X^2) < \infty \implies E(|X|) < \infty \iff X \in \mathcal{L}_1.$$

Etwas allgemeiner gilt für alle $a \in \mathbb{R}$

$$Var(X) = E((X - a)^2) - (E(X) - a)^2,$$

woraus sofort folgt, dass $E(X)$ die eindeutige Minimalstelle der Abbildung $a \mapsto E((X - a)^2)$ ist. \square

5.18 Satz Sind $X_1, \dots, X_n \in \mathcal{L}_2$ unabhängige Zufallsvariablen, so gilt

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i).$$

\square

5.19 Satz und Definition Seien $X, Y \in \mathcal{L}_2$. Dann gilt $XY \in \mathcal{L}_1$ und man definiert die **Kovarianz von X und Y** als

$$Cov(X, Y) := E((X - E(X))(Y - E(Y))) = E((X - E(X))Y) = E(XY) - E(X)E(Y).$$

Besitzen X und Y positive Varianzen, so wird die **Korrelation von X und Y** definiert als

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

X und Y heißen **unkorreliert**, falls $Cov(X, Y) = 0$. \square

5.20 Korollar Sind $X_1, \dots, X_n \in \mathcal{L}_2$ Zufallsvariablen, so gilt

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i) + 2 \sum_{1 \leq i < j \leq n} Cov(X_i, X_j).$$

\square

5.22 Bemerkung Sind $X, Y \in \mathcal{L}_1$ stochastisch unabhängig, so sind sie auch unkorreliert. Die Umkehrung gilt i.allg. nicht!

Die Korrelation misst nur den “linearen Zusammenhang” zwischen zwei Zufallsvariablen. Für die standardisierten Zufallsvariablen

$$X_0 := \frac{X - E(X)}{\sqrt{Var(X)}}, \quad Y_0 := \frac{Y - E(Y)}{\sqrt{Var(Y)}}$$

mit $E(X_0) = E(Y_0) = 0$ und $Var(X_0) = Var(Y_0) = 1$ ist $\rho := \rho(X, Y) = Cov(X_0, Y_0)$ gerade die eindeutige Minimalstelle der Abbildung $r \mapsto E((X_0 - rY_0)^2)$ und es gilt $E((X_0 - \rho Y_0)^2) = 1 - \rho^2$. \square

5.23 Satz (Tschebyschev'sche Ungleichung) Für $X \in \mathcal{L}_2(P)$ und $\varepsilon > 0$ gilt

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{Var(X)}{\varepsilon^2}.$$

\square

6 Grenzwertsätze

Der folgende Satz kann als Gegenstück zum empirischen Gesetz der Großen Zahlen in der mathematischen Modellwelt aufgefasst werden:

6.1 Satz (Schwaches Gesetz der Großen Zahlen) Seien $X_i : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathbb{B})$, $i \in \mathbb{N}$, (paarweise) unkorrelierte Zufallsvariablen mit identischem Erwartungswert $\mu \in \mathbb{R}$. Existiert eine Konstante $M < \infty$, so dass $\text{Var}(X_i) \leq M$ für alle $i \in \mathbb{N}$, so gilt für alle $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right\} = 0.$$

□

6.2 Korollar Sind $Y_i : (\Omega, \mathcal{A}, P) \rightarrow (\tilde{\Omega}, \tilde{\mathcal{A}})$, $i \in \mathbb{N}$, i.i.d., so gilt für alle $\tilde{A} \in \tilde{\mathcal{A}}$ und alle $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n 1_{\tilde{A}}(Y_i) - P\{Y_1 \in \tilde{A}\}\right| \geq \varepsilon\right\} = 0.$$

□

Dem Schwachen Gesetz der Großen Zahlen liegt der folgende Konvergenzbegriff zugrunde:

6.3 Definition Man sagt, dass Zufallsvariablen $X_n : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^k, \mathbb{B}^k)$, $n \in \mathbb{N}$, **P -stochastisch** gegen eine Zufallsvariable $Y : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^k, \mathbb{B}^k)$ **konvergieren**, falls

$$\lim_{n \rightarrow \infty} P\{\|X_n - Y\| \geq \varepsilon\} = 0.$$

In dem Fall schreibt man $X_n \rightarrow Y$ P -stochastisch oder $X_n \xrightarrow{P} Y$.

Dabei bezeichnet $\|\cdot\|$ eine beliebige Norm auf \mathbb{R}^k .

□

6.4 Bemerkung (i) Die stochastische Konvergenz hängt nicht von der Wahl der Norm ab, da alle Normen auf \mathbb{R}^k äquivalent sind.

(ii) Aus $X_n \rightarrow Y$ P -stochastisch folgt *nicht*, dass $X_n(\omega) \rightarrow Y(\omega)$ auch nur für ein $\omega \in \Omega$ gelten muss.

□

Das Schwache Gesetz der Großen Zahlen besagt, dass unter den obigen Bedingungen

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow E(X_1) \quad P\text{-stochastisch.}$$

6.7 Satz Seien $X, Y, X_n, Y_n : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^k, \mathbb{B}^k)$, $n \in \mathbb{N}$, Zufallsvariablen und $g : (\mathbb{R}^k, \mathbb{B}^k) \rightarrow (\mathbb{R}^l, \mathbb{B}^l)$. Dann gilt:

$$(i) \quad X_n \xrightarrow{P} X, \quad X_n \xrightarrow{P} Y \quad \implies \quad X = Y \quad P\text{-f.s.}$$

$$(ii) \quad X_n \xrightarrow{P} X, \quad g \text{ stetig auf einer Menge } M \text{ mit } P\{X \in M\} = 1 \quad \implies \quad g(X_n) \xrightarrow{P} g(X)$$

- (iii) $X_n = (X_{n,1}, \dots, X_{n,k}) \xrightarrow{P} X = (X_1, \dots, X_k) \iff X_{n,i} \xrightarrow{P} X_i \forall 1 \leq i \leq k$
- (iv) $X_n \xrightarrow{P} X, Y_n \xrightarrow{P} Y \implies X_n \pm Y_n \xrightarrow{P} X \pm Y, X_n^t Y_n \xrightarrow{P} X^t Y$
- (v) $k = 1, X_n \xrightarrow{P} X, Y_n \xrightarrow{P} Y, P\{Y = 0\} = 0 \implies \frac{X_n}{Y_n} 1_{\mathbb{R} \setminus \{0\}}(Y_n) \xrightarrow{P} \frac{X}{Y} 1_{\mathbb{R} \setminus \{0\}}(Y) \quad \square$

Zu i.i.d. Zufallsvariablen $X_i, i \in \mathbb{N}$, mit $E(X_i^2) < \infty$ definiere

$$Y_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - E(X_i)), \quad n \in \mathbb{N}.$$

Wegen der Tschebyschev'schen Ungleichung ist dann die Folge der Y_n stochastisch beschränkt im Sinne der folgenden Definition.

6.8 Definition Eine Folge von \mathbb{R}^k -wertigen Zufallsvariablen $Y_n, n \in \mathbb{N}$, heißt **stochastisch beschränkt**, falls

$$\lim_{c \rightarrow \infty} \limsup_{n \rightarrow \infty} P\{\|Y_n\| \geq c\} = 0.$$

In dem Fall schreibt man $Y_n = O_P(1)$. \square

Es soll im Folgenden untersucht werden, ob die Folge der Zufallsvariablen $Y_n, n \in \mathbb{N}$, sogar in einem geeigneten Sinne konvergiert.

6.9 Definition (i) Seien $F, F_n, n \in \mathbb{N}$, Verteilungsfunktionen auf \mathbb{R}^k . F_n **konvergiert schwach gegen F** (i.Z. $F_n \rightarrow F$ schwach), falls

$$F_n(x) \rightarrow F(x) \quad \text{für alle } x \in \mathcal{C}_F := \{y \in \mathbb{R}^k \mid F \text{ stetig in } y\}.$$

(ii) Seien $X, X_n : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^k, \mathbb{B}^k)$ mit Verteilungsfunktion F bzw. $F_n, n \in \mathbb{N}$. Die Zufallsvariablen X_n **konvergieren in Verteilung gegen X** , falls $F_n \rightarrow F$ schwach. In dem Fall sagt man auch, dass P^{X_n} **schwach gegen P^X konvergiert**. \square

6.10 Bemerkung (i) Während die stochastische Konvergenz von Zufallsvariablen eine Eigenschaft der Zufallsvariablen *als Abbildungen* sind, hängt die Konvergenz in Verteilung in der Tat nur von der *Verteilung* der Zufallsvariablen ab.

(ii) In der Definition der schwachen Konvergenz wäre es nicht sinnvoll, $F_n(x) \rightarrow F(x)$ für *alle* $x \in \mathbb{R}^k$ zu fordern. Sind z.B. $X_n = a_n, n \in \mathbb{N}$, und $X = a$ für reelle Konstanten $a_n \downarrow a$, so konvergiert $F_n(a) = 0$ für $n \rightarrow \infty$ nicht gegen $F(a) = 1$, d.h. selbst $X_n(\omega) \rightarrow X(\omega)$ für alle $\omega \in \Omega$ würde nicht die schwache Konvergenz implizieren. \square

6.11 Satz (Zentraler Grenzwertsatz von Lindeberg-Levy) Seien $X_i : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathbb{B})$, $i \in \mathbb{N}$, i.i.d. Zufallsvariablen mit $\mu := E(X_1) \in \mathbb{R}$ und $\sigma^2 := \text{Var}(X_1) \in (0, \infty)$. Dann gilt für die standardisierten Summen

$$S_n^* := \frac{1}{\sqrt{n\sigma^2}} \sum_{i=1}^n (X_i - \mu)$$

$$P^{S_n^*} \longrightarrow \mathcal{N}_{(0,1)} \quad \text{schwach,}$$

d.h.

$$P\{S_n^* \leq x\} \longrightarrow \Phi(x) = \int_{-\infty}^x \varphi(t) dt = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \quad \forall x \in \mathbb{R}.$$

□

6.12 Bemerkung Die Summen sind gerade so standardisiert, dass $E(S_n^*) = 0$ und $\text{Var}(S_n^*) = 1$. Ebenso gilt für eine $\mathcal{N}_{(0,1)}$ -verteilte Zufallsvariable Y gerade $E(Y) = 0$ und $\text{Var}(Y) = 1$. □

6.13 Korollar (Zentraler Grenzwertsatz von Moivre-Laplace) Für alle $p \in (0, 1)$ und $-\infty \leq a < b \leq \infty$ gilt

$$\lim_{n \rightarrow \infty} \mathcal{B}_{(n,p)} \left\{ k \in \{0, \dots, n\} \mid a < \frac{k - np}{\sqrt{np(1-p)}} \leq b \right\} = \Phi(b) - \Phi(a),$$

wobei wie üblich $\Phi(-\infty) := 0$ und $\Phi(\infty) := 1$ gesetzt wird. □

6.16 Lemma (Slutsky) Gilt für zwei Folgen von \mathbb{R} -wertigen Zufallsvariablen $X_n \rightarrow X$ in Verteilung und $Y_n \rightarrow a \in \mathbb{R}$ stochastisch, so folgt

$$\begin{aligned} X_n + Y_n &\longrightarrow X + a \quad \text{in Verteilung} \\ X_n \cdot Y_n &\longrightarrow X \cdot a \quad \text{in Verteilung} \end{aligned}$$

□

6.17 Satz In der Situation von Definition 6.9 gilt

$$\begin{aligned} X_n &\rightarrow X \text{ in Verteilung} \\ \iff E(g(X_n)) &\rightarrow E(g(X)) \text{ für alle } g : \mathbb{R}^k \rightarrow \mathbb{R} \text{ stetig, beschränkt.} \end{aligned}$$

□

6.18 Bemerkung (i) Für stetige, aber *unbeschränkte* Funktionen $g : \mathbb{R}^k \rightarrow \mathbb{R}$ folgt aus $X_n \rightarrow X$ in Verteilung i.allg. *nicht* $E(g(X_n)) \rightarrow E(g(X))$. Insbesondere gilt dann i.allg. *nicht* $E(X_n) \rightarrow E(X)$.

(ii) Die Charakterisierung aus Satz 6.17 wird auch verwendet, um die Konvergenz in Verteilung für Folgen von Zufallsvariablen mit Werten in beliebigen metrischen (oder topologischen) Räumen zu definieren. □

6.19 Korollar (“continuous mapping theorem”) Seien $X, X_n : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^k, \mathbb{B}^k)$ Zufallsvariablen und $h : \mathbb{R}^k \rightarrow \mathbb{R}^m$ eine stetige Abbildung, so folgt aus $X_n \rightarrow X$ in Verteilung auch $h(X_n) \rightarrow h(X)$ in Verteilung. \square

Für den Nachweis des Zentralen Grenzwertsatzes und allgemeiner der schwachen Konvergenz von Verteilungen auf dem \mathbb{R}^k sind gewisse solche Verteilungen eindeutig charakterisierende Funktionen nützlich.

Im folgenden wird der Raum \mathbb{C} der komplexen Zahlen mittels $z = x + iy \mapsto (x, y)^t$ (also der Zerlegung in Real- und Imaginärteil) mit dem \mathbb{R}^2 identifiziert und mit der entsprechenden Borel- σ -Algebra versehen. Eine \mathbb{C} -wertige Zufallsvariable Z ist daher immer in der Form $Z = X + iY : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{C}, \mathcal{B}^2)$ darstellbar mit \mathbb{R} -wertigen Zufallsvariablen X und Y . Man definiert dann

$$E(Z) := E(X) + iE(Y),$$

falls $X, Y \in \mathcal{L}_1(P)$.

6.20 Definition Zu einer \mathbb{R}^k -wertigen Zufallsvariablen definiert man die **charakteristische Funktion** (oder **Fourier-Transformierte**) $\varphi_X : \mathbb{R}^k \rightarrow \mathbb{C}$ durch

$$\varphi_X(t) := E(e^{i\langle t, X \rangle}) = E(\cos(\langle t, X \rangle)) + iE(\sin(\langle t, X \rangle))$$

definiert, wobei

$$\langle t, x \rangle := \sum_{i=1}^k t_i x_i$$

das übliche Skalarprodukt bezeichnet. \square

6.22 Satz Für unabhängige \mathbb{R}^k -wertige Zufallsvariablen X_j , $1 \leq j \leq n$, und $S_n := \sum_{j=1}^n X_j$ gilt

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

\square

6.23 Beispiel (i) Ist X standardnormalverteilt, so gilt $\varphi_X(t) = e^{-t^2/2}$ für alle $t \in \mathbb{R}$. \square

6.24 Satz Ist X eine \mathbb{R} -wertige Zufallsvariable mit $E(|X|^k) < \infty$, so ist φ_X k -mal stetig differenzierbar mit $\varphi^{(k)}(0) = i^k E(X^k)$. \square

6.25 Satz (Stetigkeitssatz von Levy-Cr amer) F ur \mathbb{R}^k -wertige Zufallsvariablen X, X_n , $n \in \mathbb{N}$, sind  quivalent

(i) $X_n \rightarrow X$ in Verteilung

(ii) $\lim_{n \rightarrow \infty} \varphi_{X_n}(t) = \varphi_X(t) \quad \forall t \in \mathbb{R}^k$ \square

6.26 Korollar (i) Die charakteristische Funktion bestimmt eine Verteilung eindeutig, d.h. sind X, Y \mathbb{R}^k -wertige Zufallsvariablen, so dass $\varphi_X(t) = \varphi_Y(t)$ für alle $t \in \mathbb{R}^k$, so gilt $P^X = P^Y$.

(ii) (“Cramér-Wold-device”) Für \mathbb{R}^k -wertige Zufallsvariablen $X, X_n, n \in \mathbb{N}$, sind äquivalent

(a) $X_n \rightarrow X$ in Verteilung

(b) $\langle u, X_n \rangle \rightarrow \langle u, X \rangle$ in Verteilung für alle $u \in \mathbb{R}^k$. □

6.27 Bemerkung In der Situation von Satz 6.11 gilt unter der Bedingung $E(|X_1|^3) < \infty$ sogar die Ungleichung von Berry-Esseen:

$$\sup_{x \in \mathbb{R}} |P\{S_n^* \leq x\} - \Phi(x)| \leq 0,8 \frac{E(|X_1 - \mu|^3)}{\sigma^3} \cdot \frac{1}{\sqrt{n}}.$$

□

Oft werden Varianten des Zentralen Grenzwertsatzes benötigt, die nicht notwendig identisch verteilte Zufallsvariablen voraussetzen. Wir betrachten dazu im folgenden ein *Dreiecksschema* von Zufallsvariablen $(X_{n,j})_{1 \leq j \leq k_n, n \in \mathbb{N}}$ mit $k_n \rightarrow \infty$ für $n \rightarrow \infty$, die zeilenweise unabhängig sind.

6.28 Satz (Zentraler Grenzwertsatz von Lindeberg-Feller) Sei $(X_{n,j})_{1 \leq j \leq k_n, n \in \mathbb{N}}$ ein Dreiecksschema von \mathbb{R} -wertigen Zufallsvariablen, so dass für jedes feste $n \in \mathbb{N}$ die Zufallsvariablen $X_{n,j}$, $1 \leq j \leq k_n$, stochastisch unabhängig sind mit $E(|X_{n,j}|) < \infty$ und $\text{Var}(X_{n,j}) \in (0, \infty)$ für alle $1 \leq j \leq k_n, n \in \mathbb{N}$. Gilt für $\sigma_n := \sqrt{\sum_{j=1}^{k_n} \text{Var}(X_{n,j})}$ die folgende **Lindeberg-Bedingung**

$$L_n(\delta) := \frac{1}{\sigma_n^2} \sum_{j=1}^{k_n} E\left(\left(X_{n,j} - E(X_{n,j})\right)^2 1_{\{|X_{n,j} - E(X_{n,j})| > \delta \sigma_n\}}\right) \rightarrow 0 \quad \forall \delta > 0$$

für $n \rightarrow \infty$, dann

$$S_n^* := \frac{1}{\sigma_n} \sum_{j=1}^{k_n} (X_{n,j} - E(X_{n,j})) \rightarrow S \quad \text{in Verteilung}$$

für eine $\mathcal{N}_{(0,1)}$ -verteilte Zufallsvariable S . □

6.29 Bemerkung Die Lindeberg-Bedingung ist erfüllt, wenn ein $\varepsilon > 0$ existiert, so dass $E(|X_{n,j}|^{2+\varepsilon}) < \infty$ für alle $1 \leq j \leq k_n, n \in \mathbb{N}$, und die folgende *Ljapunov-Bedingung* erfüllt ist:

$$\frac{1}{\sigma_n^{2+\varepsilon}} \sum_{j=1}^{k_n} E(|X_{n,j} - E(X_{n,j})|^{2+\varepsilon}) \rightarrow 0$$

für $n \rightarrow \infty$. □

Ist eine Zufallsvariable (wie z.B. die empirische Varianz) darstellbar als Funktion von mehreren Summen unterschiedlicher Zufallsvariablen, so wird für die Approximation ihrer Verteilung eine multivariate (d.h. mehrdimensionale) Version des Zentralen Grenzwertsatzes benötigt.

6.31 Definition Sei X eine \mathbb{R}^m -wertige Zufallsvariable. Dann heißt

$$E(X) := (E(X_i))_{1 \leq i \leq m}$$

der **Erwartungswertvektor** und

$$\text{Cov}(X) := (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq m}$$

die **(Varianz-)Kovarianzmatrix** von X , falls die rechten Seiten jeweils wohldefiniert sind. \square

6.32 Bemerkung Jede Kovarianzmatrix ist symmetrisch und positiv semidefinit, denn für all $a \in \mathbb{R}^m$ gilt

$$a^t \text{Cov}(X) a = \sum_{i,j=1}^m a_i \text{Cov}(X_i, X_j) a_j = \text{Var}\left(\sum_{i=1}^m a_i X_i\right) \geq 0.$$

\square

6.33 Definition Seien X_1, \dots, X_k i.i.d. und $\mathcal{N}_{(0,1)}$ -verteilt, $X := (X_1, \dots, X_k)$, A eine $m \times k$ -Matrix, $\mu \in \mathbb{R}^m$ und $\Sigma := AA^t$. Dann heißt die Verteilung der \mathbb{R}^m -wertigen Zufallsvariable $Y = AX + \mu$ die **multivariate Normalverteilung mit Mittelwertvektor μ und Kovarianzmatrix Σ** . \square

6.34 Bemerkung (i) Man kann zeigen, dass P^Y nur von μ und Σ abhängt. Es gilt $E(Y) = \mu$ und $\text{Cov}(Y) = \Sigma$.

(ii) Ist Σ invertierbar, so hat $\mathcal{N}_{(\mu, \Sigma)}$ die Dichte

$$\varphi_{(\mu, \Sigma)}(y) = \frac{1}{\sqrt{(2\pi)^m \det(\Sigma)}} \exp\left(-\frac{1}{2}(y - \mu)^t \Sigma^{-1}(y - \mu)\right), \quad y \in \mathbb{R}^m.$$

Daraus ergibt sich zusammen mit (i), dass die Komponenten von Y genau dann stochastisch unabhängig sind, wenn sie unkorreliert sind und daher Σ und Σ^{-1} Diagonalmatrizen sind. \square

6.35 Satz Ist Y eine $\mathcal{N}_{(\mu, \Sigma)}$ -verteilte \mathbb{R}^m -wertige Zufallsvariable, B eine $l \times m$ -Matrix und $c \in \mathbb{R}^l$, so ist $Z = BY + c$ gerade $\mathcal{N}_{(c+B\mu, B\Sigma B^t)}$ -verteilt. \square

6.36 Satz (Multivariater Zentraler Grenzwertsatz) Seien X_j , $j \in \mathbb{N}$, i.i.d. \mathbb{R}^m -wertige Zufallsvariablen mit $\mu := E(X_1)$ und $\Sigma := \text{Cov}(X_1)$. Dann gilt für eine $\mathcal{N}_{(0, \Sigma)}$ -verteilte Zufallsvariable Y

$$\tilde{S}_n := \frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j - \mu) \longrightarrow Y \quad \text{in Verteilung.}$$

\square