

# Statistik III

Skript zur Vorlesung von Prof. Dr. Leonhard Held  
Wintersemester 2005/06

6. Februar 2007

Dieses Skript wurde im Wintersemester 2005/06 von Christiane Dargatz und Leonhard Held an der Universität München erstellt. Als Grundlage diente ein Skript zur gleichlautenden Vorlesung im WS04/05, das von Sebastian Petry freundlicherweise im L<sup>A</sup>T<sub>E</sub>X-Format zur Verfügung gestellt wurde.

Verbesserungen und Anregungen ausdrücklich erwünscht  
an [christiane.dargatz@stat.uni-muenchen.de](mailto:christiane.dargatz@stat.uni-muenchen.de) oder [leonhard.held@ifspm.uzh.ch](mailto:leonhard.held@ifspm.uzh.ch)!

# Inhaltsverzeichnis

<b>1</b>	<b>Ereignisse und deren Wahrscheinlichkeiten</b>	<b>2</b>
1.1	Einführung . . . . .	2
1.1.1	Formaler Rahmen . . . . .	2
1.1.2	Interpretation von Wahrscheinlichkeiten . . . . .	2
1.2	Ereignisse als Mengen . . . . .	2
1.3	Wahrscheinlichkeit . . . . .	5
1.4	Bedingte Wahrscheinlichkeit . . . . .	7
1.5	Unabhängigkeit . . . . .	14
1.6	Vollständigkeit und Produkträume . . . . .	15
1.6.1	Vollständigkeit . . . . .	16
1.6.2	Produkträume . . . . .	16
1.7	Ausführliche Beispiele . . . . .	16
<b>2</b>	<b>Zufallsvariablen und deren Verteilungen</b>	<b>18</b>
2.1	Zufallsvariablen . . . . .	18
2.2	Das Gesetz vom Durchschnitt (The Law of Averages) . . . . .	20
2.3	Diskrete und stetige Zufallsvariablen . . . . .	20
2.4	Zufallsvektoren . . . . .	22
2.5	Monte Carlo Simulation . . . . .	25
<b>3</b>	<b>Diskrete Zufallsvariablen</b>	<b>27</b>
3.1	Wahrscheinlichkeitsfunktion . . . . .	27
3.2	Unabhängigkeit . . . . .	28
3.3	Erwartungswert . . . . .	29
3.4	Beispiele für diskrete Zufallsvariablen . . . . .	35
3.5	Lineare stochastische Abhängigkeit . . . . .	37
3.6	Kovarianz- und Korrelationsmatrizen . . . . .	40
3.7	Bedingte Verteilungen und bedingte Erwartungen . . . . .	42
3.8	Summen von Zufallsvariablen . . . . .	49
<b>4</b>	<b>Stetige Zufallsvariablen</b>	<b>51</b>
4.1	Dichtefunktion . . . . .	51
4.2	Unabhängigkeit . . . . .	52
4.3	Erwartungswert . . . . .	52
4.4	Beispiele für stetige Zufallsvariablen . . . . .	54
4.4.1	Stetige Gleichverteilung . . . . .	54
4.4.2	Exponentialverteilung . . . . .	54
4.4.3	Normalverteilung . . . . .	55
4.4.4	Gammaverteilung . . . . .	55
4.4.5	Cauchy-Verteilung . . . . .	57
4.4.6	$t$ -Verteilung / Studentverteilung . . . . .	58
4.4.7	Betaverteilung . . . . .	58
4.5	Jensensche Ungleichung und Informationsungleichung . . . . .	59
4.6	Stochastische Abhängigkeit . . . . .	61
4.7	Bedingte Verteilungen und bedingte Erwartungen . . . . .	65

4.8	Funktionen von Zufallsvariablen . . . . .	72
4.9	Summen von Zufallsvariablen . . . . .	78
4.10	Multivariate Normalverteilung . . . . .	80
4.11	Von der Normalverteilung abgeleitete Verteilungen . . . . .	83
4.12	Simulation aus stetigen Verteilungen . . . . .	85
<b>5</b>	<b>Erzeugende Funktionen und deren Anwendungen</b>	<b>88</b>
5.1	Wahrscheinlichkeitserzeugende Funktionen . . . . .	88
5.2	Anwendung: Verzweigungsprozesse . . . . .	95
5.2.1	Verzweigungsprozesse mit Migration . . . . .	97
5.3	Allgemeiner Erwartungswertbegriff . . . . .	97
5.3.1	Notation . . . . .	97
5.3.2	Abstrakte Integration . . . . .	98
5.4	Momentenerzeugende Funktion und Kumulanten . . . . .	101
5.5	Charakteristische Funktionen . . . . .	103
5.6	Beispiele für charakteristische Funktionen . . . . .	106
5.7	Umkehrformel und Stetigkeitssatz . . . . .	107
5.8	Zwei Grenzwertsätze . . . . .	108
<b>6</b>	<b>Konvergenz von Zufallsvariablen</b>	<b>114</b>
6.1	Einführung . . . . .	114
6.2	Konvergenzarten von Zufallsvariablen . . . . .	114
6.3	Einige zusätzliche Resultate . . . . .	117
6.4	Gesetze der großen Zahlen . . . . .	119
	<b>Stichwortverzeichnis</b>	<b>121</b>

# 1 Ereignisse und deren Wahrscheinlichkeiten

## 1.1 Einführung

### 1.1.1 Formaler Rahmen

Die Menge  $\Omega$  aller möglichen Ergebnisse eines Experimentes heißt

- *Grundmenge*,
- *Ergebnisraum*,
- *Grundraum*.

Eine Menge  $A \subset \Omega$  heißt *Ereignis*. Mengen  $\{\omega\}$  mit Elementen  $\omega \in \Omega$  heißen *Elementarereignisse*. Jedem Ereignis  $A$  ordnen wir eine Zahl  $P(A) \in [0, 1]$  zu, die *Wahrscheinlichkeit* von  $A$ .

### 1.1.2 Interpretation von Wahrscheinlichkeiten

Wahrscheinlichkeit kann auf verschiedene Arten interpretiert werden. Die zwei wichtigsten Arten sind:

- (1) **Wahrscheinlichkeit als Wetteinsatz.** Wenn im Falle  $\omega \in A$  ein Gewinn  $G$  ausbezahlt wird, dann ist man (maximal) bereit, den Einsatz  $P(A) \cdot G$  zu setzen.

Bei dieser Interpretation handelt es sich um eine *subjektivistische Deutung* von Wahrscheinlichkeit. BRUNO DE FINETTI (1906-1985) kann als Vater dieser Interpretation gesehen werden.

- (2) **Wahrscheinlichkeit als Grenzwert.** Nimmt man an, dass ein Experiment beliebig oft wiederholt werden kann, wobei die einzelnen Ausgänge voneinander unabhängig sind, dann gilt folgendes Postulat:

*Die relative Häufigkeit des Eintretens von  $A$  bei  $n$  unabhängigen Wiederholungen des Zufallsexperiments,  $\hat{P}_n(A) = \frac{\#\{i: \omega_i \in A\}}{n}$ , konvergiert für  $n \rightarrow \infty$  gegen  $P(A)$ .*<sup>1</sup>

Als wichtigster Vertreter dieser *frequentistischen Interpretation* von Wahrscheinlichkeit ist RICHARD VON MISES (1883-1953) zu nennen.

Diese beiden Ansätze sind extrem konträr. Jeder hat seine eigenen Vorteile. Die zugrunde liegende mathematische Theorie ist aber von der Interpretation unabhängig.

## 1.2 Ereignisse als Mengen

**Definition 1.2.1** (Grundraum). Die Menge aller möglichen Ergebnisse eines Zufallsexperiments heißt *Grundraum* (engl.: *sample space*) und wird üblicherweise mit  $\Omega$  bezeichnet.

**Beispiel 1.2.2** (Werfen einer Münze). Beim Werfen einer Münze gibt es zwei mögliche Ergebnisse: Kopf ( $K$ ) oder Zahl ( $Z$ ). Somit ergibt sich der Grundraum zu  $\Omega = \{K, Z\}$ . Von Interesse könnten die Wahrscheinlichkeiten folgender Ereignisse sein:

- (a) Das Ergebnis ist Kopf.

---

<sup>1</sup>vergleiche Gesetz der Großen Zahlen, Kapitel 2.2

- (b) Das Ergebnis ist Kopf oder Zahl.
- (c) Das Ergebnis ist Kopf und Zahl.
- (d) Das Ergebnis ist nicht Kopf.

**Beispiel 1.2.3** (Einmaliger Würfelwurf). Es ist leicht ersichtlich, dass  $\Omega = \{1, 2, 3, 4, 5, 6\}$  ist. Von Interesse könnten folgende Ereignisse sein:

- (a) Das Ergebnis ist die Zahl 1.
- (b) Das Ergebnis ist eine gerade Zahl.
- (c) Das Ergebnis ist gerade, aber nicht größer 3.
- (d) Das Ergebnis ist nicht gerade.

Ereignisse sind Teilmengen  $A$  von den jeweiligen Grundräumen  $\Omega$ . Für das Beispiel 1.2.2 ergeben sich folgende Lösungen:

- (a)  $A = \{K\}$  oder, in anderer Schreibweise,  $A = K$ .
- (b)  $A = \{K\} \cup \{Z\} = \{K, Z\} = \Omega$ .
- (c)  $A = \{K\} \cap \{Z\} = \emptyset$ .
- (d)  $A = \{K\}^c = \overline{\{K\}}$  oder, in anderer Schreibweise,  $A = K^c = \bar{K}$ .

Für das Beispiel 1.2.3 folgt analog:

- (a)  $A = \{1\}$ .
- (b)  $A = \{2, 4, 6\}$ .
- (c)  $A = \{2, 4, 6\} \cap \{4, 5, 6\}^c = \{2, 4, 6\} \cap \{1, 2, 3\} = \{2\}$ .
- (d)  $A = \{2, 4, 6\}^c = \{1, 3, 5\}$ .

**Definition 1.2.4** (Disjunkte Ereignisse). Zwei Ereignisse  $A, B$  heißen *disjunkt*, wenn ihre Schnittmenge gleich der leeren Menge ist:

$$A \cap B = \emptyset.$$

Außerdem gelten folgende **Bezeichnungen**:

- $\emptyset$  heißt das *unmögliche Ereignis*.
- $\Omega$  heißt das *sichere Ereignis*.
- $A^c$  heißt *Komplement* oder *Gegenereignis* von  $A$ .
- $A \setminus B$  heißt *Differenz* ("A, aber nicht B", "A ohne B").
- $A \triangle B$  heißt *symmetrische Differenz* ("entweder A oder B, aber nicht beide").

**Beachte:** Nicht alle Teilmengen von  $\Omega$  sind notwendigerweise Ereignisse. Bei überabzählbarem  $\Omega$  wäre es in diesem Fall nicht möglich, Wahrscheinlichkeiten auf allen Ereignissen vernünftig zu definieren<sup>2</sup>. Man beschränkt sich daher auf eine Mengenfamilie  $\mathcal{F}$  von Ereignissen, also auf eine Menge von Teilmengen von  $\Omega$ , die "hinreichend reichhaltig" ist in dem Sinne, dass alle interessierenden Ereignisse in  $\mathcal{F}$  enthalten sind und wir wie oben mit Mengenoperationen arbeiten können. Das bedeutet, wir fordern, dass mit zwei Ereignissen  $A$  und  $B$  auch deren Vereinigung  $A \cup B$  und Durchschnitt  $A \cap B$  sowie das Gegenereignis  $A^c = \Omega \setminus A$  in  $\mathcal{F}$  enthalten sind, außerdem das sichere und unmögliche Ereignis. Sind diese Eigenschaften erfüllt, nennt man  $\mathcal{F}$  eine *Algebra*. Man folgert leicht, dass für  $A_1, A_2, \dots, A_n \in \mathcal{F}$  auch  $\bigcup_{i=1}^n A_i \in \mathcal{F}$  und  $\bigcap_{i=1}^n A_i \in \mathcal{F}$ . Dies gilt jedoch nur für *endliche* Vereinigungen und Durchschnitte! In einigen Fällen, wie dem folgenden Beispiel, ist dies noch nicht ausreichend. Daher verallgemeinert man den Begriff der Algebra zu dem der  $\sigma$ -Algebra.

**Beispiel 1.2.5.** Eine Münze wird solange geworfen, bis zum ersten Mal Kopf erscheint, also  $\Omega = \{\omega_1, \omega_2, \dots\}$ , wobei  $\omega_i$  das Ergebnis "beim  $i$ -ten Versuch Kopf, alle vorhergehenden Versuche Zahl" bezeichnet. Man könnte sich zum Beispiel für das Ereignis  $A$  "Kopf erscheint nach ungerader Anzahl von Zahl-Würfen",  $A = \{\omega_2, \omega_4, \omega_6, \dots\}$ , interessieren.

**Definition 1.2.6** ( $\sigma$ -Algebra). Besitzt eine Mengenfamilie  $\mathcal{F}$  von Teilmengen von  $\Omega$  die Eigenschaften

- (a)  $\emptyset \in \mathcal{F}$ ,
- (b)  $A \in \mathcal{F} \implies A^c \in \mathcal{F}$ ,
- (c)  $A_1, A_2, \dots \in \mathcal{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ ,

dann nennt man  $\mathcal{F}$  eine  $\sigma$ -Algebra (engl.:  $\sigma$ -field) über  $\Omega$ .

**Bemerkung:** Man zeigt leicht, dass dann auch  $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$ .

**Beispiel 1.2.7** (Kleinste  $\sigma$ -Algebra). Die kleinste  $\sigma$ -Algebra über einem Grundraum  $\Omega$  ist  $\mathcal{F} = \{\emptyset, \Omega\}$ . Sie heißt auch *triviale  $\sigma$ -Algebra*.

**Beispiel 1.2.8.** Falls  $A \subset \Omega$  ist, so ist  $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$  eine  $\sigma$ -Algebra über  $\Omega$ .

**Beispiel 1.2.9** (Potenzmenge). Die *Potenzmenge* von  $\Omega$  (Schreibweise:  $\{0, 1\}^{\Omega}$  oder  $\mathcal{P}(\Omega)$ ) ist definiert als die Menge aller Teilmengen von  $\Omega$  und ist eine  $\sigma$ -Algebra. Ist  $\Omega$  überabzählbar, so ist die Potenzmenge "zu groß", um auf jedem Element der Potenzmenge Wahrscheinlichkeiten zu definieren. Bei abzählbarem  $\Omega$  dagegen kann man immer  $\mathcal{P}(\Omega)$  als  $\sigma$ -Algebra wählen.

---

<sup>2</sup>siehe auch das Buch von Krenzel

**Satz 1.2.10** (Gesetze von DE MORGAN, 1806–1871). Für eine Mengenfamilie  $\{A_i : i \in I\}$  gilt:

$$\begin{aligned}\left(\bigcup_{i \in I} A_i\right)^c &= \left(\bigcap_{i \in I} A_i^c\right), \\ \left(\bigcap_{i \in I} A_i\right)^c &= \left(\bigcup_{i \in I} A_i^c\right).\end{aligned}$$

Speziell für zwei Mengen gilt:

$$(A \cup B)^c = A^c \cap B^c \quad \text{und} \quad (A \cap B)^c = A^c \cup B^c.$$

*Beweis.* Für die erste Gleichung zeigt man sowohl " $\subseteq$ " als auch " $\supseteq$ ". Die zweite Gleichung ergibt sich durch Anwendung der ersten auf  $A_i^c$ . Der komplette Beweis wird auf dem ersten Übungsblatt gezeigt.  $\square$

### 1.3 Wahrscheinlichkeit

**Ziel:** Erweiterung des Maßraums  $(\Omega, \mathcal{F})$  zum Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$ .

**Definition 1.3.1** (Axiome von KOLMOGOROV, 1903-1987). Ein *Wahrscheinlichkeitsmaß* auf  $(\Omega, \mathcal{F})$  ( $\Omega \neq \emptyset$ ,  $\mathcal{F}$  ist  $\sigma$ -Algebra über  $\Omega$ ) ist eine Funktion  $P : \mathcal{F} \rightarrow [0, 1]$  mit den Eigenschaften

- (a)  $P(\emptyset) = 0$  ("Positivität"),
- (b)  $P(\Omega) = 1$  ("Normiertheit"),
- (c) Für  $A_1, A_2, \dots$  paarweise disjunkt (d.h.  $A_i \cap A_j = \emptyset$  für  $i \neq j$ ) gilt

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad (\text{"}\sigma\text{-Additivität"})$$

$(\Omega, \mathcal{F}, P)$  heißt *Wahrscheinlichkeitsraum*.

**Beispiel 1.3.2** (Münzwurf). Es ist  $\Omega = \{K, Z\}$  und somit  $\mathcal{F} = \{\emptyset, K, Z, \Omega\}$ . Ein mögliches Wahrscheinlichkeitsmaß ist nach Definition 1.3.1 durch

$$\begin{aligned}P(\emptyset) &= 0 \\ P(\Omega) &= 1 \\ P(K) &= p \\ P(Z) &= 1 - p\end{aligned}$$

gegeben, wobei  $p \in (0, 1)$ .

**Beispiel 1.3.3** (Würfelwurf). Es ist  $\Omega = \{1, 2, 3, 4, 5, 6\}$  und  $\mathcal{F} = \{0, 1\}^\Omega = \mathcal{P}(\Omega)$ . Ein Wahrscheinlichkeitsmaß  $P$  ist, Definition 1.3.1 folgend, durch

$$\begin{aligned}P(\emptyset) &= 0 \\ P(\Omega) &= 1 \\ P(A) &= \sum_{i \in A} p_i \quad \forall A \subseteq \mathcal{F}\end{aligned}$$

gegeben, wobei  $p_i$  Zahlen im Intervall  $[0, 1]$  mit  $\sum_{i=1}^6 p_i = 1$  sind.

**Lemma 1.3.4.** In einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$  gilt:

(a)  $P(A^c) = 1 - P(A)$  für alle  $A \in \mathcal{F}$ .

(b)  $B \supseteq A \Rightarrow P(B) \geq P(A)$  ("Monotonie").

(c)  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  für alle  $A, B \in \mathcal{F}$ .

(d) (**Siebformel von SYLVESTER, 1814-1897, und POINCARÉ, 1854-1912**) Seien  $A_1, A_2, \dots, A_n$  Ereignisse, dann gilt:

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n).$$

*Beweis.*

(a) Es ist  $A \cup A^c = \Omega$  und  $A \cap A^c = \emptyset$  ( $A$  und  $A^c$  sind disjunkt), daher:

$$P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1.$$

(b) Aus  $B \supseteq A$  folgt  $B = A \cup (B \setminus A)$ . Da  $A$  und  $B \setminus A$  disjunkt, gilt

$$P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A) \geq P(A).$$

(c) Da  $A \cup B = A \cup (B \setminus (A \cap B))$  (disjunkte Vereinigung!), erhält man

$$P(A \cup B) = P(A) + P(B \setminus (A \cap B)).$$

Im Beweis von Teil (b) hatten wir  $P(D \setminus C) = P(D) - P(C)$  für zwei Ereignisse  $C$  und  $D$  gesehen, also

$$P(B \setminus (A \cap B)) = P(B) - P(A \cap B),$$

und damit erhält man die gewünschte Formel.

(d) Siehe Übungsblatt 2. □

**Lemma 1.3.5** (Stetigkeit des Wahrscheinlichkeitsmaßes  $P$ ). Sei  $A_1, A_2, \dots$  eine wachsende (auch: isotone) Folge von Ereignissen, das heißt  $A_1 \subseteq A_2 \subseteq \dots$ . Betrachte den Grenzwert

$$A := \bigcup_{i=1}^{\infty} A_i = \lim_{i \rightarrow \infty} A_i.$$

Dann gilt:

$$P(A) = \lim_{i \rightarrow \infty} P(A_i).$$

*Beweis.* Man überlegt sich leicht, dass  $A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \dots$ . Also lässt sich  $A$  darstellen als Vereinigung von einer disjunkten Mengenfamilie von Ereignissen. Daher folgt mit Definition 1.3.1.c:

$$\begin{aligned} P(A) &= P(A_1) + \sum_{i=1}^{\infty} P(A_{i+1} \setminus A_i) \\ &= P(A_1) + \lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} (P(A_{i+1}) - P(A_i)) \\ &= \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

□

**Bemerkung:** Eine ähnliche Aussage gilt für fallende (auch: *antitone*) Mengenfolgen im Sinne von  $B_1 \supseteq B_2 \supseteq B_3 \supseteq \dots$ , wobei  $B := \bigcap_{i=1}^{\infty} B_i$ . Es gilt:

$$P(B) = \lim_{i \rightarrow \infty} P(B_i).$$

Abschließend seien noch zwei Definitionen gegeben:

**Definition 1.3.6** (Nullereignis). Ein Ereignis  $A$  heißt *null*, wenn  $P(A) = 0$ . Daraus folgt *nicht*  $A = \emptyset$ . Man spricht auch von einem *Nullereignis*.

**Definition 1.3.7** (fast sicher). Ein Ereignis  $A$  heißt *fast sicher*, wenn  $P(A) = 1$ . Daraus folgt *nicht*  $A = \Omega$ , sondern lediglich, dass  $A^c$  ein Nullereignis ist.

**Aufgabe:** Seien  $A, B$  Ereignisse mit  $P(A) = 3/4$  und  $P(B) = 1/3$ . Zeigen Sie, dass  $1/12 \leq P(A \cap B) \leq 1/3$ .

**Lösung:** Die obere Grenze ergibt sich durch die Vorstellung, dass  $B$  ganz in  $A$  liegt, die untere aus

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \leq 1 \\ \implies P(A \cap B) &\geq P(A) + P(B) - 1 = \frac{3}{4} + \frac{1}{3} - 1 = \frac{1}{12}. \end{aligned}$$

## 1.4 Bedingte Wahrscheinlichkeit

Vorab eine heuristische Betrachtung: Angenommen, man weiß, dass ein Ereignis  $B$  bereits eingetreten ist, dann ist die Wahrscheinlichkeit von  $A$  gegeben diese Information offensichtlich proportional zu  $P(A \cap B)$ , das heißt gleich  $\alpha \cdot P(A \cap B)$  mit  $\alpha \in \mathbb{R}$ . Aus  $\alpha \cdot P(\Omega \cap B) = 1$  folgt, dass  $\alpha = 1/P(B)$ , also  $P(A|B) = P(A \cap B)/P(B)$ .

**Definition 1.4.1** (Bedingte Wahrscheinlichkeit). Seien  $A, B$  zwei Ereignisse mit  $P(B) > 0$ . Dann ist die *bedingte Wahrscheinlichkeit von  $A$  gegeben  $B$*  definiert durch

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

**Beispiel 1.4.2.** Ein Ehepaar hat zwei Kinder. Wie groß ist die Wahrscheinlichkeit, dass es zwei Jungen sind, wenn wir bereits wissen, dass eines der Kinder ein Junge ist?

**1. Schritt:** Definiere den Grundraum

$$\Omega = \{JJ, MM, JM, MJ\}.$$

**2. Schritt:** Wahrscheinlichkeiten der Elementarereignisse festlegen mit

$$P(JJ) = P(MM) = P(JM) = P(MJ) = \frac{1}{4}.$$

**3. Schritt:** Definiere die notwendigen Ereignisse:

$$\begin{aligned} A &:= \text{„zwei Jungen“} \\ B &:= \text{„ein Kind ist ein Junge“} \\ A &= \{JJ\} \\ B &= \{JJ, JM, MJ\} \\ A \cap B &= \{JJ\}. \end{aligned}$$

**4. Schritt:** Anwenden von Definition 1.4.1:

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(\{JJ\})}{P(\{JJ, JM, MJ\})} \\ &= \frac{P(\{JJ\})}{P(\{JJ\}) + P(\{JM\}) + P(\{MJ\})} \\ &= \frac{\frac{1}{4}}{3 \cdot \frac{1}{4}} \\ &= \frac{1}{3}. \end{aligned}$$

**Definition 1.4.3** (Partition). Eine Familie  $B_1, B_2, \dots, B_n$  von Ereignissen heißt *Partition* (oder *disjunkte Zerlegung*) von  $\Omega$ , wenn gilt:

1.  $B_i \cap B_j = \emptyset \quad \forall i \neq j$  ("  $B_i$ 's paarweise disjunkt"),
2.  $\bigcup_{i=1}^n B_i = \Omega$ .

**Lemma 1.4.4** (Satz von der totalen Wahrscheinlichkeit). Für beliebige Ereignisse  $A$  und  $B$  mit  $0 < P(B) < 1$  gilt

$$P(A) = P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c).$$

Allgemein gilt: Ist  $B_1, B_2, \dots, B_n$  eine Partition von  $\Omega$  mit  $P(B_i) > 0$  für alle  $i = 1, \dots, n$ , dann gilt

$$P(A) = \sum_{i=1}^n P(A|B_i) \cdot P(B_i).$$

*Beweis.* Die Ereignisse  $(A \cap B)$  und  $(A \cap B^c)$  sind disjunkt, und  $A = (A \cap B) \cup (A \cap B^c)$ . Mit Definition 1.4.1 gilt daher:

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) \\ &= P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c). \end{aligned}$$

Der Beweis der zweiten Aussage verläuft analog. □

**Beispiel 1.4.5.** Gegeben seien zwei Urnen mit folgendem Inhalt:

**Urne I:** 2 weiße, 3 blaue Bälle,

**Urne II:** 3 weiße, 4 blaue Bälle.

Ein zufällig aus Urne I gezogener Ball wird in Urne II gelegt. Dann wird zufällig aus Urne II ein Ball gezogen. Wie groß ist die Wahrscheinlichkeit, dass man einen blauen Ball zieht?

*Lösung:* Definiere die Ereignisse

$$\begin{aligned} A &= \text{"Der zweite gezogene Ball ist blau."} \\ B &= \text{"Der erste gezogene Ball ist blau."} \end{aligned}$$

Somit ergibt sich nach dem Satz der totalen Wahrscheinlichkeit (Lemma 1.4.4) folgendes Ergebnis:

$$\begin{aligned} P(A) &= P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c) \\ &= \frac{5}{8} \cdot \frac{3}{5} + \frac{4}{8} \cdot \frac{2}{5} \\ &= \frac{23}{40}. \end{aligned}$$

**Beispiel 1.4.6** (Prisoners' paradox). Drei Gefangene sind zum Tode verurteilt. Einer wird zufällig ausgewählt und begnadigt. Gefangener  $a$  weiß daher, dass die Wahrscheinlichkeit zu überleben gleich  $1/3$  ist. Nun bittet er den Wärter, ihm den Namen von einem der beiden anderen Gefangenen zu nennen, der nicht begnadigt wird. Der Wärter sagt: " $b$  wird nicht begnadigt". Der Gefangene  $a$  argumentiert nun: Nur er selbst oder  $c$  werden nun begnadigt, daher steigt die Wahrscheinlichkeit, dass er überlebt, auf  $1/2$ . Hat er recht?

*Lösung:* Definiere folgende Ereignisse:

$$\begin{aligned} A &:= \text{"}a \text{ wird begnadigt"} \\ B &:= \text{"}b \text{ wird begnadigt"} \\ C &:= \text{"}c \text{ wird begnadigt"} \\ W_b &:= \text{"Wärter sagt, } b \text{ wird nicht begnadigt"} \\ W_c &:= \text{"Wärter sagt, } c \text{ wird nicht begnadigt"} \end{aligned}$$

Gesucht ist  $P(A|W_b)$ . Als Hilfsmittel wird der Satz der totalen Wahrscheinlichkeit herangezogen. Hieraus ergibt sich eine formale Zerlegung von  $\Omega$  in  $A \cap W_b, B \cap W_b, C \cap W_b, A \cap W_c, B \cap W_c, C \cap W_c$ . Nun stellt sich die Frage nach der Festlegung der Wahrscheinlichkeiten auf  $\Omega$ .

1.  $P(A) = P(B) = P(C) = \frac{1}{3}$

2.

$$\begin{aligned}P(W_b|B) &= 0 \\P(W_c|B) &= 1 \\P(W_b|C) &= 1 \\P(W_c|C) &= 0 \\P(W_b|A) &= \beta \geq 0 \\P(W_c|A) &= \gamma \geq 0\end{aligned}$$

mit  $\beta + \gamma = 1$ .

Mit dem Satz der totalen Wahrscheinlichkeit ergeben sich folgende Wahrscheinlichkeiten auf  $\Omega$ :

$\cap$	$A$	$B$	$C$	
$W_b$	$\frac{\beta}{3}$	0	$\frac{1}{3}$	$\frac{1+\beta}{3}$
$W_c$	$\frac{\gamma}{3}$	$\frac{1}{3}$	0	$\frac{1+\gamma}{3}$
	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	

Mit diesen Informationen erhält man

$$P(A|W_b) = \frac{P(A \cap W_b)}{P(W_b)} = \frac{\frac{\beta}{3}}{\frac{1+\beta}{3}} = \frac{\beta}{1+\beta}.$$

Da  $0 \leq \beta \leq 1$ , folgt nur  $0 \leq P(A|W_b) \leq 1/2$ . Nimmt man gleiche Wahrscheinlichkeiten für jedes der beiden Ereignisse  $P(W_b|A) = \beta$  und  $P(W_c|A) = \gamma$  an, also  $\beta = \gamma = 1/2$ , erhält man  $P(A|W_b) = 1/3$ .

Im Folgenden wird ein weiterer Satz der Statistik hergeleitet: Da  $A \cap B = B \cap A$ , folgt aus  $P(A|B) = P(A \cap B)/P(B)$  (für  $P(B) > 0$ ) und  $P(B|A) = P(B \cap A)/P(A)$  (für  $P(A) > 0$ ) sofort der *Satz von BAYES*.

**Satz 1.4.7** (Satz von Bayes, 1701–1761). *Seien  $A$  und  $B$  zwei Ereignisse mit  $0 < P(A) < 1$  und  $P(B) > 0$ . Dann gilt:*

$$\begin{aligned}P(A|B) &= \frac{P(B|A) \cdot P(A)}{P(B)} \\&= \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)}.\end{aligned}$$

*Allgemein: Sei  $B$  ein Ereignis mit  $P(B) > 0$  und  $A_1, A_2, \dots, A_n$  eine Partition von  $\Omega$  mit  $P(A_i) > 0$  für alle  $i = 1, \dots, n$ . Dann gilt:*

$$P(A_j|B) = \frac{P(B|A_j) \cdot P(A_j)}{\sum_{i=1}^n P(B|A_i) \cdot P(A_i)}.$$

Als Standardbeispiel für den Satz von Bayes kann ein diagnostischer Test genannt werden. Eine weitere Bedeutung wird durch folgende Umformulierung ersichtlich:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad \text{und} \quad P(A^c|B) = \frac{P(B|A^c) \cdot P(A^c)}{P(B)}.$$

Division der ersten durch die zweite Gleichung liefert den folgenden Satz.

**Satz 1.4.8** (Variante des Satzes von Bayes). *Mit  $A$  und  $B$  wie in Satz 1.4.7 gilt:*

$$\frac{P(A|B)}{P(A^c|B)} = \frac{P(A)}{P(A^c)} \cdot \frac{P(B|A)}{P(B|A^c)}.$$

$$\textit{Posteriori Chance} = \textit{Priori Chance} \cdot \textit{Likelihood Quotient}$$

Zum einen ist die Angabe von Chancen in vielen Bereichen (wie etwa der Medizin) sehr beliebt, zum anderen bietet diese Schreibweise ein Tor zur Likelihood-Inferenz.

Wir bemerken an dieser Stelle, dass zwischen Wahrscheinlichkeiten  $\pi$  und den zugehörigen Chancen  $\gamma$  offensichtlich folgende Zusammenhänge bestehen:

$$\gamma = \frac{\pi}{1 - \pi} \tag{1}$$

$$\pi = \frac{\gamma}{1 + \gamma} \tag{2}$$

**Beispiel 1.4.9** (Fall-Kontroll-Studie). Die zwei Ereignisse  $E$  (*Exposition*) und  $K$  (*Krankheit*) stehen dafür, dass eine Person einem Risiko ausgesetzt bzw. an einer bestimmten Krankheit leidet. Zum Beispiel könnte  $E$  das Ereignis sein, dass eine Person raucht, und  $K$ , dass sie an Lungenkrebs erkrankt ist. Man interessiert sich für den Zusammenhang zwischen diesen beiden Ereignissen. In Fall-Studien schätzt man  $P(E|K)$  und  $P(E|K^c)$  und berechnet daraus das *Expositions-Odds Ratio* (engl. *exposure odds ratio*)

$$\frac{P(E|K)}{P(E^c|K)} \bigg/ \frac{P(E|K^c)}{P(E^c|K^c)}.$$

Von eigentlichem Interesse ist aber das *Krankheits-Odds Ratio* (engl. *disease odds ratio*)

$$\frac{P(K|E)}{P(K^c|E)} \bigg/ \frac{P(K|E^c)}{P(K^c|E^c)}.$$

Die beiden Odds Ratios sind gleich, da nach Satz 1.4.8

$$\frac{P(E|K)}{P(E^c|K)} = \frac{P(K|E)}{P(K^c|E)} \cdot \frac{P(E)}{P(E^c)} \tag{3}$$

und

$$\frac{P(E|K^c)}{P(E^c|K^c)} = \frac{P(K^c|E)}{P(K^c|E^c)} \cdot \frac{P(E)}{P(E^c)}. \tag{4}$$

Dividieren von (3) durch (4) liefert

$$\frac{P(E|K)}{P(E^c|K)} \bigg/ \frac{P(E|K^c)}{P(E^c|K^c)} = \frac{P(K|E)}{P(K^c|E)} \bigg/ \frac{P(K|E^c)}{P(K^c|E^c)}.$$

Da im Allgemeinen sehr seltene Krankheiten betrachtet werden, kann man  $P(K^c|E) \approx 1$  und  $P(K^c|E^c) \approx 1$  annehmen. In diesem Fall vereinfacht sich das Krankheits-Odds-Ratio zum *relativen Risiko*  $P(K|E)/P(K|E^c)$ .

Bevor nun ein Beispiel zum Satz von Bayes folgt, sei noch einiges angemerkt.

**Bemerkung** (Rechenregeln): Bedingte Wahrscheinlichkeiten  $P(A|B)$  verhalten sich (für  $P(B) > 0$ ) wie gewöhnliche Wahrscheinlichkeiten. Dies folgt, wie man leicht sieht, aus der Definition von bedingter Wahrscheinlichkeit,  $P(A|B) = P(A \cap B)/P(B)$ . Zum Beispiel kann man die Siebformel wie folgt verallgemeinern:

$$P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C) \quad \text{für } P(C) > 0.$$

Andere Formeln und Rechenregeln gelten analog. Man "schleppt" das bedingende Ereignis und dessen Wahrscheinlichkeit einfach mit. Für weitere Umformungen bieten sich vor allem der Satz der totalen Wahrscheinlichkeit (Lemma 1.4.4) sowie die Definition der bedingten Wahrscheinlichkeit (Definition 1.4.1) an.

Im Weiteren gilt folgende **Schreibweise**:  $P(A|B, C) := P(A|B \cap C)$ .

Durch Anwendung der Rechenregeln lässt sich der Satz von Bayes noch in folgender Variante darstellen:

**Satz 1.4.10** (Satz von Bayes bei zwei bedingten Ereignissen). *Unter der Voraussetzung, dass die beteiligten Wahrscheinlichkeiten definiert sind, gilt:*

$$\frac{P(A|B, C)}{P(A^c|B, C)} = \frac{P(A|C)}{P(A^c|C)} \cdot \frac{P(B|A, C)}{P(B|A^c, C)} \quad (5)$$

oder auch

$$\frac{P(A|B, C)}{P(A^c|B, C)} = \frac{P(A|B)}{P(A^c|B)} \cdot \frac{P(C|A, B)}{P(C|A^c, B)}. \quad (6)$$

*Beweis.* Die Gültigkeit von (5) folgt direkt, indem alle in Satz 1.4.8 auftauchenden Wahrscheinlichkeiten noch zusätzlich bzgl.  $C$  bedingt werden. Variante (6) ist zu (5) äquivalent.  $\square$

**Beispiel 1.4.11** (O. J. SIMPSON PROZESS). Der Amerikaner O.J. Simpson, berühmter Footballspieler und später auch Schauspieler, wurde 1994 wegen Mordes an seiner Ex-Frau und deren Liebhaber angeklagt. Im Prozess wurde Simpson vorgeworfen, seine Frau früher geschlagen und vergewaltigt zu haben. Simpsons Verteidiger, Alan Dershowitz, wies diese Vorwürfe als irrelevant zurück, da nur jeder 1000. Mann, der seine Frau schlägt, sie schließlich auch umbringen würde. Der emeritierte Statistikprofessor I.J. Good hielt dagegen, dass es hier nicht um die Wahrscheinlichkeit gehe, dass ein Mann seine Frau umbringe, wenn er sie zuvor geschlagen habe. Gesucht sei vielmehr die Wahrscheinlichkeit, dass ein Mann seine Frau umgebracht hat, wenn er sie zuvor geschlagen hat und wenn diese Frau dann tatsächlich von jemandem umgebracht worden ist. Auf der Grundlage eigener Schätzungen und von der Verteidigung gelieferter Zahlen berechnete Good diese Wahrscheinlichkeit als keineswegs verschwindend gering. Er schickte seinen Artikel sowohl an die Zeitschrift *Nature* als auch Simpsons Verteidigung und die Polizei von Los Angeles. Es ist jedoch davon auszugehen, dass nicht alle Empfänger die wahrscheinlichkeitstheoretischen Überlegungen gleichermaßen verstanden haben.

Simpson wurde im Strafprozess von den Geschworenen freigesprochen, von einem Zivilgericht jedoch zu Schadensersatzzahlungen an die Hinterbliebenen der Opfer verurteilt.

Im Folgenden wollen wir uns mit den Berechnungen eines Artikels von J.F. Merz und J.P. Caulkins<sup>3</sup> und denen von Good<sup>4</sup> befassen. Dazu werden drei Ereignisse definiert:

$A$	:	"Mann hat seine Frau geschlagen"	[Abuse]
$M$	:	"Frau wurde umgebracht"	[Murder]
$G$	:	"Mann hat seine Frau umgebracht"	[Guilty]

Folgende bedingte Wahrscheinlichkeiten werden von Merz & Caulkins verwendet:

$$\begin{aligned} P(G|M) &= \frac{1430}{4936} = 0.29 \\ \Rightarrow P(G^c|M) &= 0.71 \\ P(A|G, M) &= 0.5 \\ P(A|G^c, M) &= 0.05. \end{aligned}$$

Diese Wahrscheinlichkeiten basieren auf folgenden empirischen Zahlen: Von den 4936 Frauen, die 1992 ermordet wurden, wurden 1430 von Ihren Ehemännern ermordet, daher  $P(G|M) = 0.29$ . In einem Zeitungsartikel zitieren Sie Dershowitz wie folgt: "It is, of course, true that, among the small number of men who do kill their present or former mates, a considerable number did first assault them." Merz & Caulkins interpretieren "a considerable number" als 50%, so dass  $P(A|G, M) = 0.5$  folgt. Schließlich nehmen Sie an, dass  $P(A|G^c, M)$  gleich der Wahrscheinlichkeit ist, dass eine zufällige ausgewählte Frau geschlagen wird. Empirische Daten schätzen den Anteil von Frauen, die geschlagen werden, auf 5%, also  $P(A|G^c, M) = 0.05$ .

Unter Anwendung von Satz 1.4.10 ergibt sich

$$\begin{aligned} \frac{P(G|A, M)}{P(G^c|A, M)} &= \frac{P(G|M)}{P(G^c|M)} \cdot \frac{P(A|G, M)}{P(A|G^c, M)} \\ &= \frac{0.29}{0.71} \cdot \frac{0.5}{0.05} \\ &= 4.08. \end{aligned}$$

Somit folgt mit (2)  $P(G|A, M) = 4.08/(1 + 4.08) \approx 0.8$ . Das heißt mit einer Wahrscheinlichkeit von 0.8 hat ein Mann seine Frau umgebracht, wenn er sie zuvor geschlagen hat und wenn diese Frau dann tatsächlich von jemandem umgebracht worden ist.

Überraschend ist das Ergebnis, wenn man die Methode von Good verwendet, die ebenfalls auf Satz 1.4.10 beruht: Er verwendet

$$\frac{P(G|A, M)}{P(G^c|A, M)} = \frac{P(G|A)}{P(G^c|A)} \cdot \frac{P(M|G, A)}{P(M|G^c, A)} \quad (7)$$

und schätzt  $P(G|A) = 1/10000$ . Hintergrund dieser Schätzung ist die Aussage von Dershowitz, dass nur jeder 1000. Mann, der seine Frau schlägt, sie schließlich auch umbringen würde. Good nimmt also an, dass das (mindestens) mit Wahrscheinlichkeit 1/10 in dem fraglichen Jahr passieren wird. Da  $P(M|G, A) = 1$ , bleibt nur noch  $P(M|G^c, A)$  zu quantifizieren. Offensichtlich gilt  $P(M|G^c, A) = P(M|G^c) \approx P(M)$ . Da es jährlich in den Vereinigten Staaten ca. 25,000 Morde gibt, folgt bei 250,000,000 Einwohnern, dass

$$P(M|G^c, A) = \frac{25000}{250000000} = \frac{1}{10000}.$$

<sup>3</sup>J.F. Merz and J.P. Caulkins (1995): "Propensity to abuse—propensity to murder?", *Chance*, 8(2), 14.

<sup>4</sup>I.J. Good (1995): "When batterer turns murderer", *Nature*, 375(6532), 541.

Setzt man diese Zahlen in (7) ein, so erhält man

$$\frac{P(G|A, M)}{P(G^c|A, M)} = \frac{\frac{1}{10000}}{\frac{9999}{10000}} \cdot \frac{1}{\frac{1}{10000}} \approx 1$$

Da  $P(G|A, M) + P(G^c|A, M) = 1$ , erhält man  $P(G|A, M) = 0.5$ .

## 1.5 Unabhängigkeit

**Idee:** Gilt  $P(A|B) = P(A)$ , so nennt man  $A$  und  $B$  unabhängig.

**Beachte:** Für  $P(A) > 0$  und  $P(B) > 0$  gilt im Falle von Unabhängigkeit:

$$P(A|B) = P(A) \iff P(B|A) = P(B).$$

Denn:

$$\begin{aligned} P(A) &= P(A|B) = P(A \cap B)/P(B) \\ \Leftrightarrow P(B) &= P(A \cap B)/P(A) = P(B|A). \end{aligned}$$

In folgender Definition kann man die Bedingung  $P(A) > 0$  bzw.  $P(B) > 0$  sogar weglassen.

**Definition 1.5.1** (Unabhängigkeit von Ereignissen). Zwei Ereignisse  $A$  und  $B$  heißen (*stochastisch*) *unabhängig*, wenn

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

gilt.

**Bemerkung:** Wenn  $A$  und  $B$  unabhängig, dann auch

$$\begin{aligned} A &\text{ und } B^c, \\ A^c &\text{ und } B, \\ A^c &\text{ und } B^c. \end{aligned}$$

In Worten: Wenn zwei Ereignisse voneinander unabhängig sind, sind sie auch jeweils von den Komplementen des anderen Ereignisses unabhängig.

Verallgemeinert gilt folgende Definition:

**Definition 1.5.2** (Unabhängigkeit einer Familie). Man nennt eine Familie  $\{A_i : i \in I\}$  von Ereignissen (*stochastisch*) *unabhängig*, falls

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

für alle endlichen Teilmengen  $J \subset I$  gilt.

**Bemerkung:** Falls für eine Familie  $\{A_i : i \in I\}$

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$$

für alle  $i \neq j$  gilt, so nennt man diese *paarweise unabhängig*. Daraus folgt *nicht* notwendigerweise die Unabhängigkeit der Familie.

**Beispiel 1.5.3.** Sei  $\Omega = \{abc, acb, cab, cba, bca, bac, aaa, bbb, ccc\}$  mit  $P(\omega_i) = 1/9$  für  $i = 1, \dots, 9$ . Sei  $A_k$  das Ereignis "k-ter Buchstabe ist ein a" ( $k = 1, 2, 3$ ). Dann ist

$$\begin{aligned} A_1 &= \{abc, acb, aaa\} \\ A_2 &= \{cab, bac, aaa\} \\ A_3 &= \{cba, bca, aaa\} \end{aligned}$$

und  $P(A_k) = 1/3$  für  $k = 1, 2, 3$ . Es gilt

$$A_1 \cap A_2 = A_1 \cap A_3 = A_2 \cap A_3 = \{aaa\}$$

und daher

$$P(A_i \cap A_j) = \frac{1}{9} = \frac{1}{3} \cdot \frac{1}{3} = P(A_i) \cdot P(A_j) \quad \forall i \neq j.$$

Das heißt, die Ereignisse sind paarweise unabhängig. *Aber:*

$$P(A_1 \cap A_2 \cap A_3) = P(\{aaa\}) = \frac{1}{9} \neq \left(\frac{1}{3}\right)^3 = P(A_1) \cdot P(A_2) \cdot P(A_3).$$

Das heißt, die Ereignisse sind *nicht* unabhängig. Oder genauer: Die Familie der Ereignisse ist nicht unabhängig.

Der Sachverhalt, dass aus paarweiser Unabhängigkeit nicht die Unabhängigkeit von Familien von Ereignissen folgt, sei hier nochmals betont!

**Definition 1.5.4** (Bedingte Unabhängigkeit). Sei  $C$  ein Ereignis mit  $P(C) > 0$ . Man nennt die Ereignisse  $A$  und  $B$  *bedingt unabhängig gegeben  $C$* , falls

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

Eine analoge Definition gilt für Familien  $\{A_i : i \in I\}$  von Ereignissen.

**Beachte:** Aus der Unabhängigkeit folgt im Allgemeinen *nicht* die bedingte Unabhängigkeit, genauso folgt aus der bedingten Unabhängigkeit im Allgemeinen nicht die Unabhängigkeit.

## 1.6 Vollständigkeit und Produkträume

**Lemma 1.6.1.** *Seien  $\mathcal{F}$  und  $\mathcal{G}$  zwei  $\sigma$ -Algebren über  $\Omega$ . Dann ist auch  $\mathcal{F} \cap \mathcal{G}$  eine  $\sigma$ -Algebra über  $\Omega$ .*

Allgemeiner: *Ist  $\{\mathcal{F}_i : i \in I\}$  eine Familie von  $\sigma$ -Algebren über  $\Omega$ , dann ist auch*

$$\mathcal{G} := \bigcap_{i \in I} \mathcal{F}_i$$

*eine  $\sigma$ -Algebra über  $\Omega$ .*

*Beweis.* Siehe Übungsblatt 3. □

**Bemerkung:**  $\mathcal{F} \cup \mathcal{G}$  ist im Allgemeinen *keine*  $\sigma$ -Algebra. Man kann  $\mathcal{F} \cup \mathcal{G}$  jedoch erweitern zu  $\sigma(\mathcal{F} \cup \mathcal{G})$ , der *von  $\mathcal{F} \cup \mathcal{G}$  erzeugten  $\sigma$ -Algebra*. Diese ist die kleinste  $\sigma$ -Algebra, die  $\mathcal{F} \cup \mathcal{G}$  erfasst, und sie ist eindeutig.

Konstruktion: Betrachte  $\{\mathcal{H}_i : i \in I\}$ , die Familie aller  $\sigma$ -Algebren, die sowohl  $\mathcal{F}$  als auch  $\mathcal{G}$  beinhalten. Diese Familie ist nicht leer, da auf jeden Fall die Potenzmenge  $\mathcal{P}(\Omega)$  von  $\Omega$  darin enthalten sein muss. Dann ist  $\mathcal{H} = \bigcap_{i \in I} \mathcal{H}_i$  die eindeutige kleinste  $\sigma$ -Algebra, die  $\mathcal{F} \cup \mathcal{G}$  enthält.  $\mathcal{F} \cup \mathcal{G}$  heißt dann auch *Erzeugendensystem* von  $\sigma(\mathcal{F} \cup \mathcal{G})$ .

### 1.6.1 Vollständigkeit

**Definition 1.6.2** (Vollständigkeit). Ein Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$  heißt *vollständig*, falls alle Teilmengen von Nullereignissen wieder Ereignisse sind.

**Bemerkung:** Man kann jeden unvollständigen Wahrscheinlichkeitsraum wie folgt vervollständigen: Sei  $\mathcal{N}$  die Menge aller Teilmengen von Nullereignissen in  $\mathcal{F}$ . Dann ist  $\mathcal{G} = \sigma(\mathcal{F} \cup \mathcal{N})$  wieder eine  $\sigma$ -Algebra über  $\Omega$ . Das Wahrscheinlichkeitsmaß  $P$  kann in naheliegender Weise von  $\mathcal{F}$  auf  $\mathcal{G}$  verallgemeinert werden.  $(\Omega, \mathcal{G}, P)$  ist dann vollständig.

### 1.6.2 Produkträume

Betrachte zwei Wahrscheinlichkeitsräume  $(\Omega_1, \mathcal{F}_1, P_1)$  und  $(\Omega_2, \mathcal{F}_2, P_2)$ . Wie kann man diese kombinieren?

1.  $\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1 \text{ und } \omega_2 \in \Omega_2\}$ .
2. **Problem:**  $\mathcal{F}_1 \times \mathcal{F}_2 = \{A_1 \times A_2 : A_1 \in \mathcal{F}_1 \text{ und } A_2 \in \mathcal{F}_2\}$  ist nicht notwendigerweise wieder eine  $\sigma$ -Algebra.  
**Lösung:** Verwende  $\mathcal{G} = \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$ .
3. Definiere  $P_{12} : \mathcal{F}_1 \times \mathcal{F}_2 \rightarrow [0, 1]$  via  $P_{12}(A_1 \times A_2) = P_1(A_1) \cdot P_2(A_2)$  für alle  $A_1 \in \mathcal{F}_1$  und  $A_2 \in \mathcal{F}_2$ .  $P_{12}$  lässt sich auf  $\mathcal{G}$  erweitern.

Man nennt  $(\Omega_1 \times \Omega_2, \mathcal{G}, P_{12})$  den *Produktraum* von  $(\Omega_1, \mathcal{F}_1, P_1)$  und  $(\Omega_2, \mathcal{F}_2, P_2)$ .  $P_{12}$  heißt *Produktmaß*.

#### Bemerkungen:

1.  $P_{12}$  nimmt an, dass die den beiden Wahrscheinlichkeitsräumen zugrundeliegenden Zufallsexperimente unabhängig sind. Auch andere Wahrscheinlichkeitsmaße können auf  $(\Omega_1 \times \Omega_2, \mathcal{G})$  konstruiert werden.
2. Für  $\Omega_1$  und  $\Omega_2$  endlich ist diese technische Diskussion unnötig.

### 1.7 Ausführliche Beispiele

**Beispiel 1.7.1** (Random Walk). Ein Jaguar-Händler und ein Kunde machen folgendes Spiel, da der Kunde nicht die erforderlichen  $N$  Euro für sein Traumauto dabei hat: Sie werfen eine faire Münze. Bei Kopf ( $K$ ) kriegt der Kunde einen Euro vom Händler, bei Zahl ( $Z$ ) muss der Kunde dem Händler einen Euro geben. Ausgehend von einem Anfangsvermögen von  $k$  [Euro] stellt sich folgende Frage: Wie groß ist die Wahrscheinlichkeit dafür, dass der Kunde ohne Geld und Auto (also bankrott) das Autohaus verlässt?

- $A$  : "Mann ist letztendlich bankrott",  
 $B$  : "erster Wurf zeigt  $K$ " = "Mann gewinnt einen Euro im ersten Spiel".

Schreibe nun  $P(\text{"bankrott"}) = P_k(\text{"bankrott"})$  explizit als Funktion des Anfangsvermögens  $k$ :

$$\begin{aligned} P_k(A) &= P_k(A|B) \cdot P(B) + P_k(A|B^c) \cdot P(B^c) \\ &= P_{k+1}(A) \cdot 0.5 + P_{k-1}(A) \cdot 0.5. \end{aligned}$$

Mit  $p_k := P_k(A)$  folgt:

$$p_k = 0.5(p_{k+1} + p_{k-1}) \quad \forall 0 < k < N.$$

Definiert man unter den Nebenbedingungen  $p_0 = 1$  und  $p_N = 0$  nun  $b_k = p_k - p_{k-1}$  ( $0 < k \leq N$ ), so folgt  $b_k = b_{k-1}$  und damit  $b_k = b_1$  für alle  $0 < k \leq N$ . Daraus ergibt sich folgende Rekursion:

$$\begin{aligned} p_k &= b_k + p_{k-1} \\ &= 2b_1 + p_{k-2} \\ &\quad \vdots \\ &= kb_1 + p_0 \\ &= kb_1 + 1. \end{aligned}$$

Aus  $p_N = 0$  folgt schließlich  $b_1 = -1/N$  und damit  $p_k = 1 - k/N$  für  $k = 1, \dots, N$ .

## 2 Zufallsvariablen und deren Verteilungen

### 2.1 Zufallsvariablen

Zufallsvariablen sind Funktionen von  $\Omega$  in die reellen Zahlen  $\mathbb{R}$ .

**Beispiel 2.1.1** (Zweimaliger Münzwurf).  $\Omega = \{ZZ, KK, KZ, ZK\}$ . Sei  $X(\omega)$  die Anzahl an  $K$ 's:

$$X(KK) = 2, \quad X(KZ) = X(ZK) = 1, \quad X(ZZ) = 0.$$

Die wichtigste Größe zur Beschreibung von Zufallsvariablen ist die *Verteilungsfunktion*  $F(x)$ , die Wahrscheinlichkeit, dass die Zufallsvariable  $X$  nicht größer als  $x$  ist. Genauer: Sei  $(\Omega, \mathcal{F}, P)$  ein Wahrscheinlichkeitsraum. Dann ist  $F(x) = P(A(x))$ , wobei  $A(x) \subseteq \Omega$  mit  $A(x) := \{\omega \in \Omega | X(\omega) \leq x\}$ . Damit diese Aussage Sinn hat, muss zusätzlich gelten:  $A(x) \in \mathcal{F}$  für alle  $x \in \mathbb{R}$ .

**Definition 2.1.2** (Zufallsvariable). Eine *Zufallsvariable* ist eine Funktion  $X : \Omega \rightarrow \mathbb{R}$  mit der Eigenschaft, dass  $\{\omega \in \Omega | X(\omega) \leq x\} \in \mathcal{F}$  für alle  $x \in \mathbb{R}$ . Eine solche Funktion nennt man  *$\mathcal{F}$ -messbar*.

**Notation:**  $X, Y, Z$  etc. sind Zufallsvariablen,  $x, y, z$  deren (reelle) Realisationen.

**Definition 2.1.3** (Verteilungsfunktion). Die *Verteilungsfunktion* einer Zufallsvariablen  $X$  ist die Funktion  $F : \mathbb{R} \rightarrow [0, 1]$  mit  $F(x) = P(X \leq x)$ .

**Bemerkungen:**

1. Folgende Abkürzungen werden häufig verwendet:

$$\{\omega \in \Omega | X(\omega) \leq x\} = \{\omega | X(\omega) \leq x\} = \{X \leq x\}.$$

Beachte: Der erste Ausdruck ist ein Ereignis, also  $\in \mathcal{F}$ , da  $X$  eine Zufallsvariable ist.

2. Will man betonen, dass  $F(x)$  die Verteilungsfunktion der Zufallsvariablen  $X$  ist, so schreibt man explizit

$$F_X(x) = F(x).$$

**Beispiel** (Fortsetzung von Beispiel 2.1.1): Hier ergibt sich bei fairer Münze, die zweimal unabhängig voneinander geworfen wird, folgende Verteilungsfunktion:

$$F_X(x) = \begin{cases} 0 & \text{für } x < 0, \\ \frac{1}{4} & \text{für } 0 \leq x < 1, \\ \frac{3}{4} & \text{für } 1 \leq x < 2, \\ 1 & \text{für } 2 \leq x. \end{cases}$$

**Lemma 2.1.4.** Eine Verteilungsfunktion  $F$  hat folgende Eigenschaften:

- (a) *Normiertheit:*  $\lim_{x \rightarrow -\infty} F(x) = 0$  und  $\lim_{x \rightarrow +\infty} F(x) = 1$ .
- (b) *Monotonie:* Für  $x < y$  ist  $F(x) \leq F(y)$ .
- (c) *Rechtsstetigkeit:*  $\lim_{h \downarrow 0} F(x+h) = F(x)$  für alle  $x \in \mathbb{R}$ .

*Beweis zu Teil (a).* Sei  $B_n = \{\omega \in \Omega | X(\omega) \leq -n\} = \{X \leq -n\}$  für  $n = 1, 2, \dots$ . Es gilt:

$$B_1 \supseteq B_2 \supseteq \dots \text{ mit } \bigcap_{n=1}^{\infty} B_n = \emptyset.$$

Mit der Stetigkeit von  $P$  folgt

$$P(B_n) \longrightarrow P(\emptyset) = 0 \quad \text{für } n \rightarrow \infty.$$

Der zweite Teil der Aussage lässt sich analog beweisen. □

**Beispiel 2.1.5.**  $X(\omega) = c \in \mathbb{R}$  für alle  $\omega \in \Omega$  ist eine konstante (deterministische) Zufallsvariable mit

$$F(x) = \begin{cases} 0 & \text{für } x < c, \\ 1 & \text{für } c \leq x. \end{cases}$$

Allgemein nennt man  $X$  *fast sicher konstant* (Schreibweise: f.s. const.), falls es ein  $c \in \mathbb{R}$  gibt mit  $P(X = c) = 1$ .

**Beispiel 2.1.6** (Bernoullivariablen). Wir betrachten einen Münzwurf mit  $P(K) = \pi$  (vergleiche hierzu Beispiel 1.3.2). Definiere

$$\begin{aligned} X(K) &= 1, \\ X(Z) &= 0. \end{aligned}$$

Hierbei handelt es sich um eine *Bernoullivariablen*. Sie ist die einfachste nicht-triviale Zufallsvariable und hat die Verteilungsfunktion

$$F(x) = \begin{cases} 0 & \text{für } x < 0, \\ 1 - \pi & \text{für } 0 \leq x < 1, \\ 1 & \text{für } 1 \leq x. \end{cases}$$

Man sagt auch,  $X$  ist *Bernoulliverteilt* mit Parameter  $\pi \in [0, 1]$ . Abkürzung:  $X \sim B(\pi)$ .

**Beispiel 2.1.7** (Indikatorfunktion). Zu einem Ereignis  $A \in \mathcal{F}$  wird die *Indikatorfunktion*  $I_A : \Omega \rightarrow \mathbb{R}$  von  $A$  definiert durch

$$I_A(\omega) = \begin{cases} 1 & \text{falls } \omega \in A, \\ 0 & \text{falls } \omega \notin A \text{ bzw. } \omega \in A^c. \end{cases}$$

$I_A$  ist also Bernoulliverteilt mit Parameter  $\pi = P(A)$ .

Nützlich ist folgende Identität: Sei  $\{B_i | i \in I\}$  eine Familie von disjunkten Ereignissen mit  $A \subseteq \bigcup_{i \in I} B_i$ . Dann gilt:

$$I_A = \sum_{i \in I} I_{A \cap B_i}$$

**Bemerkung:**  $\bigcup_{i \in I} B_i$  muss nicht notwendigerweise gleich  $\Omega$  sein.

**Lemma 2.1.8.** Sei  $F(x)$  die Verteilungsfunktion der Zufallsvariable  $X$ . Dann gilt:

- (a)  $P(X > x) = 1 - F(x)$ .
- (b)  $P(x < X \leq y) = F(y) - F(x)$ .
- (c)  $P(X = x) = F(x) - \lim_{y \uparrow x} F(y)$ .

## 2.2 Das Gesetz vom Durchschnitt (The Law of Averages)

Der nun folgende Abschnitt wird häufig auch mit *Gesetz der großen Zahlen* (engl. *Law of Large Numbers*) überschrieben. Vergleiche hierzu Kapitel 1.1.

Betrachte eine Folge  $A_1, A_2, \dots, A_n$  von unabhängigen Ereignissen mit  $P(A_i) = p$  für  $0 \leq p \leq 1$ . Betrachte nun die Summe  $S_n$  der Indikatorfunktionen  $I_{A_1}, I_{A_2}, \dots, I_{A_n}$ ,  $S_n = \sum_{i=1}^n I_{A_i}$ .  $S_n$  ist als Summe von Zufallsvariablen wieder eine Zufallsvariable, und es gilt folgender Satz:

**Satz 2.2.1.** *Die Zufallsvariable  $S_n/n$  konvergiert gegen  $p$  in dem Sinne, dass für beliebiges  $\varepsilon > 0$  gilt:*

$$P\left(p - \varepsilon \leq \frac{1}{n}S_n \leq p + \varepsilon\right) \longrightarrow 1 \quad \text{für } n \rightarrow \infty.$$

*Beweis.* Siehe Literatur. □

Die axiomatisch eingeführte Wahrscheinlichkeitstheorie steht also im Einklang mit der frequentistischen Interpretation von Wahrscheinlichkeiten (vgl. Kapitel 1.1).

## 2.3 Diskrete und stetige Zufallsvariablen

**Definition 2.3.1** (diskrete Zufallsvariable). Eine Zufallsvariable heißt *diskret*, falls sie nur Werte in einer maximal abzählbaren Teilmenge  $\{x_1, x_2, \dots\}$  von  $\mathbb{R}$  annimmt. Die diskrete Zufallsvariable  $X$  besitzt die *Wahrscheinlichkeitsfunktion* (auch *Dichtefunktion*, engl. *probability function, mass function, density function*)

$$f : \mathbb{R} \rightarrow [0, 1] \quad \text{mit } f(x) = P(X = x).$$

Man nennt  $\mathcal{T}_X = \{x | f(x) > 0\}$  den *Träger* der Zufallsvariablen  $X$ .

Diskrete Zufallsvariablen besitzen also maximal abzählbaren Träger.

**Beachte:** Die Verteilungsfunktion einer diskreten Zufallsvariable ist immer stückweise konstant mit Sprungstellen an allen Elementen ihres Trägers.

**Definition 2.3.2** (stetige Zufallsvariable). Eine Zufallsvariable  $X$  heißt *stetig*, wenn ihre Verteilungsfunktion  $F$  in der Form

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

geschrieben werden kann, wobei

$$f : \mathbb{R} \rightarrow [0, \infty)$$

integrierbar sein muss.  $f(x)$  heißt *Dichte(funktion)* von  $X$ . Auch hier nennt man  $\mathcal{T}_X = \{x | f(x) > 0\}$  den *Träger* von  $X$ .

**Beachte:** Man verwendet also sowohl für die Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariablen als auch für die Dichtefunktion einer stetigen Zufallsvariablen das gleiche Symbol  $f(x)$ . Diese haben aber unterschiedliche Eigenschaften, insbesondere gilt zwar in beiden Fällen

$$f(x) \geq 0 \quad \text{für alle } x \in \mathbb{R},$$

bei der diskreten Zufallsvariable muss aber zusätzlich noch

$$f(x) \leq 1 \quad \text{für alle } x \in \mathbb{R}$$

gelten.

**Beispiel 2.3.3** (für eine diskrete Zufallsvariable). Vergleiche hierzu Beispiel 2.1.1.

**Beispiel 2.3.4** (für eine stetige Zufallsvariable). Ein gerader Stab wird zufällig auf eine Ebene geworfen. Gemessen wird der Winkel  $\omega$  zwischen der Längsachse des Stabes und einer Referenzrichtung (z.B. Norden). Als Ergebnis scheint  $\omega \in [0, 2\pi) =: \Omega$  sinnvoll, ebenso das Wahrscheinlichkeitsmaß  $P((a, b)) = (b - a)/2\pi$ .

Die Frage nach der  $\sigma$ -Algebra  $\mathcal{F}$  wird später behandelt. Alle "angenehmen" Teilmengen von  $\Omega$  sind darin enthalten, insbesondere alle offenen Intervalle  $(a, b)$  mit  $0 \leq a < b < 2\pi$ . Definiere nun die beiden Zufallsvariablen

$$\begin{aligned} X(\omega) &= \omega \\ Y(\omega) &= \omega^2 = [X(\omega)]^2, \text{ oder kurz: } Y = X^2. \end{aligned}$$

Wie lauten die zugehörigen Verteilungsfunktionen?

$$\begin{aligned} F_X(x) &= P(\{\omega \in \Omega | X(\omega) \leq x\}) \\ &= P(\{\omega \in \Omega | 0 \leq X(\omega) \leq x\}) \\ &= P(\{\omega \in \Omega | 0 \leq \omega \leq x\}) \\ &= \frac{x}{2\pi} \quad \text{für } 0 \leq x \leq 2\pi. \end{aligned}$$

Insgesamt:

$$F_X(x) = \begin{cases} 0 & \text{für } x < 0, \\ \frac{x}{2\pi} & \text{für } 0 \leq x < 2\pi, \\ 1 & \text{für } 2\pi \leq x. \end{cases}$$

Analog:

$$\begin{aligned} F_Y(y) &= P(\{\omega \in \Omega | Y(\omega) \leq y\}) \\ &= P(\{\omega \in \Omega | 0 \leq Y(\omega) \leq y\}) \\ &= P(\{\omega \in \Omega | 0 \leq \omega^2 \leq y\}) \\ &= P(\{\omega \in \Omega | 0 \leq \omega \leq \sqrt{y}\}) \\ &= P(X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) \\ &= \frac{\sqrt{y}}{2\pi} \quad \text{für } 0 \leq y \leq 4\pi^2. \end{aligned}$$

Insgesamt:

$$F_Y(y) = \begin{cases} 0 & \text{für } y < 0, \\ \frac{\sqrt{y}}{2\pi} & \text{für } 0 \leq y < 4\pi^2, \\ 1 & \text{für } 4\pi^2 \leq y. \end{cases}$$

Nach Definition 2.3.2 folgt mit Differentiation:

$$f_X(u) = \begin{cases} \frac{1}{2\pi} & \text{für } 0 \leq u \leq 2\pi, \\ 0 & \text{sonst,} \end{cases}$$

$$f_Y(u) = \begin{cases} \frac{1}{\sqrt{u}4\pi} & \text{für } 0 < u \leq 4\pi^2, \\ 0 & \text{sonst.} \end{cases}$$

**Beispiel 2.3.5** (für eine gemischt *diskret-stetige* Zufallsvariable). Es wird eine Münze geworfen. Beim Ergebnis Kopf (Wahrscheinlichkeit  $p$ ) wird zusätzlich ein Stab wie in Beispiel 2.3.4 geworfen und der Winkel  $x$  gemessen. Es ist

$$\Omega = \{Z\} \cup \{(K, x) | 0 \leq x < 2\pi\}.$$

Betrachte nun die Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  mit

$$X(Z) = -1,$$

$$X((K, x)) = x.$$

Man sagt: "Die Zufallsvariable ist stetig mit Ausnahme eines Punktmaßes im Punkte  $x = -1$ ."

## 2.4 Zufallsvektoren

$$\begin{array}{ccc} & X_1 \nearrow & \mathbb{R} \\ (\Omega, \mathcal{F}, P) & & \\ & X_2 \searrow & \mathbb{R} \end{array}$$

**Frage:** Wie stehen die Zufallsvariablen  $X_1$  und  $X_2$  zueinander in Beziehung?

**Definition 2.4.1.** Ein *Zufallsvektor*  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  ist eine Abbildung von  $\Omega$  in  $\mathbb{R}^n$ , wobei jede Komponente  $X_i$  von  $\mathbf{X}$  eine Zufallsvariable ist und  $\mathbf{X}$  wieder  $\mathcal{F}$ -messbar sein soll.

**Definition 2.4.2.** Die *gemeinsame Verteilungsfunktion*  $F_{\mathbf{X}}$  eines Zufallsvektors  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  ist die Funktion  $F_{\mathbf{X}} : \mathbb{R}^n \rightarrow [0, 1]$  mit

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) &= P(\mathbf{X} \leq \mathbf{x}) \\ &= P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \\ &= P(\{\omega \in \Omega | X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \dots, X_n(\omega) \leq x_n\}), \end{aligned}$$

wobei  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  eine Realisierung von  $\mathbf{X}$  ist. Es sei darauf hingewiesen, dass

$$\{\omega \in \Omega | X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \dots, X_n(\omega) \leq x_n\} \in \mathcal{F}$$

ist, da  $\mathbf{X}$   $\mathcal{F}$ -messbar.

**Lemma 2.4.3.** Die gemeinsame Verteilungsfunktion  $F_{X,Y}(x, y)$  eines bivariaten Zufallsvektors  $(X, Y)$  hat folgende Eigenschaften:

(a) *Positivität und Normiertheit:*

$$\begin{aligned} \text{Positivität: } \lim_{x, y \rightarrow -\infty} F_{X,Y}(x, y) &= 0, \\ \text{Normiertheit: } \lim_{x, y \rightarrow +\infty} F_{X,Y}(x, y) &= 1. \end{aligned}$$

(b) *Monotonie:* Für  $(x_1, y_1) \leq (x_2, y_2)$ , d.h.  $x_1 \leq x_2$  und  $y_1 \leq y_2$ , gilt:

$$F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2).$$

(c) *Stetigkeit von oben:*  $F$  ist stetig von oben (engl.: "continuous from above"), das heißt

$$\lim_{u, v \downarrow 0} F_{X,Y}(x+u, y+v) \rightarrow F_{X,Y}(x, y).$$

(d) *Randverteilungen:*

$$\begin{aligned} \lim_{y \rightarrow \infty} F_{X,Y}(x, y) &= F_X(x), \\ \lim_{x \rightarrow \infty} F_{X,Y}(x, y) &= F_Y(y). \end{aligned}$$

Man nennt  $F_X(x)$  und  $F_Y(y)$  die *Rand-* oder *Marginalverteilungen* des Zufallsvektors  $(X, Y)$ . Diese sind nach Teil (d) von Lemma 2.4.3 aus  $F_{X,Y}(x, y)$  bestimmbar. Jedoch kann man im Allgemeinen *nicht*  $F_{X,Y}(x, y)$  aus  $F_X(x)$  und  $F_Y(y)$  berechnen.

**Bemerkung:** Ähnliche Resultate wie die von Lemma 2.4.3 gelten auch für  $n$ -dimensionale Zufallsvektoren.

**Definition 2.4.4.** Die Zufallsvariablen  $X$  und  $Y$ , definiert auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$ , heißen (*gemeinsam*) *diskret*, falls der Vektor  $(X, Y)$  nur maximal abzählbar viele Werte annehmen kann. Die (*gemeinsame*) *Wahrscheinlichkeitsfunktion* von  $(X, Y)$  ist dann definiert durch

$$f(x, y) = P(X = x, Y = y).$$

**Beispiel 2.4.5.** Ein Lehrer bittet seine Schüler, eine faire Münze zweimal zu werfen und das Ergebnis zu notieren. Ein strebsamer Schüler macht das tatsächlich, ein fauler wirft nur einmal und notiert das eine Ergebnis zweimal.

	1. Wurf	2. Wurf
Strebsamer Schüler	$X_S$	$Y_S$
Fauler Schüler	$X_F$	$Y_F$

Mit Definition 2.4.4 ergibt sich folglich:

Strebsamer Schüler			Fauler Schüler		
$f(x, y)$	$Y_S = 0$	$Y_S = 1$	$f(x, y)$	$Y_F = 0$	$Y_F = 1$
$X_S = 0$	$\frac{1}{4}$	$\frac{1}{4}$	$X_F = 0$	$\frac{1}{2}$	0
$X_S = 1$	$\frac{1}{4}$	$\frac{1}{4}$	$X_F = 1$	0	$\frac{1}{2}$

Obwohl  $X_S, Y_S, X_F, Y_F$  Zufallsvariablen mit identischen Wahrscheinlichkeitsfunktionen sind, sind die Verteilungsfunktionen der Zufallsvektoren  $(X_S, Y_S)$  und  $(X_F, Y_F)$  unterschiedlich. Zum Beispiel gilt:

$$\begin{aligned} P(X_S = Y_S = 1) &= \frac{1}{4}, \\ P(X_F = Y_F = 1) &= \frac{1}{2}. \end{aligned}$$

**Beispiel 2.4.6** ( $n$ -maliger Wurf einer 3-seitigen Münze). Die drei Ereignisse "Kopf", "Zahl" und "Rand" treten jeweils mit gleicher Wahrscheinlichkeit auf:

Ereignis	Wahrscheinlichkeit des Ereignisses
$K$ : Kopf	$P(K) = \frac{1}{3}$
$Z$ : Zahl	$P(Z) = \frac{1}{3}$
$R$ : Rand	$P(R) = \frac{1}{3}$

Seien  $K_n, Z_n, R_n$  die Anzahlen der entsprechenden Ergebnisse (Kopf, Zahl, Rand) bei  $n$  unabhängigen Würfen dieser Münze. Dann ist  $(K_n, Z_n, R_n)$  ein diskreter Zufallsvektor mit der Wahrscheinlichkeitsfunktion

$$P(K_n = k, Z_n = z, R_n = r) = \frac{n!}{k!z!r!} \left(\frac{1}{3}\right)^n, \quad (8)$$

wobei  $k, z, r \in \{0, 1, \dots, n\}$  und  $k + z + r = n$  gelten muss.

**Beachte:**  $K_n, Z_n$  und  $R_n$  sind (stochastisch) abhängig! Die Wahrscheinlichkeitsfunktion (8) ist ein Spezialfall der *Trinomialverteilung* (11), da  $P(K) = P(Z) = P(R) = \frac{1}{3}$ . Die Trinomialverteilung ist wiederum ein Spezialfall der *Multinomialverteilung* (12). Alle sind Verallgemeinerungen der *Binomialverteilung* (9) bzw. (10).

*Binomialverteilung* (mit Parametern  $n$  und  $\pi$ ): Für alle  $x \in \{0, 1, \dots, n\}$  gilt:

$$P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}. \quad (9)$$

Für andere  $x$  ist  $P(X = x) = 0$ .

Die folgende Aussage ist zur vorherigen äquivalent, jedoch zur Veranschaulichung der Verwandtschaft mit den folgenden Wahrscheinlichkeitsfunktionen dienlich:

Für alle  $\mathbf{x} = (x_1, x_2) \in \mathbb{N}_0^2$  mit der Eigenschaft  $x_1 + x_2 = n$  gilt:

$$P(\mathbf{X} = \mathbf{x}) = \frac{n!}{x_1!x_2!} \pi_1^{x_1} \pi_2^{x_2} \quad \text{mit } \pi_1 + \pi_2 = 1, \pi_1, \pi_2 \geq 0. \quad (10)$$

Für andere  $\mathbf{x}$  ist  $P(\mathbf{X} = \mathbf{x}) = 0$ .

*Trinomialverteilung* (mit Parametern  $n$  und  $\pi_1, \pi_2, \pi_3$ ): Für alle  $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{N}_0^3$  mit der Eigenschaft  $x_1 + x_2 + x_3 = n$  gilt:

$$P(\mathbf{X} = \mathbf{x}) = \frac{n!}{x_1!x_2!x_3!} \pi_1^{x_1} \pi_2^{x_2} \pi_3^{x_3} \quad \text{mit } \pi_1 + \pi_2 + \pi_3 = 1, \pi_1, \dots, \pi_3 \geq 0. \quad (11)$$

Für andere  $\mathbf{x}$  ist  $P(\mathbf{X} = \mathbf{x}) = 0$ .

*Multinomialverteilung* (mit Parametern  $n$  und  $\pi_1, \dots, \pi_p$ ): Für alle  $\mathbf{x} = (x_1, \dots, x_p)$  mit  $x_i \in \{0, 1, \dots, n\}$  und  $\sum_{i=1}^p x_i = n$  gilt:

$$P(\mathbf{X} = \mathbf{x}) = \frac{n!}{\prod_{i=1}^p x_i!} \prod_{i=1}^p \pi_i^{x_i} \quad \text{mit } \pi_1 + \dots + \pi_p = 1, \pi_1, \dots, \pi_p \geq 0. \quad (12)$$

Für andere  $\mathbf{x}$  ist  $P(\mathbf{X} = \mathbf{x}) = 0$ .

**Definition 2.4.7.** Die Zufallsvariablen  $X$  und  $Y$  nennt man (*gemeinsam*) *stetig*, falls deren (gemeinsame) Verteilungsfunktion geschrieben werden kann als

$$F_{X,Y}(x, y) = \int_{v=-\infty}^y \int_{u=-\infty}^x f_{X,Y}(u, v) du dv$$

für alle  $x, y \in \mathbb{R}$ , wobei  $f_{X,Y} : \mathbb{R}^2 \rightarrow [0, \infty)$  integrierbar sein muss und (*gemeinsame*) *Dichtefunktion* von  $(X, Y)$  heißt.

**Beispiel 2.4.8** (Wurf eines Dartpfeils auf eine Scheibe). Man wirft einen Dartpfeil auf eine Scheibe mit dem Radius  $\rho$ . Es wird angenommen, dass die Wahrscheinlichkeit, eine beliebige Region auf der Scheibe zu treffen, proportional zu ihrer Fläche ist. Weiterhin wird die Scheibe immer getroffen. Sei  $R$  der Abstand vom Mittelpunkt und  $\Theta$  der Winkel bzgl. einer Referenzrichtung (z.B. "oben"). Man kann zeigen, dass dann gilt:

$$P(R \leq r) = \frac{r^2}{\rho^2} \quad \text{und} \quad P(\Theta \leq \theta) = \frac{\theta}{2\pi}$$

für  $0 \leq r \leq \rho$  und  $0 \leq \theta \leq 2\pi$ , und ferner

$$P(R \leq r, \Theta \leq \theta) = P(R \leq r) \cdot P(\Theta \leq \theta) = \frac{r^2 \theta}{\rho^2 \cdot 2\pi}.$$

Es folgt daher durch Differentiation, dass die gemeinsame Dichtefunktion von  $R$  und  $\Theta$  gleich

$$f_{R,\Theta}(r, \theta) = \begin{cases} \frac{r}{\pi \rho^2} & \text{für } 0 \leq r \leq \rho \text{ und } 0 \leq \theta \leq 2\pi, \\ 0 & \text{sonst} \end{cases}$$

ist.

## 2.5 Monte Carlo Simulation

**Einfachstes Beispiel:** Wiederholter Wurf einer Münze, um die Wahrscheinlichkeit für Kopf bzw. Zahl empirisch zu bestimmen.

**Beispiel 2.5.1** (Buffon's Nadel). Eine Nadel der Länge 1 wird auf eine Ebene mit parallelen Geraden im Abstand 1 geworfen. Man kann zeigen:

$$P(\text{"Nadel kreuzt eine Linie"}) = \frac{2}{\pi}$$

für  $\pi = 3.1415926\dots$ . Folglich ist

$$\hat{\pi} = \frac{2}{\text{Relative Häufigkeit des Ereignisses "Nadel kreuzt Linie" bei } n \text{ Versuchen}}$$

ein Monte-Carlo-Schätzer für  $\pi$ .

**Beispiel 2.5.2** (Monte Carlo Integration). Dieses Verfahren stellt häufig bei höher-dimensionalen Integrationsberechnungen die beste Methode zur approximativen Bestimmung des Integrals von analytisch sehr schwer zu integrierenden Funktionen dar.

Sei  $g : [0, 1] \rightarrow [0, 1]$  stetig, aber analytisch "schwer" zu integrieren. Die Aufgabe lautet: Berechne  $\int_0^1 g(u) du$ . Hierzu bietet sich die *Hit or Miss*-Technik an: Der Computer kann leicht gleichverteilte (Pseudo-)Zufallszahlen  $(X, Y)$  in einem Einheitsquadrat erzeugen, d.h.  $P((x, y) \in A) = |A|$  für alle  $A \subseteq [0, 1]^2$ . Für jede Zufallszahl/Realisation  $(x, y)$  kann leicht überprüft werden, ob  $y \leq g(x)$  ist. Die relative Häufigkeit des Ereignisses  $A = \{y \leq g(x)\}$  konvergiert für unendlich viele Realisationen gegen  $|A|$ , also gegen  $\int_0^1 g(u) du$ .

**Beispiel 2.5.3.** Berechne für  $g(x) = \arctan(x)$  das Integral auf dem Einheitsintervall  $[0, 1]$ . Monte Carlo Integration mit  $n$  Zufallszahlen liefert (zum Beispiel)

n	Monte Carlo Integration
10	0.6
100	0.54
1000	0.439
10000	0.4355
100000	0.44065
1000000	0.439235

Die analytische Lösung lautet

$$\begin{aligned}
 \int_0^1 \arctan(x) dx &= \left[ x \arctan(x) - \frac{1}{2} \ln(1 + x^2) \right]_0^1 \\
 &= \frac{\pi}{4} - \frac{1}{2} \ln(2) \\
 &= 0.4388246.
 \end{aligned}$$

**Bemerkung:** Auch der Fehler der Schätzung kann berechnet werden.

### 3 Diskrete Zufallsvariablen

#### 3.1 Wahrscheinlichkeitsfunktion

Zu Beginn seien zwei Aussagen wiederholt:

1. Die Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariable  $X$  ist die Funktion  $f : \mathbb{R} \rightarrow [0, 1]$  mit  $f(x) = P(X = x)$  (vgl. Definition 2.3.1).
2. Der Zusammenhang zwischen Wahrscheinlichkeitsfunktion und Verteilungsfunktion stellt sich im diskreten Fall wie folgt dar:

$$\begin{aligned} F(x) &= \sum_{i | x_i \leq x} f(x_i), \\ f(x) &= F(x) - \lim_{y \uparrow x} F(y). \end{aligned}$$

**Lemma 3.1.1.** *Die diskrete Wahrscheinlichkeitsfunktion  $f$  erfüllt folgende Eigenschaften:*

(a) *Der Träger  $\mathcal{T}_X = \{x | f(x) > 0\}$  ist abzählbar.*

(b)  $\sum_{x_i \in \mathcal{T}_X} f(x_i) = 1$ .

**Beispiel 3.1.2** (Binomialverteilung). Eine Münze werde  $n$ -mal geworfen. Die Erfolgswahrscheinlichkeit (etwa die Wahrscheinlichkeit für das Ergebnis "Kopf") sei  $\pi \in [0, 1]$ . Definiere die Zufallsvariable  $X :=$  "Anzahl der Würfe mit dem Ergebnis Kopf". Der Träger von  $X$  ist  $\mathcal{T} = \{0, 1, 2, \dots, n\}$ , und

$$f(x) = \begin{cases} \binom{n}{x} \pi^x (1 - \pi)^{n-x} & \text{für } x = 0, \dots, n, \\ 0 & \text{sonst.} \end{cases}$$

Man schreibt  $X \sim \text{Bin}(n, \pi)$  oder auch  $X \sim B(n, \pi)$ .

**Bemerkung:**  $X$  ist die Summe von  $n$  unabhängigen Bernoullivariablen  $Y_i \sim B(\pi)$ ,  $i = 1, 2, \dots, n$ . Also:

$$Y_i \sim B(\pi) \text{ unabhängig} \implies \sum_i^n Y_i \sim B(n, \pi).$$

**Beispiel 3.1.3** (Poissonverteilung). Die *Poissonverteilung* kann als ein Zählvorgang innerhalb eines bestimmten Zeitraumes aufgefasst werden. Gezählt werden etwa gemeldete Schadensfälle bei einer Versicherung innerhalb eines Jahres. Träger der Verteilung ist  $\mathcal{T} = \{0, 1, 2, \dots\} = \mathbb{N}_0$ . Die Wahrscheinlichkeitsfunktion lautet

$$f(x) = \begin{cases} \frac{\lambda^x}{x!} \exp(-\lambda) & x \in \mathbb{N}_0, \\ 0 & \text{sonst,} \end{cases}$$

wobei der Parameter  $\lambda > 0$  *Rate* (der Schadensfälle o.Ä.) heißt. Man schreibt  $X \sim \text{Po}(\lambda)$ .

### 3.2 Unabhängigkeit

**Definition 3.2.1.** Zwei diskrete Zufallsvariablen  $X$  und  $Y$  heißen *unabhängig*, falls die Ereignisse  $\{X = x\} = \{\omega \in \Omega | X(\omega) = x\}$  und  $\{Y = y\}$  für alle  $x \in \mathbb{R}$  und  $y \in \mathbb{R}$  unabhängig sind. Kurz: Gilt

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) \quad \text{für alle } x, y \in \mathbb{R},$$

so sind  $X$  und  $Y$  unabhängig. (Man kann sich hier auch auf  $x \in \mathcal{T}_X$  und  $y \in \mathcal{T}_Y$  beschränken.)

**Beispiel 3.2.2.**  $N$ -maliger unabhängiger Münzwurf mit Wahrscheinlichkeit  $p$  für Kopf und  $q = 1 - p$  für Zahl. Betrachte:

$X$ : Anzahl der Kopfwürfe,  
 $Y$ : Anzahl der Zahlwürfe.

Wählt man den Stichprobenumfang als  $N \sim \text{Po}(\lambda)$ ,  $\lambda > 0$ , so sind  $X$  und  $Y$  unabhängig. Denn:

Das Modell lautet:

$$\begin{aligned} N &\sim \text{Po}(\lambda) \\ X|N &\sim B(N, p) \\ Y|N &\sim B(N, q = 1 - p). \end{aligned}$$

Dann:

$$\begin{aligned} P(X = x, Y = y) &= P(X = x, Y = y, N = x + y) \\ &= P(X = x, Y = y | N = x + y) \cdot P(N = x + y) \\ &= P(X = x | N = x + y) \cdot P(N = x + y) \\ &= \binom{x + y}{x} p^x q^y \cdot \frac{\lambda^{x+y}}{(x + y)!} e^{-\lambda} \\ &= \frac{(\lambda p)^x (\lambda q)^y}{x! y!} e^{-\lambda}. \end{aligned}$$

Ferner gilt nach dem Satz der totalen Wahrscheinlichkeit:

$$\begin{aligned} P(X = x) &= \sum_{n \geq x} P(X = x | N = n) \cdot P(N = n) \\ &= \sum_{n \geq x} \binom{n}{x} p^x q^{n-x} \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \sum_{n \geq x} \frac{1}{x!(n-x)!} p^x q^{n-x} \lambda^n e^{-\lambda} \end{aligned}$$

$$\begin{aligned}
&= \frac{(\lambda p)^x}{x!} e^{-\lambda} \sum_{n \geq x} \frac{(\lambda q)^{n-x}}{(n-x)!} \\
&= \frac{(\lambda p)^x}{x!} e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda q)^n}{n!} \\
&= \frac{(\lambda p)^x}{x!} e^{-\lambda} e^{\lambda q} \\
&= \frac{(\lambda p)^x}{x!} e^{-\lambda(q-1)} \\
&= \frac{(\lambda p)^x}{x!} e^{-\lambda p}.
\end{aligned}$$

Analog ergibt sich:

$$P(Y = y) = \frac{(\lambda q)^y}{y!} e^{-\lambda q},$$

also  $X \sim \text{Po}(\lambda p)$  und  $Y \sim \text{Po}(\lambda q)$ . Insgesamt:

$$\begin{aligned}
P(X = x) \cdot P(Y = y) &= \frac{(\lambda p)^x}{x!} e^{-\lambda p} \cdot \frac{(\lambda q)^y}{y!} e^{-\lambda q} \\
&= \frac{(\lambda p)^x}{x!} e^{-\lambda p} \cdot \frac{(\lambda q)^y}{y!} e^{-\lambda(1-p)} \\
&= \frac{(\lambda p)^x (\lambda q)^y}{x! y!} e^{-\lambda} \\
&= P(X = x, Y = y).
\end{aligned}$$

Damit ist die Bedingung für stochastische Unabhängigkeit erfüllt.

**Bemerkung:** Ist der Stichprobenumfang  $N$  fest, sind  $X$  und  $Y$  natürlich abhängig. Zum Beispiel für  $N = 1$ :

$$P(X = Y = 1) = 0 \neq P(X = 1) \cdot P(Y = 1) = p \cdot q.$$

Der folgende Satz hat große Bedeutung:

**Satz 3.2.3.** Seien  $X$  und  $Y$  unabhängige Zufallsvariablen und  $g, h : \mathbb{R} \rightarrow \mathbb{R}$  beliebige Funktionen. Dann sind auch  $g(X)$  und  $h(X)$  unabhängig.

**Bemerkung:** Die Unabhängigkeit einer Familie von Zufallsvariablen  $\{X_i | i \in I\}$  definiert man analog zu Definition 3.2.1. Auch bei Zufallsvariablen folgt aus paarweiser Unabhängigkeit im Allgemeinen *nicht* Unabhängigkeit. Die Definition von bedingter Unabhängigkeit von Zufallsvariablen gegeben ein Ereignis  $C$  erfolgt ebenfalls analog zu Definition 3.2.1.

### 3.3 Erwartungswert

**Definition 3.3.1.** Der *Erwartungswert* einer diskreten Zufallsvariablen  $X$  mit Wahrscheinlichkeitsfunktion  $f(x)$  ist definiert als

$$E(X) = \sum_{x \in \mathcal{I}} x f(x),$$

falls diese Summe *absolut konvergent* ist, d.h. falls

$$\sum_{x \in \mathcal{T}} |x|f(x) < \infty.$$

**Bemerkungen:**

1. Häufige Abkürzung:  $\mathbf{E} X = \mathbf{E}(X)$ .
2. Häufige Schreibweise:  $\sum_{x \in \mathcal{T}} xf(x) = \sum_{x \in \mathbb{R}} xf(x)$ . Hierbei ist zu beachten, dass es sich bei der linken Seite um eine Summe über eine abzählbare Menge handelt, wohingegen rechts die Summe eigentlich über eine überabzählbare Menge geht, wobei aber nur abzählbar viele Elemente ungleich null sind.

**Beispiel 3.3.2.** Es folgen zwei Beispiele für diskrete Zufallsvariablen ohne Erwartungswert.

- (a) Sei  $f(x) = \frac{1}{x(x-1)}$  für  $x = 2, 3, \dots$ . Dann ist  $f(x)$  wegen  $\sum_{x=2}^{\infty} f(x) = 1$  fraglos eine Wahrscheinlichkeitsfunktion. Aber es gilt:

$$E(X) = \sum_{x=2}^{\infty} xf(x) = \sum_{x=2}^{\infty} \frac{1}{x-1} = \sum_{x=1}^{\infty} \frac{1}{x} = \infty,$$

da die letzte Summe die *harmonische Reihe* ist, die bekanntermaßen nicht konvergiert.

- (b) Sei  $f(x) = A/x^2$  für  $x = \pm 1, \pm 2, \pm 3, \dots$  die Wahrscheinlichkeitsfunktion einer Zufallsvariable  $X$ , wobei  $A = [\sum_{x \neq 0} \frac{1}{x^2}]^{-1}$ . Die Reihe  $\sum_{x \neq 0} xf(x)$  ist nicht absolut konvergent, da

$$\sum_{x \neq 0} |xf(x)| = \sum_{x \neq 0} A|\frac{1}{x}| = \infty.$$

Bei  $\sum_{x \neq 0} |\frac{1}{x}| = \infty$  handelt es sich um das doppelte der harmonischen Reihe.

Es bleibt fest zu halten: Obwohl  $f(x)$  symmetrisch um 0 ist, ist  $\mathbf{E}(X) \neq 0$ , da  $\mathbf{E}(X)$  nicht existiert. Generell gilt: Aus Symmetrie um  $c$  folgt nicht automatisch  $\mathbf{E}(X) = c$ !

Sei  $X$  nun eine diskrete Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  und  $g : \mathbb{R} \rightarrow \mathbb{R}$  eine reelle Funktion. Wie laut der Erwartungswert von  $Y = g(X)$ ?

**Methode 1:** Berechne zunächst  $f_Y(y)$  und dann

$$\mathbf{E}(Y) = \sum_{y \in \mathcal{T}_Y} yf_Y(y).$$

**Methode 2:** Nach folgendem Lemma 3.3.3:

$$\mathbf{E}(Y) = \mathbf{E}(g(X)) = \sum_{x \in \mathcal{T}_X} g(x) \cdot f(x).$$

**Lemma 3.3.3.** Sei  $X$  eine diskrete Zufallsvariable mit Wahrscheinlichkeitsfunktion  $f(x)$  und  $g: \mathbb{R} \rightarrow \mathbb{R}$ , so ist

$$E(g(X)) = \sum_{x \in \mathcal{T}_X} g(x) \cdot f(x),$$

falls diese Summe absolut konvergent ist.

**Beispiel 3.3.4.** Eine Zufallsvariable  $X$  nehme die Werte  $-2$ ,  $-1$ ,  $1$  oder  $3$  mit den Wahrscheinlichkeiten  $\frac{1}{4}$ ,  $\frac{1}{8}$ ,  $\frac{1}{4}$ ,  $\frac{3}{8}$  an. Berechne  $E(X^2)$ .

**Methode 1:**  $Y = X^2$  hat die Ausprägungen  $1$ ,  $4$ ,  $9$  mit Wahrscheinlichkeiten  $\frac{3}{8}$ ,  $\frac{1}{4}$  und  $\frac{3}{8}$ . Somit ergibt sich:

$$E(Y) = 1 \cdot \frac{3}{8} + 4 \cdot \frac{1}{4} + 9 \cdot \frac{3}{8} = 4\frac{3}{4}.$$

**Methode 2:**  $E(Y) = E(X^2) = (-2)^2 \cdot \frac{1}{4} + (-1)^2 \cdot \frac{1}{8} + 1^2 \cdot \frac{1}{4} + 3^2 \cdot \frac{3}{8} = 4\frac{3}{4}$ .

**Definition 3.3.5.** Sei  $k$  eine positive ganze Zahl, so heißt

$$m_k = E(X^k)$$

das  $k$ -te Moment der Zufallsvariablen  $X$  und

$$\sigma_k = E((X - m_1)^k)$$

das  $k$ -te zentrale Moment.

Besonders wichtig sind  $m_1 = E(X)$  und  $\sigma^2 = \sigma_2 = E((X - E(X))^2) = \text{Var}(X)$ , die Varianz von  $X$ .  $\sqrt{\sigma^2} = \sigma$  heißt Standardabweichung. Zentrale Momente lassen sich über gewöhnliche Momente berechnen, zum Beispiel  $\sigma_2 = m_2 - m_1^2$ .

**Satz 3.3.6** (Verschiebungssatz). Sei  $X$  eine Zufallsvariable, für die  $E(X)$  und  $E(X^2)$  existieren, dann gilt:

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

**Beispiel 3.3.7.** Ist  $X \sim B(\pi)$ , so ist  $E(X) = \pi$  und  $\text{Var} = \pi(1 - \pi)$ .

**Beispiel 3.3.8.** Ist  $X \sim \text{Bin}(n, \pi)$ , so ist  $E(X) = n\pi$  und  $\text{Var} = n\pi(1 - \pi)$ .

**Satz 3.3.9.** Der Erwartungswert ist ein linearer Operator auf dem Raum aller Zufallsvariablen:

(a) Falls  $X \geq 0$ , so ist  $E(X) \geq 0$ .

(b) Für  $a, b \in \mathbb{R}$  ist  $E(aX + bY) = aE(X) + bE(Y)$ .

(c) Die konstante Variable  $X(\omega) = 1$  für alle  $\omega \in \Omega$  hat Erwartungswert  $E(X) = 1$ .

**Lemma 3.3.10.** Sind  $X$  und  $Y$  unabhängig, so gilt

$$E(X \cdot Y) = E(X) \cdot E(Y). \tag{13}$$

*Beweis.* Bei Unabhängigkeit von  $X$  und  $Y$  gilt:

$$\begin{aligned}
 E(X \cdot Y) &= \sum_x \sum_y xy f_{X,Y}(x, y) \\
 &= \sum_x \left( x f_X(x) \sum_y y f_Y(y) \right) \\
 &= \left( \sum_x x f_X(x) \right) \left( \sum_y y f_Y(y) \right) \\
 &= E(X) \cdot E(Y).
 \end{aligned}$$

□

**Definition 3.3.11.**  $X$  und  $Y$  heißen *unkorreliert*, wenn (13) gilt.

**Bemerkung zu Lemma 3.3.10 und Definition 3.3.11:** Aus Unabhängigkeit folgt Unkorreliertheit. Der Umkehrschluss ist im Allgemeinen jedoch *nicht* möglich. Aus (13) kann also nicht auf Unabhängigkeit geschlossen werden.

**Satz 3.3.12.** Für zwei Zufallsvariablen  $X$  und  $Y$  gilt:

(a)  $\text{Var}(aX + b) = a^2 \text{Var}(X)$  für alle  $a, b \in \mathbb{R}$ .

(b)  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ , falls  $X$  und  $Y$  unkorreliert sind.

**Satz 3.3.13.** Hat die diskrete Zufallsvariable  $X$  den Träger  $\mathcal{T}_X \subseteq \mathbb{N}_0$ , so gilt:

$$E(X) = \sum_{k=1}^{\infty} P(X \geq k) \left( = \sum_{k=0}^{\infty} P(X > k) = \sum_{k=0}^{\infty} (1 - F_X(k)) \right).$$

*Beweis.*

$$\begin{aligned}
 \sum_{k=1}^{\infty} P(X \geq k) &= \sum_{k=1}^{\infty} \sum_{t=k}^{\infty} P(X = t) \\
 &\stackrel{(*)}{=} \sum_{t=1}^{\infty} \sum_{k=1}^t P(X = t) \\
 &= \sum_{t=1}^{\infty} t P(X = t) \\
 &= \sum_{t=0}^{\infty} t P(X = t) \\
 &= E(X).
 \end{aligned}$$

(\*): Die Umnummerierung lässt sich graphisch wie folgt begründen:

□

**Zusatz:** Quantile werden im Allgemeinen wie folgt definiert: Sei  $X$  eine diskrete Zufallsvariable mit Verteilungsfunktion  $F$ , und sei  $p \in [0, 1]$ . Dann heißt jeder Wert  $\tilde{x}_p$ , für den

$$F(\tilde{x}_p) = P(X \leq \tilde{x}_p) \geq p \quad \text{und} \quad P(X \geq \tilde{x}_p) \geq 1 - p$$

gilt,  $p$ -Quantil der Zufallsvariable  $X$ .

*Problem:* Das  $p$ -Quantil ist im Allgemeinen nicht eindeutig. Daher gilt im Folgenden:

**Definition 3.3.14.** Das  $p$ -Quantil  $x_p$  einer Zufallsvariablen  $X$  ist definiert als der kleinste Wert  $x$ , für den  $F(x) \geq p$  gilt, genauer:

$$x_p = \min\{x \in \mathcal{T}_X \mid F(x) \geq p\}.$$

(Das Minimum existiert, da  $F$  stetig von rechts.) Man definiert die *Quantilsfunktion*  $F^-(p) = x_p$  für  $p \in [0, 1]$ .

**Beispiel 3.3.15.** Sei  $X \sim B(\pi)$ . Dann

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 - \pi & \text{für } 0 \leq x < 1 \\ 1 & \text{für } x \geq 1. \end{cases}$$

Für die Quantilsfunktion ergibt sich

$$F^-(p) = \begin{cases} 0 & \text{für } 0 \leq p \leq 1 - \pi \\ 1 & \text{für } 1 - \pi < p \leq 1. \end{cases}$$

Es gilt folgender Satz:

**Satz 3.3.16** (Inversionsverfahren). *Falls  $U$  stetig gleichverteilt auf dem Intervall  $[0, 1]$  ist, d.h.  $U \sim U(0, 1)$ , so hat  $F^-(U)$  die Verteilungsfunktion  $F$ .*

*Beweis.* Für eine beliebige Verteilungsfunktion  $F$  und ihre Quantilsfunktion  $F^-$  gilt:

1.  $F^-(u)$  ist das kleinste  $y$  mit  $F(y) \geq u$ , also insbesondere

$$F(F^-(u)) \geq u. \quad (14)$$

2. Da  $F$  rechtsstetig ist, gilt für  $x \geq \min(T_X)$

$$\min\{y \in \mathcal{T}_X | F(y) \geq F(x)\} = \min\{y \in \mathbb{R} | F(y) \geq F(x)\}.$$

In der Menge  $\{y \in \mathbb{R} | F(y) \geq F(x)\}$  ist auch  $x$  enthalten, deshalb

$$F^-(F(x)) = \min\{y \in \mathcal{T}_X | F(y) \geq F(x)\} = \min\{y \in \mathbb{R} | F(y) \geq F(x)\} \leq x. \quad (15)$$

Mit Hilfe dieser zwei Aussagen erhalten wir nun für  $x \geq \min(T_X)$

$$F^-(u) \leq x \iff u \leq F(x), \quad (16)$$

denn (" $\Rightarrow$ ")

$$\begin{aligned} F^-(u) &\leq x \\ \implies F(F^-(u)) &\leq F(x) \\ \stackrel{(14)}{\implies} &u \leq F(x) \end{aligned}$$

und (" $\Leftarrow$ ")

$$\begin{aligned} u &\leq F(x) \\ \implies F^-(u) &\leq F^-(F(x)) \\ \stackrel{(15)}{\implies} &F^-(u) \leq x, \end{aligned}$$

da sowohl  $F$  als auch  $F^-$  monoton wachsend ist. Aus (16) folgt schließlich für  $x \geq \min(T_X)$

$$P(F^-(U) \leq x) = P(U \leq F(x)) = F(x).$$

Für  $x < \min(T_X)$  gilt  $F(x) = 0$  nach Definition des Trägers und  $P(F^-(U) \leq x) = 0$ , da  $F^-(U) \geq \min(T_X)$  nach Definition von  $F^-$ .

Insgesamt erhalten wir  $P(F^-(U) \leq x) = F(x)$  für alle  $x \in \mathbb{R}$ , d.h.  $F$  ist die Verteilungsfunktion von  $F^-(U)$ .  $\square$

Im Beispiel Bernoulliverteilung (Bsp. 3.3.15) ergibt sich der folgende intuitiv einleuchtende Algorithmus:

```

U ~ U(0, 1)
if U ≤ (1 - π) return(0)
else return(1).

```

### 3.4 Beispiele für diskrete Zufallsvariablen

**Beispiel 3.4.1** (Trinomialverteilung). Betrachte  $n$  unabhängige Versuche eines Experiments mit drei möglichen Ausgängen mit den Wahrscheinlichkeiten  $\pi_1, \pi_2$  und  $\pi_3$ , wobei  $\pi_1 + \pi_2 + \pi_3 = 1$ . Dann heißt der Zufallsvektor  $\mathbf{X} = (X_1, X_2, X_3)^T$  *trinomialverteilt* und hat die Wahrscheinlichkeitsfunktion

$$f(\mathbf{X} = \mathbf{x}) = P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1!x_2!x_3!} \pi_1^{x_1} \pi_2^{x_2} \pi_3^{x_3}$$

für  $x_1 + x_2 + x_3 = n$ , wobei  $x_1, x_2, x_3 \geq 0$ . Man schreibt hierfür  $\mathbf{X} \sim M_3(n, (\pi_1, \pi_2, \pi_3)^T)$ .

**Eigenschaften:** Jede Komponente von  $\mathbf{X} \sim M_3(n, (\pi_1, \pi_2, \pi_3)^T)$  ist binomialverteilt:

$$\begin{aligned} X_1 &\sim B(n, p) \\ X_2 &\sim B(n, q) \\ X_3 &\sim B(n, r). \end{aligned}$$

Daher ergibt sich

$$\begin{aligned} E(X_1) &= np \\ E(X_2) &= nq \\ E(X_3) &= nr, \end{aligned}$$

Der Erwartungswert eines Zufallsvektors ist der Vektor der Erwartungswerte der einzelnen Komponenten, also

$$E(\mathbf{X}) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ E(X_3) \end{pmatrix} = n \begin{pmatrix} p \\ q \\ r \end{pmatrix}.$$

**Beispiel 3.4.2** (Poissonverteilung, vgl. Beispiel 3.1.3). Für den Erwartungswert einer mit Parameter  $\lambda$  Poissonverteilten Zufallsvariable  $X$  gilt

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x}{x!} \exp(-\lambda) \\ &= \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} \exp(-\lambda) \\ &= \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \exp(-\lambda) \\ &\stackrel{x-1=\tilde{x}}{=} \lambda \underbrace{\sum_{\tilde{x}=0}^{\infty} \frac{\lambda^{\tilde{x}}}{\tilde{x}!} \exp(-\lambda)}_{=\sum_{x=0}^{\infty} f(x)=1} \\ &= \lambda. \end{aligned}$$

Weiterhin

$$\begin{aligned}
 E(X^2) &= \sum_{x=0}^{\infty} x^2 \cdot \frac{\lambda^x}{x!} \exp(-\lambda) \\
 &= \sum_{x=1}^{\infty} x \frac{\lambda^x}{(x-1)!} \exp(-\lambda) \\
 &\stackrel{x-1=\tilde{x}}{=} \lambda \sum_{\tilde{x}=0}^{\infty} (\tilde{x}+1) \frac{\lambda^{\tilde{x}}}{\tilde{x}!} \exp(-\lambda) \\
 &= \lambda \left( \underbrace{\sum_{\tilde{x}=0}^{\infty} \tilde{x} \frac{\lambda^{\tilde{x}}}{\tilde{x}!} \exp(-\lambda)}_{\stackrel{\text{s.o.}}{=} E(X)=\lambda} + \underbrace{\sum_{\tilde{x}=0}^{\infty} \frac{\lambda^{\tilde{x}}}{\tilde{x}!} \exp(-\lambda)}_{=\sum_{x=0}^{\infty} f(x)=1} \right) \\
 &= \lambda[\lambda + 1] \\
 &= \lambda^2 + \lambda.
 \end{aligned}$$

Mit dem Verschiebungssatz (Satz 3.3.6) folgt:

$$\begin{aligned}
 \text{Var}(X) &= E(X^2) - [E(X)]^2 \\
 &= \lambda^2 + \lambda - [\lambda]^2 \\
 &= \lambda.
 \end{aligned}$$

Es ist charakteristisch für die Poissonverteilung, dass

$$E(X) = \text{Var}(X) = \lambda.$$

Eine Anwendung der Poissonverteilung ist die Approximation der Binomialverteilung:

**Satz 3.4.3.** Sei  $X \sim B(n, \pi)$ . Dann gilt für  $n \rightarrow \infty$ ,  $\pi \rightarrow 0$  mit  $n\pi = \lambda$ :

$$f_X(x) \longrightarrow \frac{\lambda^x}{x!} \exp(-\lambda) \quad \text{für } x = 0, 1, 2, \dots$$

Also:  $B(n, \pi) \approx \text{Po}(n\pi)$  für  $n$  "groß" und  $\pi$  "klein".

**Definition 3.4.4** (Geometrische Verteilung). Wiederhole unabhängige Versuche eines Bernoulliexperimentes mit Erfolgswahrscheinlichkeit  $\pi$ . Beobachte die Zufallsvariable

$X$  : "Anzahl der Experimente, bis zum ersten Mal ein Erfolg eintritt".

$X$  heißt dann *geometrisch verteilt* und hat die Wahrscheinlichkeitsfunktion

$$f_X(x) = \pi (1 - \pi)^{x-1} \quad \text{für alle } x \in \mathcal{T} = \{1, 2, \dots\}.$$

**Beachte:** Da hier

$$\begin{aligned}
 P(X > x) &= (1 - \pi)^x \\
 \implies F(x) &= 1 - P(X > x) \\
 &= 1 - (1 - \pi)^x
 \end{aligned}$$

für alle  $x \in \mathcal{T}$  gilt, folgt mit Satz 3.3.13:

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} [1 - F_X(x)] \\ &= \sum_{x=0}^{\infty} (1 - \pi)^x \\ &= \frac{1}{1 - (1 - \pi)} \\ &= \frac{1}{\pi}. \end{aligned}$$

Man kann ferner zeigen:

$$\text{Var}(X) = \frac{1 - \pi}{\pi^2}.$$

**Beachte:** Häufig wird auch die Anzahl  $Y$  der Versuche *vor* dem ersten Erfolg betrachtet. Es gilt also mit  $Y = X - 1$  und  $y \in \mathcal{T}_Y = \{0, 1, 2, \dots\}$

$$f_Y(y) = \pi (1 - \pi)^y.$$

Desweiteren:

$$E(Y) = E(X) - 1 = \frac{1}{\pi} - 1 = \frac{1 - \pi}{\pi}$$

und

$$\text{Var}(Y) = \text{Var}(X) = \frac{1 - \pi}{\pi^2}.$$

**Zusatz (Diskrete Zufallsvariablen in R):** Im Programmpaket R stehen für diverse Verteilungen Funktionen zur Berechnung der

- Wahrscheinlichkeitsfunktion  $f$  (erster Buchstabe: **d** für *distribution*),
- Verteilungsfunktion  $F$  (erster Buchstabe: **p** für *probability*) und
- Quantilsfunktion  $F^{-}$  (erster Buchstabe: **q** für *quantile*)

sowie zum

- Simulieren von Zufallszahlen aus einer gegebenen Verteilung (erster Buchstabe: **r** für *random*)

zur Verfügung. Zum Beispiel: **dbinom**, **pbinom**, **qbinom**, **rbinom**. Analog: **dpois**, **dgeom**, **dhyper**, **dnegbinom**. Für die Multinomialverteilung können nur Dichtefunktionen (**dmultinom**) und Zufallszahlen (**rmultinom**) berechnet werden.

### 3.5 Lineare stochastische Abhängigkeit

Sei  $(X, Y)$  ein bivariater Zufallsvektor. Zur Quantifizierung der linearen Abhängigkeit zwischen  $X$  und  $Y$  wird die Kovarianz und die Korrelation verwendet. Zur Berechnung benötigen wir

**Lemma 3.5.1.** Sei  $g: \mathbb{R}^2 \rightarrow \mathbb{R}$  eine Funktion. Dann gilt:

$$E(g(X, Y)) = \sum_{x, y} g(x, y) \cdot f_{X, Y}(x, y).$$

Insbesondere:

$$E(X \cdot Y) = \sum_{x, y} xy \cdot f_{X, Y}(x, y).$$

**Definition 3.5.2** (Kovarianz). Die Kovarianz von  $X$  und  $Y$  ist definiert durch

$$\begin{aligned} \text{Cov}(X, Y) &= E((X - E(X))(Y - E(Y))) \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

**Beachte:**  $\text{Cov}(X, X) = \text{Var}(X)$ .

**Definition 3.5.3** (Korrelation). Die Korrelation von  $X$  und  $Y$  ist definiert durch

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

wobei  $\text{Var}(X) > 0$  und  $\text{Var}(Y) > 0$  erfüllt sein muss.

Eine wichtige Eigenschaft der Korrelation beschreibt

**Lemma 3.5.4.** Für die Korrelation  $\rho = \rho(X, Y)$  gilt:

$$|\rho(X, Y)| \leq 1.$$

Zum Beweis benötigt man

**Satz 3.5.5** (Cauchy-Schwarz-Ungleichung; CAUCHY, 1789-1857, SCHWARZ, 1843-1921).  
Für zwei Zufallsvariablen  $X$  und  $Y$  mit  $E(X^2) < \infty$  und  $E(Y^2) < \infty$  gilt:

$$[E(XY)]^2 \leq E(X^2) E(Y^2).$$

*Beweis.* Sei o.B.d.A.  $E(X^2) > 0$ , da ansonsten  $P(X = 0) = P(XY = 0) = 1$ , und die obige Ungleichung wäre eine Gleichung. Mit  $a \in \mathbb{R}$  gilt

$$\begin{aligned} 0 &\leq E((aX - Y)^2) \\ &= E(a^2 X^2 - 2aXY + Y^2) \\ &= a^2 E(X^2) - 2a E(XY) + E(Y^2). \end{aligned}$$

Einsetzen von  $a = E(XY)/E(X^2)$  ergibt

$$\begin{aligned} 0 &\leq \frac{[E(XY)]^2}{E(X^2)} - 2 \frac{[E(XY)]^2}{E(X^2)} + E(Y^2) \\ &= E(Y^2) - \frac{[E(XY)]^2}{E(X^2)} \\ \Leftrightarrow [E(XY)]^2 &\leq E(X^2)E(Y^2). \end{aligned}$$

□

*Beweis zu Lemma 3.5.4.* Wendet man Satz 3.5.5 auf  $X - E(X)$  und  $Y - E(Y)$  an, so erhält man

$$[E((X - E(X))(Y - E(Y)))]^2 \leq E((X - E(X))^2) E((Y - E(Y))^2).$$

Division der linken durch die rechte Seite ergibt

$$\rho^2 = \frac{[\text{Cov}(X, Y)]^2}{\text{Var}(X) \text{Var}(Y)} \leq 1.$$

□

Man kann ferner zeigen, dass die Grenzen  $\rho = +1$  bzw.  $\rho = -1$  erreicht werden, wenn  $P(aX + bY = c) = 1$  für bestimmte  $a, b, c \in \mathbb{R}$ . Falls  $Y$  linear in  $X$  steigt bzw. fällt, ist die Korrelation  $+1$  bzw.  $-1$ .

**Beispiel 3.5.6.** Gegeben sei folgende gemeinsame Wahrscheinlichkeitsfunktion von zwei Zufallsvariablen  $X$  und  $Y$ :

$f_{X,Y}(x, y)$	$Y = -1$	$Y = 0$	$Y = 2$	$\sum$
$X = 1$	$\frac{1}{18}$	$\frac{3}{18}$	$\frac{2}{18}$	$\frac{6}{18}$
$X = 2$	$\frac{2}{18}$	$0$	$\frac{3}{18}$	$\frac{5}{18}$
$X = 3$	$0$	$\frac{4}{18}$	$\frac{3}{18}$	$\frac{7}{18}$
$\sum$	$\frac{3}{18}$	$\frac{7}{18}$	$\frac{8}{18}$	$1$

Nun sind folgende Fragen zu beantworten:

1. Sind  $X$  und  $Y$  unabhängig?

Gemäß Definition 3.2.1 muss nur ein Gegenbeispiel für die Gleichung

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) \text{ für alle } x, y$$

gefunden werden. Dies ist leicht, etwa mit  $X = 2$  und  $Y = -1$ , was zu

$$\frac{2}{18} \neq \frac{3}{18} \cdot \frac{5}{18}$$

führt.

2. Berechnen Sie  $\rho(X, Y)$ !

Gemäß Definition 3.5.3 gilt:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}. \quad (17)$$

Aus der obigen Verteilung erhält man folgende Erwartungswerte:

$$\begin{aligned} E(X) &= \frac{37}{18} & E(Y) &= \frac{13}{18} & E(XY) &= \frac{29}{18} \\ E(X^2) &= \frac{89}{18} & E(Y^2) &= \frac{35}{18}. \end{aligned}$$

Mit Hilfe des Verschiebungssatzes (Satz 3.3.6) sowie der Definition der Kovarianz (Definiton 3.5.2) erhält man

$$\text{Var}(X) = 0.719 \quad \text{Var}(Y) = 1.423 \quad \text{Cov}(XY) = 0.127.$$

Durch Einsetzen in (17) erhält man nun aus Definition 3.5.3:

$$\rho(X, Y) = 0.125.$$

### 3.6 Kovarianz- und Korrelationsmatrizen

**Satz 3.6.1** (Lineare Transformationen von Kovarianz und Korrelation). *Für zwei Zufallsvariablen  $X$  und  $Y$  gilt:*

1.  $\text{Cov}(aX + b, cY + d) = ac \cdot \text{Cov}(X, Y)$  für alle  $a, b, c, d \in \mathbb{R}$ ,
2.  $|\rho(aX + b, cY + d)| = |\rho(X, Y)|$  für alle  $a, b, c, d \in \mathbb{R}$ , wobei  $a, c \neq 0$ .

**Definition 3.6.2** (Kovarianzmatrix). Sei  $\mathbf{X} = (X_1, \dots, X_p)^T$  ein  $p$ -dimensionaler Zufallsvektor. Definiere  $\sigma_{ij} := \text{Cov}(X_i, X_j)$  für alle  $i$  und  $j$  (d.h. insbesondere  $\sigma_{ii} = \text{Var}(X_i)$  für alle  $i$ ). Die Matrix

$$\Sigma = \Sigma_{\mathbf{X}} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{pmatrix}$$

heißt *Kovarianzmatrix* von  $\mathbf{X}$ . Gebräuchlich ist auch die Schreibweise  $\sigma_{ii} = \sigma_i^2$ .

**Satz 3.6.3** (Eigenschaften der Kovarianzmatrix).

1.  $\Sigma$  ist symmetrisch, d.h.  $\sigma_{ij} = \sigma_{ji}$  für alle  $i$  und  $j$ .
2.  $\Sigma$  ist positiv semi-definit, d.h.  $\mathbf{a}^T \Sigma \mathbf{a} \geq 0$  für alle  $\mathbf{a} \in \mathbb{R}^p$ .
3. Sei  $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$  eine lineare Transformation von  $\mathbf{X}$  ( $\mathbf{Y} \in \mathbb{R}^{r \times 1}$ ,  $\mathbf{A} \in \mathbb{R}^{r \times p}$ ,  $\mathbf{X} \in \mathbb{R}^{p \times 1}$ ,  $\mathbf{b} \in \mathbb{R}^{r \times 1}$ ). Dann gilt

$$\begin{aligned} \mathbf{E}(\mathbf{Y}) &= \mathbf{A} \mathbf{E}(\mathbf{X}) + \mathbf{b} \\ \text{und} \quad \Sigma_{\mathbf{Y}} &= \mathbf{A} \Sigma_{\mathbf{X}} \mathbf{A}^T. \end{aligned}$$

4.  $\Sigma_{\mathbf{X}} = \mathbf{E}((\mathbf{X} - \mathbf{E}(\mathbf{X}))(\mathbf{X} - \mathbf{E}(\mathbf{X}))^T)$ .

**Beispiel 3.6.4.** Im zweidimensionalen Fall ergibt sich Satz 3.6.3.4 aus

$$\begin{aligned} & \mathbf{E} \left( \begin{pmatrix} X_1 - \mathbf{E}(X_1) \\ X_2 - \mathbf{E}(X_2) \end{pmatrix} \begin{pmatrix} X_1 - \mathbf{E}(X_1) \\ X_2 - \mathbf{E}(X_2) \end{pmatrix}^T \right) \\ &= \mathbf{E} \begin{pmatrix} [X_1 - \mathbf{E}(X_1)]^2 & [X_1 - \mathbf{E}(X_1)][X_2 - \mathbf{E}(X_2)] \\ [X_2 - \mathbf{E}(X_2)][X_1 - \mathbf{E}(X_1)] & [X_2 - \mathbf{E}(X_2)]^2 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{E}([X_1 - \mathbf{E}(X_1)]^2) & \mathbf{E}([X_1 - \mathbf{E}(X_1)][X_2 - \mathbf{E}(X_2)]) \\ \mathbf{E}([X_2 - \mathbf{E}(X_2)][X_1 - \mathbf{E}(X_1)]) & \mathbf{E}([X_2 - \mathbf{E}(X_2)]^2) \end{pmatrix} \\ &= \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \\ &= \Sigma_{\mathbf{X}}. \end{aligned}$$

Für  $\mathbf{X} \in \mathbb{R}^{p \times 1}$  und eine beliebige Matrix  $\mathbf{A} \in \mathbb{R}^{r \times p}$  gilt

$$\begin{aligned}\Sigma_{\mathbf{A}\mathbf{X}} &= \mathbb{E}([\mathbf{A}\mathbf{X} - \mathbb{E}(\mathbf{A}\mathbf{X})][\mathbf{A}\mathbf{X} - \mathbb{E}(\mathbf{A}\mathbf{X})]^T) \\ &= \mathbb{E}(\mathbf{A}[\mathbf{X} - \mathbb{E}(\mathbf{X})][\mathbf{X} - \mathbb{E}(\mathbf{X})]^T \mathbf{A}^T) \\ &= \mathbf{A}\Sigma_{\mathbf{X}}\mathbf{A}^T.\end{aligned}$$

Im Sonderfall  $\mathbf{A} = \mathbf{a}^T \in \mathbb{R}^p$  gilt wegen  $\mathbf{a}^T \mathbf{X} = \sum_{i=1}^p a_i X_i$

$$\text{Var}(\mathbf{a}^T \mathbf{X}) = \mathbf{a}^T \Sigma_{\mathbf{X}} \mathbf{a} \geq 0.$$

Dies ergibt sich auch ohne Wissen von Satz 3.6.3.2, da bekannt ist, dass die Varianz immer größer oder gleich null ist. Hieraus lässt sich dann wiederum die positive Semi-Definitheit von  $\Sigma_{\mathbf{X}}$  folgern.

Ein weiterer Spezialfall: Sei  $\mathbf{A} = (1, 1)$  und  $Y = \mathbf{A}\mathbf{X} = X_1 + X_2$ . Dann gilt:

$$\begin{aligned}\text{Var}(Y) &= \Sigma_Y \\ &= \mathbf{A}\Sigma_{\mathbf{X}}\mathbf{A}^T \\ &= (1, 1) \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \sigma_1^2 + \sigma_2^2 + 2\sigma_{12} \\ &= \text{Var}(X_1) + \text{Var}(X_2) + 2\text{Cov}(X_1, X_2).\end{aligned}$$

**Beispiel 3.6.5** (Kovarianzmatrix der Trinomialverteilung). Definiere den Zufallsvektor  $\mathbf{X} = (X_1, X_2, X_3)^T \sim M_3(n, \boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)^T)$ . Wie lautet die Kovarianz von  $X_i$  und  $X_j$ ,  $i, j = 1, 2, 3$  ?

Betrachte o.B.d.A.  $X_1$  und  $X_2$ :

$$\begin{aligned}\mathbb{E}(X_1 X_2) &= \sum_{\substack{\mathbf{x} = (x_1, x_2, x_3) \text{ mit} \\ x_1 + x_2 + x_3 = n}} x_1 x_2 \frac{n!}{x_1! x_2! x_3!} \pi_1^{x_1} \pi_2^{x_2} \pi_3^{x_3} \\ &= \sum_{\substack{\mathbf{x} = (x_1, x_2, x_3) \text{ mit} \\ x_1 + x_2 + x_3 = n}} \frac{n!}{(x_1 - 1)! (x_2 - 1)! x_3!} \pi_1^{x_1} \pi_2^{x_2} \pi_3^{x_3} \\ &= n(n-1)\pi_1\pi_2 \underbrace{\sum_{\substack{\mathbf{x} = (x_1, x_2, x_3) \text{ mit} \\ (x_1 - 1) + (x_2 - 1) + x_3 = n - 2}} \frac{(n-2)!}{(x_1 - 1)! (x_2 - 1)! x_3!} \pi_1^{x_1 - 1} \pi_2^{x_2 - 1} \pi_3^{x_3}}_{=1} \\ &= n(n-1)\pi_1\pi_2.\end{aligned}$$

Damit ergibt sich:

$$\begin{aligned}\text{Cov}(X_1, X_2) &= \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1) \mathbb{E}(X_2) \\ &= n(n-1)\pi_1\pi_2 - n\pi_1 \cdot n\pi_2 \\ &= -n\pi_1\pi_2.\end{aligned}$$

Allgemein gilt also

$$\text{Cov}(X_i, X_j) = \begin{cases} -n\pi_i\pi_j & \text{für } i \neq j \\ n\pi_i(1 - \pi_i) & \text{für } i = j \end{cases} \quad (18)$$

bzw.

$$\Sigma_{\mathbf{X}} = n(\text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^T), \text{ hier (bei der Trinomialverteilung) } \boldsymbol{\pi} = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix}. \quad (19)$$

Die Korrelation zwischen  $X_i$  und  $X_j$  ergibt sich zu

$$\rho(X_i, X_j) = \begin{cases} \frac{-n\pi_i\pi_j}{\sqrt{n\pi_i(1-\pi_i) \cdot n\pi_j(1-\pi_j)}} = -\sqrt{\frac{\pi_i}{1-\pi_i} \cdot \frac{\pi_j}{1-\pi_j}} & \text{für } i \neq j \\ 1 & \text{für } i = j. \end{cases} \quad (20)$$

**Bemerkung:** Man zeigt leicht, dass die Formeln (18), (19) und (20) auch allgemein für die Multinomialverteilung gelten.

Zuletzt wird nun noch die Korrelationsmatrix definiert:

**Definition 3.6.6** (Korrelationsmatrix). Sei  $\mathbf{X} = (X_1, \dots, X_p)^T$ , dann ist

$$\mathbf{R} = \mathbf{R}_{\mathbf{X}} = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \rho_{23} & \dots & \rho_{2p} \\ \rho_{31} & \rho_{32} & 1 & \dots & \rho_{3p} \\ \vdots & \vdots & & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \rho_{p3} & \dots & 1 \end{pmatrix}$$

die *Korrelationsmatrix* von  $\mathbf{X}$ , wobei  $\rho_{ij} = \rho(X_i, X_j)$ .

**Satz 3.6.7.** Für die Korrelationsmatrix eines Zufallsvektors  $\mathbf{X} \in \mathbb{R}^p$  gilt:

$$\mathbf{R} = \mathbf{P}\Sigma_{\mathbf{X}}\mathbf{P}^T,$$

wobei  $\mathbf{P} = \text{diag}((\text{Var}(X_1))^{-\frac{1}{2}}, \dots, (\text{Var}(X_p))^{-\frac{1}{2}})$ .

### 3.7 Bedingte Verteilungen und bedingte Erwartungen

Seien  $X$  und  $Y$  zwei Zufallsvariablen auf  $(\Omega, \mathcal{F}, P)$ .

**Definition 3.7.1.** Die *bedingte Verteilungsfunktion* von  $Y$  gegeben  $X = x$  ist definiert als

$$F_{Y|X}(y|x) = P(Y \leq y | X = x)$$

für alle  $x \in \mathbb{R}$  mit  $P(X = x) = f_X(x) > 0$ . Die zugehörige *bedingte Wahrscheinlichkeitsfunktion* lautet

$$\begin{aligned} f_{Y|X}(y|x) &= P(Y = y | X = x) \\ &= \frac{P(X = x, Y = y)}{P(X = x)}. \end{aligned}$$

Diese Aussage ist äquivalent zu

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

Man sieht sofort, dass folgender Satz gilt:

**Satz 3.7.2.** *X und Y sind genau dann unabhängig, wenn  $f_{Y|X}(y|x) = f_Y(y)$  für alle  $x \in \mathcal{T}_X$  und  $y \in \mathcal{T}_Y$  gilt.*

**Beispiel 3.7.3.** Sei  $\mathbf{X} = (X_1, X_2, X_3)^T \sim M_3(n, (\pi_1, \pi_2, \pi_3)^T)$ . Wie lautet die bedingte Verteilungsfunktion von  $X_1$  gegeben  $X_3 = x_3$ ?

Ausgehend von Definition 3.7.1 erhält man

$$\begin{aligned} f_{X_1, X_2|X_3}(x_1, x_2|x_3) &= \frac{f_{X_1, X_2, X_3}(x_1, x_2, x_3)}{f_{X_3}(x_3)} \\ &= \frac{\frac{n!}{x_1!x_2!x_3!} \pi_1^{x_1} \pi_2^{x_2} \pi_3^{x_3}}{\frac{n!}{x_3!(n-x_3)!} \pi_3^{x_3} (1-\pi_3)^{(n-x_3)}} \\ &\stackrel{n-x_3=x_1+x_2}{=} \frac{(n-x_3)!}{x_1!x_2!} \left(\frac{\pi_1}{1-\pi_3}\right)^{x_1} \left(\frac{\pi_2}{1-\pi_3}\right)^{x_2} \end{aligned}$$

Man erkennt leicht, dass dies äquivalent ist zu

$$X_1|\{X_3 = x_3\} \sim B\left(n - x_3, \frac{\pi_1}{1 - \pi_3}\right).$$

Nun lässt sich der bedingte Erwartungswert von  $Y$  gegeben  $X = x$  betrachten, der wie folgt definiert ist:

**Definition 3.7.4** (Bedingter Erwartungswert). Die reelle Zahl

$$\psi(x) = E(Y|X = x) = \sum_{y \in \mathcal{T}_Y} y \cdot f_{Y|X}(y|x) \quad \text{für alle } x \in \mathcal{T}_X$$

heißt *bedingter Erwartungswert von Y gegeben X = x*, falls diese Summe absolut konvergiert.

**Beachte:** Im Folgenden gehen wir davon aus, dass die beteiligten Erwartungswerte existieren, auch wenn dies nicht mehr explizit erwähnt wird!

Die Idee der bedingten Erwartung ist es,  $\psi$  als Funktion der Zufallsvariablen  $X$  zu betrachten.

**Definition 3.7.5** (Bedingte Erwartung). Sei  $\psi(x) = E(Y|X = x)$ . Dann nennt man die Zufallsvariable  $\psi(X) = E(Y|X)$  die *bedingte Erwartung* von  $Y$  gegeben  $X$ .

**Beachte:** Es sei hier nochmals hervorgehoben, dass es sich bei der bedingten Erwartung um eine Zufallsvariable handelt, bei dem bedingten Erwartungswert um eine reelle Zahl! Diese begriffliche Unterscheidung wird in der Literatur jedoch nicht durchgehend vorgenommen. Insbesondere im Englischen werden beide Bezeichnungen mit *conditional expectation* übersetzt.

Folgender Satz ist als sehr wichtig anzusehen.

**Satz 3.7.6** (Satz vom iterierten Erwartungswert). Für die bedingte Erwartung  $\psi(X) = E(Y|X)$  gilt

$$E(\psi(X)) = E(Y)$$

oder, in anderer Schreibweise,

$$E(E(Y|X)) = E(Y).$$

*Beweis.* Es ist

$$\begin{aligned} E(\psi(X)) &= \sum_x \psi(x) f_X(x) \\ &= \sum_x \sum_y y f_{Y|X}(y|x) f_X(x) \\ &= \sum_x \sum_y y f_{X,Y}(x, y) \\ &= \sum_y y \sum_x f_{X,Y}(x, y) \\ &= \sum_y y f_Y(y) \\ &= E(Y). \end{aligned}$$

□

**Beispiel 3.7.7.** Eine Henne legt  $N \sim \text{Po}(\lambda)$  Eier, von denen sie jedes unabhängig mit der Wahrscheinlichkeit  $\pi$  ausbrütet. Sei  $K$  die Anzahl der Küken. Berechne  $E(K|N)$ ,  $E(K)$  und  $E(N|K)$ .

Da  $K|\{N = n\} \sim B(n, \pi)$ , folgt sofort  $E(K|N = n) = n\pi$  und  $E(K|N) = N\pi = \psi(N)$ . Damit:

$$E(K) = E(E(K|N)) = E(N\pi) = \pi E(N) = \pi\lambda.$$

Um  $E(N|K)$  zu berechnen, muss man zunächst die entsprechende bedingte Wahrscheinlichkeitsfunktion berechnen:

$$\begin{aligned} f_{N|K}(n|k) &= \frac{f_{N,K}(n, k)}{f_K(k)} \\ &= \frac{f_{K|N}(k|n) f_N(n)}{f_K(k)}. \end{aligned}$$

Nun gilt für  $f_K(k)$  (vgl. Beispiel 3.2.2)

$$f_K(k) = \frac{(\lambda\pi)^k}{k!} \exp(-\lambda\pi) \Leftrightarrow K \sim \text{Po}(\lambda\pi).$$

Somit ergibt sich für  $n \geq k$ :

$$\begin{aligned} f_{N|K}(n|k) &= \frac{\binom{n}{k} \pi^k (1-\pi)^{n-k} \cdot \frac{\lambda^n}{n!} \exp(-\lambda)}{\frac{(\lambda\pi)^k}{k!} \exp(-\lambda\pi)} \\ &= \frac{[\lambda(1-\pi)]^{n-k}}{(n-k)!} \exp(-\lambda(1-\pi)). \end{aligned}$$

Also ist  $(N - k)|\{K = k\} \sim \text{Po}(\lambda(1 - \pi))$ . Bei  $N|\{K = k\} \sim k + \text{Po}(\lambda(1 - \pi))$  handelt es sich um eine *um  $k$  verschobene Poissonverteilung*. Daraus ergibt sich dann

$$\begin{aligned} \mathbf{E}(N|K = k) &= k + \lambda(1 - \pi), \\ \mathbf{E}(N|K) &= K + \lambda(1 - \pi). \end{aligned}$$

Zur Überprüfung bietet sich der Satz vom iterierten Erwartungswert (Satz 3.7.6) an:

$$\mathbf{E}(N) = \mathbf{E}(\mathbf{E}(N|K)) = \mathbf{E}(K + \lambda(1 - \pi)) = \lambda\pi + \lambda(1 - \pi) = \lambda.$$

Das Ergebnis stimmt mit  $N \sim \text{Po}(\lambda)$  überein.

Einige Eigenschaften der bedingten Erwartung seien im Folgenden zusammengefasst.

**Satz 3.7.8** (Eigenschaften der bedingten Erwartung). *Seien  $X, Y$  und  $Z$  Zufallsvariablen. Dann gilt*

1.  $\mathbf{E}((aY + bZ)|X) = a\mathbf{E}(Y|X) + b\mathbf{E}(Z|X)$  für alle  $a, b \in \mathbb{R}$ .
2. Für  $Y \geq 0$  gilt  $\mathbf{E}(Y|X) \geq 0$ .
3.  $\mathbf{E}(1|X) = 1$ .
4. Sind  $X, Y$  unabhängig, so gilt  $\mathbf{E}(Y|X) = \mathbf{E}(Y)$  und natürlich auch  $\mathbf{E}(X|Y) = \mathbf{E}(X)$ .
5.  $\mathbf{E}(Y \cdot g(X)|X) = g(X)\mathbf{E}(Y|X)$  für jede Funktion  $g: \mathbb{R} \rightarrow \mathbb{R}$  ("pull through property").
6.  $\mathbf{E}(\mathbf{E}(Y|X, Z)|X) = \mathbf{E}(Y|X) = \mathbf{E}(\mathbf{E}(Y|X)|X, Z)$  ("tower property").

**Satz 3.7.9.** *Die bedingte Erwartung  $\psi(X) = \mathbf{E}(Y|X)$  erfüllt*

$$\mathbf{E}(\psi(X) \cdot g(X)) = \mathbf{E}(Y \cdot g(X))$$

für jede Funktion  $g$ , für die beide Erwartungswerte existieren.

**Bemerkung:** Für  $g(X) = 1$  erhält man Satz 3.7.6.

**Definition 3.7.10** (Bedingte Varianz). Die Zahl

$$\text{Var}(Y|X = x) = \mathbf{E}([Y - \mathbf{E}(Y|X = x)]^2|X = x) \quad \text{für alle } x \in \mathcal{T}_X$$

heißt *bedingte Varianz*.

Ähnlich der bedingten Erwartung ist  $\text{Var}(Y|X)$  eine Zufallsvariable in  $X$  und keine feste Zahl.

**Satz 3.7.11** (Varianzzerlegungssatz). *Für zwei Zufallsvariablen  $X$  und  $Y$  gilt:*

$$\text{Var}(Y) = \mathbf{E}(\text{Var}(Y|X)) + \text{Var}(\mathbf{E}(Y|X)).$$

*Beweis.*

$$\begin{aligned}
 \text{Var}(Y) &= \text{E}([Y - \text{E}(Y)]^2) \\
 &= \text{E}(\text{E}([Y - \text{E}(Y)]^2|X)) \\
 &= \text{E}(\text{E}([Y - \text{E}(Y|X) + \text{E}(Y|X) - \text{E}(Y)]^2|X)) \\
 &= \text{E}(\text{E}([Y - \text{E}(Y|X)]^2|X)) + \text{E}(\text{E}([\text{E}(Y|X) - \text{E}(Y)]^2|X)) + C \\
 &= \text{E}(\text{Var}(Y|X)) + \text{E}([\text{E}(Y|X) - \text{E}(Y)]^2) + C \\
 &= \text{E}(\text{Var}(Y|X)) + \text{E}([\text{E}(Y|X) - \text{E}(\text{E}(Y|X))]^2) + C \\
 &= \text{E}(\text{Var}(Y|X)) + \text{Var}(\text{E}(Y|X)) + C.
 \end{aligned}$$

Nun bleibt zu zeigen, dass  $C = 0$ :

$$\begin{aligned}
 C &= 2 \text{E}(\text{E}([Y - \text{E}(Y|X)] \underbrace{[\text{E}(Y|X) - \text{E}(Y)]}_{=g(X); \text{ nun Satz 3.7.8.5.}} |X)) \\
 &= 2 \text{E}([\text{E}(Y|X) - \text{E}(Y)] \underbrace{\text{E}([Y - \text{E}(Y|X)]|X)}_{=D}) \\
 D &= \text{E}(Y|X) - \text{E}(\text{E}(Y|X)|X) \\
 &= \text{E}(Y|X) - \text{E}(Y|X) \\
 &= 0 \\
 \Rightarrow C &= 0.
 \end{aligned}$$

□

**Interpretation des Varianzzerlegungssatzes.** Nach Satz 3.7.11 kann die Varianz einer Zufallsvariable  $Y$  in zwei Summanden zerlegt werden. Der erste,  $\text{E}(\text{Var}(Y|X))$ , misst die mittlere Varianz von  $Y$  "innerhalb" der einzelnen Ausprägungen einer weiteren Zufallsvariable  $X$ . Der zweite Summand,  $\text{Var}(\text{E}(Y|X))$ , gibt an, wie stark  $\text{E}(Y|X)$  um  $\text{E}(Y)$  variiert, denn nach dem Satz vom iterierten Erwartungswert gilt

$$\begin{aligned}
 \text{Var}(\text{E}(Y|X)) &= \text{E}([\text{E}(Y|X) - \text{E}(\text{E}(Y|X))]^2) \\
 &= \text{E}([\text{E}(Y|X) - \text{E}(Y)]^2).
 \end{aligned}$$

Dies ist also die Varianz "zwischen" den Gruppen, die durch die Ausprägungen von  $X$  definiert werden.

**Beispiel 3.7.12.** Seien  $X$  und  $Y$  zwei Zufallsvariablen mit  $X|Y \sim \text{Po}(Y)$  und  $Y \sim \text{Po}(\lambda)$ ,  $\lambda > 0$ . Dann ist

$$\text{E}(X) = \text{E}(\text{E}(X|Y)) = \text{E}(Y) = \lambda$$

und nach dem Varianzzerlegungssatz

$$\text{Var}(X) = \text{E}(\text{Var}(X|Y)) + \text{Var}(\text{E}(X|Y)) = \text{E}(Y) + \text{Var}(Y) = \lambda + \lambda = 2\lambda.$$

Die direkte Berechnung des Erwartungswertes und der Varianz von  $X$  via

1.  $f_{X,Y}(x, y) = f_{X|Y}(x|y)f_Y(y)$
2.  $f_X(x) = \sum_{y \in \mathcal{T}_Y} f_{X,Y}(x, y)$

$$3. \ E(X) = \sum_{x \in \mathcal{T}_X} x f_X(x), \ E(X^2) = \sum_{x \in \mathcal{T}_X} x^2 f_X(x), \ \text{Var}(X) = E(X^2) - [E(X)]^2$$

wäre deutlich aufwändiger gewesen.

**Beispiel 3.7.13** (Fortsetzung von Beispiel 3.7.7). Zu Erinnerung:  $K$  ist die Anzahl der Küken,  $N$  die Anzahl der Eier,  $N \sim \text{Po}(\lambda)$  und  $K|N \sim B(N, \pi)$ . Nun berechnet man leicht:

$$\begin{aligned} \text{Var}(K) &= E(\text{Var}(K|N)) + \text{Var}(E(K|N)) \\ &= E(N\pi(1-\pi)) + \text{Var}(N\pi) \\ &= \pi(1-\pi)\lambda + \pi^2\lambda \\ &= \lambda\pi. \end{aligned}$$

Das Ergebnis überrascht nicht, da  $K \sim \text{Po}(\lambda\pi)$ . Das gleiche gilt für

$$\begin{aligned} \text{Var}(N) &= E(\text{Var}(N|K)) + \text{Var}(E(N|K)) \\ &= E((1-\pi)\lambda) + \text{Var}((1-\pi)\lambda + K) \\ &= (1-\pi)\lambda + \lambda\pi \\ &= \lambda \end{aligned}$$

in Übereinstimmung mit  $N \sim \text{Po}(\lambda)$ .

**Beispiel 3.7.14** (Verzweigungsprozesse, Verbreitung von Epidemien). Wir beobachten den Verlauf einer Epidemie zu diskreten Zeitpunkten  $t = 0, 1, 2, \dots$ . Die Zufallsvariable  $X_t$  bezeichne die Anzahl der Infizierten zum Zeitpunkt  $t$ . Es sei

$$\begin{aligned} X_0 &= 1 \\ X_t|X_{t-1} &\sim \text{Po}(\lambda \cdot X_{t-1}) \quad \text{für ein } \lambda > 0, \end{aligned}$$

also

$$E(X_t|X_{t-1}) = \lambda X_{t-1} \quad \text{und} \quad \text{Var}(X_t|X_{t-1}) = \lambda X_{t-1}.$$

Dann gilt für  $t = 1, 2, \dots$ :

$$(1) \ E(X_t) = \lambda^t \quad \text{und} \quad (2) \ \text{Var}(X_t) = \begin{cases} t & \text{für } \lambda = 1, \\ \frac{\lambda^t(\lambda^t - 1)}{\lambda - 1} & \text{für } \lambda \neq 1. \end{cases}$$

*Beweis durch Induktion.*

1. Induktionsanfang:

$$E(X_1) = E(E(X_1|X_0)) = E(\lambda) = \lambda.$$

Induktionsvoraussetzung: Die Behauptung gelte für  $t - 1$ .

Induktionsschluss:

$$E(X_t) = E(E(X_t|X_{t-1})) = E(\lambda \cdot X_{t-1}) = \lambda E(X_{t-1}) \stackrel{\text{I.V.}}{=} \lambda \cdot \lambda^{t-1} = \lambda^t.$$

2. Induktionsanfang:

$$\text{Var}(X_1) = \text{E}(\text{Var}(X_1|X_0)) + \text{Var}(\text{E}(X_1|X_0)) = \text{E}(\lambda) + \text{Var}(\lambda) = \lambda + 0 = \lambda,$$

d.h. die Formel ist korrekt für beliebiges  $\lambda > 0$  und  $t = 1$ .

Induktionsvoraussetzung: Die Behauptung gelte für  $t - 1$ .

Induktionsschluss:

$$\begin{aligned}\text{Var}(X_t) &= \text{E}(\text{Var}(X_t|X_{t-1})) + \text{Var}(\text{E}(X_t|X_{t-1})) \\ &= \text{E}(\lambda X_{t-1}) + \text{Var}(\lambda X_{t-1}) \\ &= \lambda^t + \lambda^2 \text{Var}(X_{t-1}).\end{aligned}$$

Hieraus folgt mit der Induktionsvoraussetzung

*Fall 1:*  $\lambda = 1$ .

$$\text{Var}(X_t) = 1^t + 1^2 \cdot (t - 1) = t.$$

*Fall 2:*  $\lambda \neq 1$ .

$$\begin{aligned}\text{Var}(X_t) &= \lambda^t + \lambda^2 \frac{\lambda^{t-1}(\lambda^{t-1} - 1)}{\lambda - 1} \\ &= \lambda^t + \frac{\lambda^{t+1}(\lambda^{t-1} - 1)}{\lambda - 1} \\ &= \frac{\lambda^{t+1} - \lambda^t + \lambda^{t+1}(\lambda^{t-1} - 1)}{\lambda - 1} \\ &= \frac{\lambda^{t+1} - \lambda^t + \lambda^{2t} - \lambda^{t+1}}{\lambda - 1} \\ &= \frac{\lambda^t(\lambda^t - 1)}{\lambda - 1}.\end{aligned}$$

□

**Beachte:** Die spezielle Form der Poissonverteilung geht in den Beweis nicht direkt ein, nur ihre ersten beiden Momente; es wäre auch  $\text{Var}(Z_t|Z_{t-1}) \neq \text{E}(Z_t|Z_{t-1})$  möglich, dann ergibt sich eine leicht veränderte Formel für  $\text{Var}(Z_t)$ . Dieser Beweis wird in Kapitel 5 noch einmal über wahrscheinlichkeitserzeugende Funktionen geführt.

**Beachte:** Es gilt das sogenannte *Schwellenwerttheorem*:

- Wenn  $\lambda < 1$ , dann  $\text{E}(X_t) \rightarrow 0$  und  $\text{Var}(X_t) \rightarrow 0$  für  $t \rightarrow \infty$ .
- Wenn  $\lambda = 1$ , dann  $\text{E}(X_t) \rightarrow \lambda$  und  $\text{Var}(X_t) \rightarrow \infty$  für  $t \rightarrow \infty$ .
- Wenn  $\lambda > 1$ , dann  $\text{E}(X_t) \rightarrow \infty$  und  $\text{Var}(X_t) \rightarrow \infty$  für  $t \rightarrow \infty$ .

Die *Basisreproduktionszahl* (engl. *basic reproduction number*)  $\lambda$  gibt an, wie viele Personen ein Infizierter im Schnitt ansteckt. Sie ist eine zentrale Größe in der Infektionsepidemiologie.

### 3.8 Summen von Zufallsvariablen

**Satz 3.8.1.** Sei  $(X, Y)$  ein Zufallsvektor mit gemeinsamer Wahrscheinlichkeitsfunktion  $f_{X,Y}(x, y)$ , dann gilt für  $Z = X + Y$ :

$$f_Z(z) = P(X + Y = z) = \sum_x f_{X,Y}(x, z - x).$$

*Beweis.* Da

$$\{x + y = z\} = \bigcup_x (\{X = x\} \cap \{Y = z - x\})$$

eine disjunkte Vereinigung von Ereignissen darstellt, folgt

$$P(X + Y = Z) = \sum_x P(X = x, Y = z - x) = \sum_x f_{X,Y}(x, z - x).$$

□

**Lemma 3.8.2** (Faltungssatz). Sind  $X$  und  $Y$  unabhängige Zufallsvariablen, so gilt:

$$f_Z(z) = P(X + Y = z) = \sum_x f_X(x) f_Y(z - x) = \sum_y f_Y(y) f_X(z - y).$$

Eine Summe der Form  $\sum_k u_k v_{n-k}$  nennt man eine Faltung. Unter Unabhängigkeit ist also die Wahrscheinlichkeitsfunktion  $f_{X+Y}(z) = P(X + Y = z)$  die Faltung von  $f_X(x)$  und  $f_Y(y)$ . Man schreibt dafür auch  $f_{X+Y} = f_X \star f_Y$ .

**Beispiel 3.8.3.** Seien  $X \sim \text{Po}(\lambda)$  und  $Y \sim \text{Po}(\mu)$  unabhängige Zufallsvariablen mit  $\lambda, \mu > 0$ . Dann gilt für  $Z = X + Y$ :

$$\begin{aligned} f_Z(z) &= \sum_{x=0}^z f_X(x) f_Y(z-x) \\ &= \sum_{x=0}^z \frac{\lambda^x}{x!} \exp(-\lambda) \cdot \frac{\mu^{z-x}}{(z-x)!} \exp(-\mu) \\ &= \sum_{x=0}^z \frac{1}{z!} \binom{z}{x} \lambda^x \mu^{z-x} \exp(-(\lambda + \mu)) \\ &= \sum_{x=0}^z \frac{(\lambda + \mu)^z}{z!} \binom{z}{x} \left(\frac{\lambda}{\lambda + \mu}\right)^x \left(\frac{\mu}{\lambda + \mu}\right)^{z-x} \exp(-(\lambda + \mu)) \\ &= \frac{(\lambda + \mu)^z}{z!} \exp(-(\lambda + \mu)) \underbrace{\sum_{x=0}^z \binom{z}{x} \left(\frac{\lambda}{\lambda + \mu}\right)^x \left(\frac{\mu}{\lambda + \mu}\right)^{z-x}}_{=1} \\ &= \frac{(\lambda + \mu)^z}{z!} \exp(-(\lambda + \mu)), \end{aligned}$$

also  $X + Y = Z \sim \text{Po}(\lambda + \mu)$ .

**Beispiel 3.8.4.** Seien  $X_1$  und  $X_2$  unabhängig und jeweils geometrisch verteilt mit Parameter  $\pi$ .

$$X_i \sim G(\pi) \iff f_{X_i}(x) = (1 - \pi)^{x-1} \pi \quad \text{für } i = 1, 2.$$

Gesucht ist nun die Verteilung von  $Z = X_1 + X_2$ .  $Z$  beschreibt die Anzahl der Versuche, bis zum zweiten Mal das günstige Ereignis eintritt.  $Z$  wird mit dem Faltungssatz (Satz 3.8.2) berechnet:

$$\begin{aligned} f_Z(z) &= \sum_{x=1}^{\infty} f_{X_1}(x) f_{X_2}(z-x) \\ &= \sum_{x=1}^{z-1} \pi(1-\pi)^{x-1} \pi(1-\pi)^{z-x-1} \\ &= \sum_{x=1}^{z-1} \pi^2(1-\pi)^{z-2} \\ &= (z-1)\pi^2(1-\pi)^{z-2} \quad \text{für alle } z = 2, 3, \dots \end{aligned}$$

**Bemerkung:** Dies ist eine spezielle Form der *negativen Binomialverteilung*. Man schreibt auch  $Z \sim \text{NB}(r=2, \pi)$ .

Allgemein beschreibt die negative Binomialverteilung mit Parametern  $r \in \mathbb{N}$  und  $\pi \in [0, 1]$  die Anzahl  $Z$  der Versuche bis zum (einschließlich)  $r$ -ten Eintreten des günstigen Ereignisses. Die Dichtefunktion von  $Z \sim \text{NB}(r, \pi)$  lautet dann

$$f_Z(z) = \binom{z-1}{r-1} \pi^r (1-\pi)^{z-r} \quad \text{für } z = r, r+1, \dots$$

**Übungsaufgabe:** Seien  $X$  und  $X'$  unabhängige Poissonverteilte Zufallsvariablen mit Parameter  $\lambda > 0$ . Wie lautet die Wahrscheinlichkeitsfunktion von  $Z = X - X'$ ?

## 4 Stetige Zufallsvariablen

### 4.1 Dichtefunktion

Zur Wiederholung sei der Begriff der Stetigkeit für Zufallsvariablen nochmals erläutert.

**Definition 4.1.1** (stetige Zufallsvariable). Eine Zufallsvariable  $X$  heißt *stetig*, falls die Verteilungsfunktion  $F_X(x)$  für alle  $x \in \mathbb{R}$  in der Form

$$F_X(x) = \int_{-\infty}^x f_X(u) du \quad (21)$$

für eine integrierbare Funktion  $f_X : \mathbb{R} \rightarrow [0, \infty)$  dargestellt werden kann.  $f_X$  heißt *Dichtefunktion* von  $X$ .

**Bemerkungen:**

1.  $f_X(x)$  ist nicht eindeutig.
2. Ist  $F(x)$  differenzierbar, so setzt man typischerweise  $f(x) = F'(x)$ . Dann nennt man  $\mathcal{T}_X = \{x : f(x) > 0\}$  den *Träger* der Zufallsvariable  $X$ .
3.  $f(x)$  ist *keine* Wahrscheinlichkeit. Jedoch kann man

$$f(x)dx \approx P(x < X \leq x + dx) = F(x + dx) - F(x)$$

bei genügend kleinem  $dx$  als solche interpretieren.

4. Aus (21) folgt  $P(a < X \leq b) = \int_a^b f_X(x)dx$ , was verallgemeinert werden kann zu

$$P(X \in B) = \int_B f_X(x)dx$$

wobei die Menge  $B$  gewisse Regularitätsbedingungen erfüllen muss.

Nach der letzten Bemerkung stellt sich die Frage, für welche Mengen  $B$  dies genau gilt.

**Definition 4.1.2.** Sei  $\mathcal{J}$  die Menge aller offenen Intervalle in  $\mathbb{R}$ . Dann ist  $\mathcal{B} = \sigma(\mathcal{J})$  die von  $\mathcal{J}$  erzeugte  $\sigma$ -Algebra (vgl. Bemerkung nach Lemma 1.6.1). Man nennt  $\mathcal{B}$  die *Borelsche  $\sigma$ -Algebra* und ihre Elemente  $B \in \mathcal{B}$  *Borelmengen* (BOREL, 1871-1956). Für jede Borelmenge  $B$  gilt obige Bemerkung 4, und mit dem Wahrscheinlichkeitsmaß  $P(B) = P(X \in B) = \int_B f_X(x)dx$  folgt, dass  $(\mathbb{R}, \mathcal{B}, P)$  ein Wahrscheinlichkeitsraum ist.

**Lemma 4.1.3.** *Hat  $X$  Dichtefunktion  $f_X(x)$ , so gilt*

1.  $\int_{-\infty}^{\infty} f_X(x)dx = 1$ ,
2.  $P(X = x) = 0$  für alle  $x \in \mathbb{R}$ ,
3.  $P(a < X \leq b) = \int_a^b f_X(x)dx$ .

## 4.2 Unabhängigkeit

Bei stetigen Zufallsvariablen ist eine Definition über die Unabhängigkeit der Ereignisse  $\{X = x\}$  und  $\{Y = y\}$  — anders als bei diskreten Zufallsvariablen — nicht mehr möglich, da diese Ereignisse die Wahrscheinlichkeit null haben. Daher hilft man sich mit

**Definition 4.2.1** (Unabhängigkeit von stetigen Zufallsvariablen). Zwei stetige Zufallsvariablen  $X$  und  $Y$  heißen *unabhängig*, falls  $\{X \leq x\}$  und  $\{Y \leq y\}$  für alle  $x, y \in \mathbb{R}$  unabhängige Ereignisse sind.

**Bemerkung:** So könnte man auch die Unabhängigkeit von diskreten Zufallsvariablen definieren.

Betrachte nun  $g(X)$  und  $h(Y)$ , wobei  $g$  und  $h$  beliebige "reguläre" Funktionen sind. Dann sind  $g(X)$  und  $h(X)$  Zufallsvariablen, und es gilt:

**Satz 4.2.2.** *Sind  $X$  und  $Y$  unabhängig, so auch  $g(X)$  und  $h(Y)$ .*

**Bemerkung:** Die Unabhängigkeit zweier Zufallsvariablen wird häufig über folgende Formeln beschrieben:

$$\begin{aligned}F_{X,Y}(x, y) &= F_X(x)F_Y(y), \\f_{X,Y}(x, y) &= f_X(x)f_Y(y).\end{aligned}$$

## 4.3 Erwartungswert

**Definition 4.3.1** (Erwartungswert). Der *Erwartungswert* einer stetigen Zufallsvariable  $X$  mit der Dichtefunktion  $f_X(x)$  ist gegeben durch

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx,$$

falls  $x f_X(x)$  absolut integrierbar ist, d.h. wenn  $\int_{-\infty}^{\infty} |x f_X(x)| dx = \int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$ .

**Satz 4.3.2.** *Sind  $X$  und  $g(X)$  stetige Zufallsvariablen, so gilt:*

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx,$$

*falls der Integrand absolut integrierbar ist.*

**Lemma 4.3.3.** *Besitzt  $X$  Verteilungsfunktion  $F_X(x)$ , Dichtefunktion  $f_X(x)$  und ist  $f_X(x) = 0$  für  $x < 0$ , so gilt*

$$E(X) = \int_0^{\infty} [1 - F_X(x)] dx.$$

*Beweis.*

$$\begin{aligned}\int_0^\infty [1 - F_X(x)] dx &= \int_0^\infty P(X > x) dx \\ &= \int_0^\infty \int_{y=x}^\infty f_X(y) dy dx \\ &= \int_{y=0}^\infty \int_{x=0}^y f_X(y) dx dy \\ &= \int_{y=0}^\infty y f_X(y) dy \\ &= E(X).\end{aligned}$$

□

**Beispiel 4.3.4.** Sei die Dichte einer Zufallsvariablen  $X$  gegeben durch

$$f_X(x) = \begin{cases} \frac{1}{2\pi} & \text{für } 0 \leq x \leq 2\pi, \\ 0 & \text{sonst.} \end{cases}$$

Als Erwartungswert erhalten wir

$$\begin{aligned}E(X) &= \int_0^{2\pi} x \cdot \frac{1}{2\pi} dx \\ &= \frac{1}{2\pi} \left[ \frac{x^2}{2} \right]_0^{2\pi} \\ &= \pi.\end{aligned}$$

Betrachte nun  $Y = X^2$ . Dann ergibt sich:

$$\begin{aligned}E(Y) &= \int_0^{2\pi} x^2 \cdot \frac{1}{2\pi} dx \\ &= \frac{1}{2\pi} \left[ \frac{x^3}{3} \right]_0^{2\pi} \\ &= \frac{4}{3} \pi^2.\end{aligned}$$

Alternativ kann man dies direkt über die Dichte

$$f_Y(y) \begin{cases} \frac{1}{4\pi} y^{-\frac{1}{2}} & \text{für } 0 \leq y \leq 4\pi^2, \\ 0 & \text{sonst} \end{cases}$$

berechnen. Siehe hierzu den später folgenden Satz über Transformationen von Dichten (Satz 4.8.3).

Im Fall von stetigen Zufallsvariablen gelten nun die gleichen Rechenregeln für Varianzen, Kovarianzen, Korrelationen, Momente etc. wie bei diskreten Zufallsvariablen.

## 4.4 Beispiele für stetige Zufallsvariablen

### 4.4.1 Stetige Gleichverteilung

*Notation:*  $X \sim U(a, b)$

*Parameter:*  $a, b \in \mathbb{R}$ , wobei  $a < b$

*Träger:*  $\mathcal{T}_X = [a, b] \subset \mathbb{R}$

*Dichtefunktion:*  $f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b, \\ 0 & \text{sonst} \end{cases}$

*Erwartungswert:*  $E(X) = \frac{b+a}{2}$

*Varianz* (\*):  $\text{Var}(X) = \frac{(b-a)^2}{12}$

(\*) ergibt sich aus  $E(X^2) = \frac{b^3-a^3}{3(b-a)}$ .

### 4.4.2 Exponentialverteilung

*Notation:*  $X \sim \text{Exp}(\lambda)$

*Parameter:*  $\lambda > 0$

*Träger:*  $\mathcal{T}_X = \mathbb{R}_0^+$

*Dichtefunktion:*  $f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{für } x \geq 0, \\ 0 & \text{sonst} \end{cases}$

*Erwartungswert* (\*):  $E(X) = \frac{1}{\lambda}$

*Varianz:*  $\text{Var}(X) = \frac{1}{\lambda^2}$

(\*): Es ist  $F(x) = 1 - \exp(-\lambda x)$  für  $x \geq 0$ , und somit errechnet man leicht

$$\begin{aligned} E(X) &= \int_0^\infty [1 - F(x)] dx \\ &= \int_0^\infty \exp(-\lambda x) dx \\ &= \left[ -\frac{1}{\lambda} \exp(-\lambda x) \right]_0^\infty \\ &= \frac{1}{\lambda}. \end{aligned}$$

### 4.4.3 Normalverteilung

*Notation:*  $X \sim N(\mu, \sigma^2)$

*Parameter:*  $\mu \in \mathbb{R}$  und  $\sigma^2 \in \mathbb{R}^+$

*Träger:*  $\mathcal{T}_X = \mathbb{R}$

*Dichtefunktion:*  $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right)$

*Erwartungswert:*  $E(X) = \mu$

*Varianz:*  $\text{Var}(X) = \sigma^2$

Ist  $\mu = 0$  und  $\sigma^2 = 1$ , so nennt man  $X$  *standardnormalverteilt*.

**Bemerkung:** Es gilt  $\int_{-\infty}^{\infty} f(x)dx = 1$ , da  $\int_{-\infty}^{\infty} \exp(-a^2x^2)dx = \sqrt{\pi}/a$  für  $a > 0$ . Die Verteilungsfunktion  $F(x)$  der Normalverteilung ist nicht analytisch zugänglich. Für die Dichte- und Verteilungsfunktion der Standardnormalverteilung verwendet man die Symbole  $\varphi(x)$  und  $\Phi(x)$ .

**Nachtrag:** Als *Normalisierungskonstante* bezeichnet man multiplikative Terme in der Dichtefunktion  $f(x)$ , die nicht vom Argument  $x$  abhängen (aber im Allgemeinen von den Parametern), der übrige Teil heißt *Kern*. Hier:

$$f(x) = c \cdot \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right) \quad \text{mit } c = \frac{1}{\sqrt{2\pi\sigma^2}}.$$

Kurz:

$$f(x) \propto \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right).$$

### 4.4.4 Gammaverteilung

*Notation:*  $X \sim G(a, b)$

*Parameter:*  $a, b \in \mathbb{R}^+$

*Träger:*  $\mathcal{T}_X = \mathbb{R}^+$

*Dichtefunktion* (\*):  $f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$  für  $x \in \mathcal{T}_X$

*Erwartungswert:*  $E(X) = \frac{a}{b}$

*Varianz:*  $\text{Var}(X) = \frac{a}{b^2}$

(\*): Hierbei ist  $\Gamma(a)$  die *Gammafunktion*

$$\Gamma(a) = \int_0^{\infty} x^{a-1} \exp(-x) dx \quad \text{für } a \in \mathbb{R}^+,$$

die folgende Eigenschaften hat:

$$\begin{aligned} \Gamma(a+1) &= \Gamma(a) \cdot a \quad \text{für } a \in \mathbb{R}^+, \\ \Gamma(a+1) &= a! \quad \text{für } a \in \mathbb{N}_0. \end{aligned}$$

Eigenschaften der Gammaverteilung:

$$\begin{aligned} G\left(\frac{d}{2}, \frac{1}{2}\right) &\hat{=} \chi_d^2 - \text{Verteilung mit } d \in \mathbb{N} \quad (\text{Erwartungswert } d, \text{ Varianz } 2d), \\ G(1, \lambda) &\hat{=} \text{Exp}(\lambda) \quad (\text{Exponentialverteilung}). \end{aligned}$$

Es stellt sich die Frage: Wieso ist  $f(x)$  eine Dichtefunktion?

1.  $f(x) \geq 0$  für alle  $x \in \mathbb{R}$  ergibt sich aus der Definition selbst.
2. Es gilt auch  $\int_0^{\infty} f(x) dx = 1$ .

*Beweis.* Verwende die Substitutionsregel

$$\int \tilde{f}(g(x))g'(x)dx = \int \tilde{f}(z)dz.$$

Mit  $g(x) = b \cdot x$ , also  $g'(x) = b$ , ergibt sich für die Dichtefunktion  $f$

$$\begin{aligned} f(x) &= \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) \\ &= \frac{1}{\Gamma(a)} [g(x)]^{a-1} \exp(-g(x)) \cdot b. \end{aligned}$$

Somit erhält man

$$\int_0^{\infty} f(x) dx = \frac{1}{\Gamma(a)} \int_0^{\infty} z^{a-1} \exp(-z) dz = \frac{1}{\Gamma(a)} \Gamma(a) = 1.$$

□

Wie berechnet man den Erwartungswert? Aus obigem Beweis erhalten wir

$$\int_0^{\infty} x^{a-1} \exp(-bx) dx = \frac{\Gamma(a)}{b^a}.$$

Damit:

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{\infty} x f(x) dx \\ &= \frac{b^a}{\Gamma(a)} \int_0^{\infty} x^a \exp(-bx) dx \\ &= \frac{b^a}{\Gamma(a)} \cdot \frac{\Gamma(a+1)}{b^{a+1}} \\ &= \frac{a}{b}. \end{aligned}$$

Analog berechnet man

$$E(X^2) = \frac{b^a}{\Gamma(a)} \cdot \frac{\Gamma(a+2)}{b^{a+2}} = \frac{(a+1)a}{b^2}$$

und damit  $\text{Var}(X) = a/b^2$ . Bei  $X \sim \chi_d^2$  (d.h.  $a = d/2$  und  $b = 1/2$ ) ergibt sich somit  $E(X) = d$  und  $\text{Var}(X) = 2d$ .

**Bemerkung:** Für  $a > 1$  hat die Dichtefunktion der Gammaverteilung einen eindeutigen Modus  $x_{mod} = (a-1)/b$ .

#### 4.4.5 Cauchy-Verteilung

*Notation:*  $X \sim C(\mu, \sigma^2)$

*Parameter:*  $\mu \in \mathbb{R}, \sigma \in \mathbb{R}^+$

*Träger:*  $\mathcal{T}_X = \mathbb{R}$

*Dichtefunktion:*  $f(x) = \frac{1}{\pi} \frac{1}{\sigma} \left(1 + \frac{(x-\mu)^2}{\sigma^2}\right)^{-1} = \frac{1}{\pi} \frac{\sigma}{\sigma^2 + (x-\mu)^2}$  für  $x \in \mathcal{T}_X$

*Erwartungswert* (\*): Existiert nicht!

*Varianz* (\*\*): Existiert nicht!

(\*): Die Cauchy-Verteilung besitzt *keinen* Erwartungswert, da  $\int x f(x) dx$  nicht absolut integrierbar ist:

$$\begin{aligned} \int_{-\infty}^{\infty} |x| f(x) dx &= \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2} dx \\ &= \frac{2}{\pi} \left[ \frac{1}{2} \log(1+x^2) \right]_0^{\infty} \\ &= \infty. \end{aligned}$$

(\*\*): Da  $E(X)$  nicht existiert, können auch keine höheren Momente existieren.

Im Weiteren wird die *Standard-Cauchy-Verteilung*  $C(0, 1)$  betrachtet, welche die Dichtefunktion

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

(d.h.  $\mu = 0, \sigma^2 = 1$ ) besitzt. Sie ist eine legale Dichtefunktion, da  $f(x) > 0$  für alle  $x \in \mathbb{R}$  und

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)} dx &= \left[ \frac{1}{\pi} \arctan(x) \right]_{-\infty}^{\infty} \\ &= \frac{1}{\pi} \left[ \frac{\pi}{2} + \frac{\pi}{2} \right] \\ &= 1. \end{aligned}$$

#### 4.4.6 $t$ -Verteilung / Studentverteilung

Eine stetige Zufallsvariable  $X$  heißt  $t$ -verteilt oder *Studentverteilt* mit Parametern  $\mu \in \mathbb{R}$ ,  $\sigma^2 > 0$  und  $\alpha > 0$ , wenn die Dichte die Form

$$f(x) = C \cdot \left(1 + \frac{1}{\alpha\sigma^2}(x - \mu)^2\right)^{-\frac{\alpha+1}{2}} \quad \text{für alle } x \in \mathbb{R}$$

hat. Schreibweise:  $X \sim t(\mu, \sigma^2, \alpha)$ . Die Normalisierungskonstante  $C$  ist gleich

$$C = \frac{\Gamma(\frac{1}{2}\alpha + \frac{1}{2})}{\Gamma(\frac{\alpha}{2})\Gamma(\frac{1}{2})} (\sigma^2\alpha)^{-\frac{1}{2}}.$$

Den Parameter  $\alpha$  nennt man "die *Freiheitsgrade*". Es gilt:

- $\text{Modus}(X) = \mu$ .
- $E(X) = \mu$ , falls  $\alpha > 1$ .
- $\text{Var}(X) = \sigma^2 \frac{\alpha}{\alpha-2}$ , falls  $\alpha > 2$ .
- $X \sim t(0, 1, \alpha) \Rightarrow \sigma X + \mu \sim t(\mu, \sigma^2, \alpha)$ .
- Für  $\alpha \rightarrow \infty$  konvergiert  $X$  gegen eine Normalverteilung.
- Für  $\alpha = 1$  ergibt sich die Cauchy-Verteilung.
- Für  $\mu = 0$  und  $\sigma^2 = 1$  ergibt sich die *Standard-t-Verteilung*.

#### 4.4.7 Betaverteilung

*Notation:*  $X \sim \mathcal{Be}(a, b)$

*Parameter:*  $a, b \in \mathbb{R}^+$

*Träger:*  $\mathcal{T}_X = [0, 1]$

*Dichtefunktion* (\*<sup>1</sup>):  $f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$  für  $x \in \mathcal{T}_X$

*Erwartungswert* (\*<sup>2</sup>):  $E(X) = \frac{a}{a+b}$

*Varianz* (\*<sup>3</sup>):  $\text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$

(\*<sup>1</sup>): Hierbei ist  $B(a, b)$  die *Betafunktion*

$$B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx.$$

Daher:  $\int_0^1 f(x) dx = 1$ . Es gilt  $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ .

(\*<sup>2</sup>): Es gilt

$$\begin{aligned} E(X) &= \int_0^1 xf(x)dx \\ &= \frac{1}{B(a, b)} \int_0^1 x^a(1-x)^{b-1}dx \\ &= \frac{B(a+1, b)}{B(a, b)} \\ &= \frac{\Gamma(a+1)}{\Gamma(a)} \cdot \frac{\Gamma(a+b)}{\Gamma(a+b+1)} \\ &= \frac{\Gamma(a) \cdot a}{\Gamma(a)} \cdot \frac{\Gamma(a+b)}{\Gamma(a+b) \cdot (a+b)} \\ &= \frac{a}{a+b}. \end{aligned}$$

(\*<sup>3</sup>): Analog zu (\*<sup>2</sup>) berechnet man

$$\begin{aligned} E(X^2) &= \int_0^1 x^2f(x)dx \\ &= \frac{1}{B(a, b)} \int_0^1 x^{a+1}(1-x)^{b-1}dx \\ &= \frac{B(a+2, b)}{B(a, b)} \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} \\ &= \frac{a(a+1)}{(a+b+1)(a+b)}. \end{aligned}$$

Mit dem Verschiebungssatz (Satz 3.3.6) ergibt sich

$$\text{Var}(X) = \frac{a(a+1)}{(a+b+1)(a+b)} - \frac{a^2}{(a+b)^2} = \frac{ab}{(a+b)^2(a+b+1)}.$$

## 4.5 JENSENSCHE UNGLEICHUNG UND INFORMATIONSUNGLEICHUNG

**Satz 4.5.1** (JENSENSCHE UNGLEICHUNG, 1869–1925). *Sei  $X$  eine Zufallsvariable mit endlichem Erwartungswert  $E(X)$  und  $g(x)$  eine konvexe Funktion (bei existierender zweiter Ableitung ist dies gleichbedeutend mit  $g''(x) \geq 0$  für alle  $x \in \mathbb{R}$ ). Dann gilt:*

$$E(g(X)) \geq g(E(X)).$$

*Ist  $g(x)$  strikt konvex ( $g''(x) > 0$  für alle reellen  $x$ ) und zusätzlich  $X$  fast sicher nicht konstant ("nicht degeneriert"), so gilt sogar*

$$E(g(X)) > g(E(X)).$$

*Analog gilt für (strikt) konkave Funktionen  $g(x)$*

$$E(g(X)) \leq g(E(X))$$

bzw.

$$\mathbb{E}(g(X)) < g(\mathbb{E}(X)).$$

**Beispiel 4.5.2** (Log-Normalverteilung). Sei  $X \sim N(\mu, \sigma^2)$  und  $Y = \exp(X)$ . Dann sagt man,  $Y$  hat eine *Log-Normalverteilung* mit Parametern  $\mu$  und  $\sigma^2$ :  $Y \sim \text{LN}(\mu, \sigma^2)$ . Da  $g(x) = \exp(x)$  strikt konvex ist und  $X$  nicht degeneriert ( $\sigma^2 > 0$ ), folgt

$$\mathbb{E}(Y) = \mathbb{E}(\exp(X)) > \exp(\mathbb{E}(X)) = \exp(\mu).$$

**Bemerkung:** Über momenterzeugende Funktionen (siehe Kapitel 5) lässt sich genauer zeigen:

$$\mathbb{E}(Y) = \exp\left(\mu + \frac{1}{2}\sigma^2\right).$$

**Satz 4.5.3** (Informationsungleichung). Seien  $f_X(x)$  und  $f_Y(y)$  zwei Dichtefunktionen von stetigen Zufallsvariablen  $X$  und  $Y$  mit gleichen Trägern. Dann gilt:

$$D(f_X, f_Y) := \mathbb{E}\left(\log \frac{f_X(X)}{f_Y(X)}\right) = \mathbb{E}(\log f_X(X)) - \mathbb{E}(\log f_Y(X)) \geq 0,$$

wobei  $X \sim f_X(x)$  und Gleichheit nur dann gilt, wenn  $f_X(x) = f_Y(x)$  für alle  $x \in \mathbb{R}$ .

Man betrachtet also zwei Funktionen der Zufallsvariablen  $X$ , einmal die logarithmierte Dichtefunktion  $f_X(x)$  von  $X$  und einmal die von  $Y$ ,  $f_Y(y)$ , ausgewertet an  $X$ . Die Informationsungleichung sagt aus, dass

$$\mathbb{E}(\log f_X(X)) \geq \mathbb{E}(\log f_Y(X)),$$

d.h. im Schnitt sind die logarithmierten Dichtewerte am größten, wenn man die Dichtefunktion  $f_X$  an der zugehörigen Zufallsvariablen  $X$  auswertet.

*Beweisskizze.* Definiere  $Z = f_Y(X)/f_X(X) \geq 0$ . Für  $f_X \neq f_Y$  ist  $Z$  nicht degeneriert. Dann gilt:

$$\begin{aligned} \mathbb{E}\left(\log \frac{f_X(X)}{f_Y(X)}\right) &= \mathbb{E}\left(-\log \frac{f_Y(X)}{f_X(X)}\right) \\ &= \mathbb{E}(-\log(Z)) \\ &> -\log \mathbb{E}(Z) \quad , \text{ da } -\log(z) \text{ strikt konvex ist} \\ &= -\log \int \frac{f_Y(x)}{f_X(x)} f_X(x) dx \\ &= -\log \int f_Y(x) dx \\ &= -\log(1) = 0. \end{aligned}$$

Bei  $f_X = f_Y$  gilt natürlich Gleichheit. □

Die Größe  $D(f_X, f_Y) \geq 0$  heißt *Kullback-Leibler-Distanz* (Solomon KULLBACK, 1907–1994, Richard LEIBLER, 1914–2003) bzw. auch *-Diskrepanz* und misst gewissermaßen den "Abstand" zwischen  $f_X$  und  $f_Y$ , wobei aber im Allgemeinen

$$D(f_X, f_Y) \neq D(f_Y, f_X)$$

gilt, da  $D(f_X, f_Y)$  unsymmetrisch definiert ist.

**Beispiel 4.5.4.** Seien  $X \sim N(0, 1)$  und  $Y \sim N(\mu, \sigma^2)$ , dann gilt

$$D(f_X, f_Y) = \frac{1}{2} \left( \log \sigma^2 + \frac{1 + \mu^2}{\sigma^2} - 1 \right)$$

(vgl. Übungsblatt 8).

**Bemerkung:** Die Kullback-Leibler-Distanz kann auch für diskrete Zufallsvariablen definiert werden, und die Informationsungleichung gilt dort genauso.

## 4.6 Stochastische Abhängigkeit

**Definition 4.6.1** (Gemeinsame Verteilungs- und Dichtefunktion). Die *gemeinsame Verteilungsfunktion* von zwei stetigen Zufallsvariablen  $X$  und  $Y$  ist durch

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

gegeben. Ihre *gemeinsame Dichtefunktion*  $f_{X,Y} : \mathbb{R}^2 \rightarrow [0, \infty)$  hat die Eigenschaft

$$F_{X,Y}(x, y) = \int_{v=-\infty}^y \int_{u=-\infty}^x f(u, v) du dv \quad \text{für alle } x, y \in \mathbb{R}.$$

**Korollar 4.6.2.** Falls ein  $F_{X,Y}$  gemäß Definition 4.6.1 existiert und auf seinem Träger vollständig (stetig) differenzierbar ist, so gilt für die zugehörige Dichtefunktion  $f_{X,Y}$

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y).$$

**Bemerkung:**  $F_{X,Y}(x, y)$  und  $f_{X,Y}(x, y)$  haben analoge Eigenschaften wie im univariaten Fall.

**Definition 4.6.3** (Randverteilungs- und Randdichtefunktion). Sei  $F_{X,Y}$  eine gemeinsame Verteilungsfunktion gemäß Definition 4.6.1. Dann lauten die *Randverteilungsfunktionen*:

$$\begin{aligned} F_X(x) &= P(X \leq x) = F_{X,Y}(x, \infty), \\ F_Y(y) &= P(Y \leq y) = F_{X,Y}(\infty, y). \end{aligned}$$

Die *Randdichtefunktionen* sind mit

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy, \\ f_Y(y) &= \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx \end{aligned}$$

definiert.

**Definition 4.6.4** (Erwartungswerte). Sei  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$  eine "reguläre" Funktion, dann gilt:

$$E(g(X, Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{X,Y}(x, y) dx dy.$$

Inbesondere gilt:

$$E(X \cdot Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x \cdot y f_{X,Y}(x, y) dx dy.$$

**Korollar 4.6.5.** Seien  $X$  und  $Y$  zwei entsprechend den Definitionen 4.6.3 und 4.6.4 gutartige Zufallsvariablen, dann gilt:

$$E(aX + bY) = aE(X) + bE(Y).$$

*Beweis.* Aus den Definitionen 4.6.3 und 4.6.4 und dem Satz von Fubini folgt

$$\begin{aligned} E(aX + bY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (ax + by) f_{X,Y}(x, y) dx dy \\ &= \left( \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} ax f_{X,Y}(x, y) dx dy \right) + \left( \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} by f_{X,Y}(x, y) dx dy \right) \\ &= a \left( \int_{-\infty}^{+\infty} x \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy dx \right) + b \left( \int_{-\infty}^{+\infty} y \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx dy \right) \\ &= a \int_{-\infty}^{+\infty} x f_X(x) dx + b \int_{-\infty}^{+\infty} y f_Y(y) dy \\ &= aE(X) + bE(Y). \end{aligned}$$

□

**Definition 4.6.6** (Unabhängigkeit). Zwei stetige Zufallsvariablen  $X$  und  $Y$  heißen genau dann *unabhängig*, wenn

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y) \quad \text{für alle } x, y \in \mathbb{R} \quad (22)$$

gilt.

**Korollar 4.6.7.** Seien  $X$  und  $Y$  zwei stetige Zufallsvariablen, deren gemeinsame Verteilungsfunktion im Punkt  $(x, y)$  differenzierbar ist. Dann sind  $X$  und  $Y$  unabhängig, wenn

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$$

gilt.

*Beweis.* Differenziere (22) nach  $x$  und  $y$ . Für die linke Seite ergibt sich mit Korollar 4.6.2

$$\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) = f_{X,Y}(x, y)$$

Auf der rechten Seite erhalten wir

$$\begin{aligned} \frac{\partial^2}{\partial x \partial y} F_X(x) \cdot F_Y(y) &= \frac{\partial}{\partial x} \left( \frac{\partial}{\partial y} F_X(x) \cdot F_Y(y) \right) \\ &= \frac{\partial}{\partial x} (f_X(x) \cdot f_Y(y)) \\ &= f_X(x) \cdot f_Y(y). \end{aligned}$$

□

**Definition 4.6.8.** Ein zweidimensionaler Zufallsvektor  $(X, Y)$  heißt *bivariat standardnormalverteilt*, wenn er eine Dichte  $f_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}$  hat, die wie folgt lautet:

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

für  $x, y \in \mathbb{R}$  mit  $\rho \in (-1, 1)$ .

Man zeigt leicht, dass

- $f_{X,Y}(x, y) > 0, \forall x, y \in \mathbb{R}$ ,
- $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx dy = 1$

und somit  $f_{X,Y}(x, y)$  eine legitime Dichtefunktion ist.

Ferner gilt für die Randverteilungen folgender Satz:

**Satz 4.6.9.** Sei  $(X, Y)$  ein bivariat standardnormalverteilter Zufallsvektor. Dann sind die Randverteilungen von  $X$  und  $Y$  für alle Werte von  $\rho \in (-1, 1)$  standardnormalverteilt, d.h.

$$\begin{aligned} X &\sim N(0, 1) \\ Y &\sim N(0, 1). \end{aligned}$$

Zum Beweis von Satz 4.6.9 benötigen wir

**Lemma 4.6.10.** Der Kern der Normalverteilungsdichte lässt sich wie folgt faktorisieren:

$$\begin{aligned} K(x, y) &= \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right) \\ &= \exp\left(-\frac{1}{2}y^2\right) \exp\left(-\frac{1}{2} \cdot \frac{(x - \rho y)^2}{1 - \rho^2}\right) \end{aligned} \quad (23)$$

$$= \exp\left(-\frac{1}{2}x^2\right) \exp\left(-\frac{1}{2} \cdot \frac{(y - \rho x)^2}{1 - \rho^2}\right). \quad (24)$$

*Beweis.* Es ist

$$\frac{(x - \rho y)^2}{1 - \rho^2} + y^2 = \frac{x^2 - 2\rho xy + \rho^2 y^2 + (1 - \rho^2)y^2}{1 - \rho^2} = \frac{x^2 - 2\rho xy + y^2}{1 - \rho^2},$$

daraus folgt (23). Der Beweis zu (24) folgt analog. □

*Beweis zu Satz 4.6.9.* Mit  $K(x, y)$  definiert wie in Lemma 4.6.10 folgt nun

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi} \cdot \frac{1}{\sqrt{1-\rho^2}} \cdot K(x, y) dx \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} \cdot \frac{(x - \rho y)^2}{1 - \rho^2}\right) dx}_{=1, \text{ da Integral über Dichte der } N(\rho y, 1-\rho^2)} \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right). \end{aligned}$$

Es gilt also  $Y \sim N(0, 1)$ .  $X \sim N(0, 1)$  folgt analog. □

Die Kovarianz eines bivariat standardnormalverteilten Zufallsvektors  $(X, Y)$  berechnet sich wie folgt:

$$\begin{aligned}
\text{Cov}(X, Y) &= \underbrace{\text{E}(X \cdot Y)} - \underbrace{\text{E}(X)}_{=0} \underbrace{\text{E}(Y)}_{=0} \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) dx dy \\
&\stackrel{(23)}{=} \int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) \underbrace{\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} \frac{(x-\rho y)^2}{1-\rho^2}\right) dx}_{\text{Erw.wert einer } N(\rho y, 1-\rho^2)\text{-verteilten ZV}} dy \\
&= \int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) \cdot \rho y dy \\
&= \rho \underbrace{\int_{-\infty}^{\infty} y^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) dy}_{=E(Y^2)=\text{Var}(Y)=1} \\
&= \rho.
\end{aligned}$$

Da die Randverteilungen standardnormalverteilt sind und somit  $\text{Var}(X) = \text{Var}(Y) = 1$ , gilt für den Korrelationskoeffizienten:

$$\rho(x, y) = \rho.$$

Man zeigt leicht, dass im Fall von zwei bivariat standardnormalverteilten Zufallsvariablen  $X$  und  $Y$  gilt: Sind  $X$  und  $Y$  unkorreliert, so sind sie auch unabhängig.

*Beweis.* Nach Voraussetzung gilt  $\rho = 0$ . Nun lässt sich die Dichte wie folgt faktorisieren:

$$\begin{aligned}
f_{X,Y}(x, y) &= \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) \\
&= f_X(x) \cdot f_Y(y),
\end{aligned}$$

d.h.  $X$  und  $Y$  sind unabhängig. □

**Definition 4.6.11.** Die allgemeine Form der Dichtefunktion der *bivariaten Normalverteilung* mit Parametern  $\mu, \sigma^2, \nu, \tau^2$  und  $\rho$  lautet:

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[ \left(\frac{x-\mu}{\sigma}\right)^2 - 2\rho \left(\frac{x-\mu}{\sigma}\right) \left(\frac{y-\nu}{\tau}\right) + \left(\frac{y-\nu}{\tau}\right)^2 \right]\right).$$

Für die Randdichten gilt nun

$$\begin{aligned}
X &\sim N(\mu, \sigma^2) \\
Y &\sim N(\nu, \tau^2).
\end{aligned}$$

Als Kovarianz ergibt sich  $\text{Cov}(X, Y) = \rho \cdot \sigma \cdot \tau$ , so dass  $\rho(X, Y) = \rho$  folgt.

**Beispiel 4.6.12.** Sei  $(X, Y)$  ein Zufallsvektor mit folgender Dichtefunktion:

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{x} & \text{für } 0 < y \leq x < 1 \\ 0 & \text{sonst.} \end{cases}$$

Dann ergibt sich die Randverteilung von  $X$  zu

$$f_X(x) = \int_0^x \frac{1}{x} dy = 1$$

für  $0 < x < 1$ . Folglich ist  $X \sim U(0, 1)$  mit

$$E(X) = \frac{1}{2} \quad \text{und} \quad \text{Var}(X) = \frac{1}{12} .$$

Die Randverteilung von  $Y$  lautet

$$f_Y(y) = \int_y^1 \frac{1}{x} dx = \log\left(\frac{1}{y}\right)$$

für  $0 < y < 1$ . Mit **Maple** errechnet man:

$$E(Y) = \frac{1}{4} \quad \text{und} \quad \text{Var}(Y) = \frac{7}{144} .$$

Ferner ergibt sich

$$\begin{aligned} E(X \cdot Y) &= \int_0^1 \int_0^x xy \cdot \frac{1}{x} dy dx \\ &= \int_0^1 \left( \int_0^x y dy \right) dx \\ &= \int_0^1 \frac{x^2}{2} dx \\ &= \frac{1}{6} \end{aligned}$$

und damit

$$\text{Cov}(X, Y) = \frac{1}{6} - \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{24} .$$

Abschließend berechnet sich der Korrelationskoeffizient aus bekannter Formel zu

$$\rho(X, Y) = \frac{\frac{1}{24}}{\sqrt{\frac{1}{12}} \sqrt{\frac{7}{144}}} \approx 0.65 .$$

## 4.7 Bedingte Verteilungen und bedingte Erwartungen

Diesem Unterkapitel liegt folgende Frage zu Grunde: Seien  $X$  und  $Y$  stetige Zufallsvariablen mit gemeinsamer Dichtefunktion  $f_{X,Y}(x, y)$ . Wie lautet die bedingte Verteilungsfunktion und die bedingte Dichtefunktion von  $Y$  gegeben  $X = x$ ?

Das Problem liegt in der Stetigkeit von  $X$ . Da  $X$  stetig ist, gilt  $P(X = x) = 0$ , und daher ist  $P(Y \leq y | X = x)$  nicht definiert. Zur Lösung dieses Dilemmas geht man wie folgt vor:

Sei  $f_X(x) > 0$ . Dann ist

$$\begin{aligned} P(Y \leq y | x \leq X \leq x + dx) &= \frac{P(Y \leq y, x \leq X \leq x + dx)}{P(x \leq X \leq x + dx)} \\ &\stackrel{dx \text{ klein}}{\approx} \frac{\int_{v=-\infty}^y f_{X,Y}(x, v) dx dv}{f_X(x) dx} \\ &= \int_{v=-\infty}^y \frac{f_{X,Y}(x, v)}{f_X(x)} dv \end{aligned}$$

Für  $dx \rightarrow 0$  erhält man die folgenden zwei Definitionen:

**Definition 4.7.1** (Bedingte Verteilungsfunktion). Die *bedingte Verteilungsfunktion* von  $Y$  gegeben  $X = x$  lautet

$$F_{Y|X}(y|x) = \int_{-\infty}^y \frac{f_{X,Y}(x, v)}{f_X(x)} dv$$

für alle  $x$  mit  $f_X(x) > 0$ .

**Definition 4.7.2** (Bedingte Dichtefunktion). Die *bedingte Dichtefunktion* von  $Y$  gegeben  $X = x$  lautet

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

für alle  $x$  mit  $f_X(x) > 0$ .

**Beispiel 4.7.3.** Wie in Beispiel 4.6.12 sei die gemeinsame Dichtefunktion von zwei Zufallsvariablen  $X$  und  $Y$  durch

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{x} & 0 < y \leq x < 1 \\ 0 & \text{sonst} \end{cases}$$

gegeben. Wir haben bereits die Randdichtefunktionen bestimmt:

$$\begin{aligned} f_X(x) &= 1 && \text{für } 0 \leq x \leq 1, \\ f_Y(y) &= \log\left(\frac{1}{y}\right) && \text{für } 0 \leq y \leq 1. \end{aligned}$$

Mit Definition 4.7.2 ergibt sich

$$f_{Y|X}(y|x) = \frac{\frac{1}{x}}{1} = \frac{1}{x} \quad \text{für } 0 < y \leq x,$$

und somit gilt für  $Y|X$ :

$$Y|\{X = x\} \sim U(0, x).$$

Weiterhin

$$f_{X|Y}(x|y) = \frac{\frac{1}{x}}{\log\left(\frac{1}{y}\right)} = \frac{1}{x \log\left(\frac{1}{y}\right)} \quad \text{für } y \leq x < 1.$$

Wie bei diskreten Zufallsvariablen definiert man den bedingten Erwartungswert:

**Definition 4.7.4.** Seien  $X$  und  $Y$  stetige Zufallsvariablen mit gemeinsamer Dichtefunktion. Dann heißt

$$\psi(x) = E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy$$

bedingter Erwartungswert von  $Y$  gegeben  $X = x$ , wenn die bedingte Dichtefunktion  $f_{Y|X}(y|x)$  Definition 4.7.2 genügt.

Analog zu Definition 3.7.5 kann man die bedingte Erwartung definieren, indem man  $x$  durch  $X$  ersetzt und somit eine *Zufallsvariable* statt einer reellen Zahl erhält:

**Definition 4.7.5.** Sei  $\psi(x) = E(Y|X = x)$  gemäß Definition 4.7.4. Dann nennt man die Zufallsvariable  $\psi(X) = E(Y|X)$  die *bedingte Erwartung* von  $Y$  gegeben  $X$ .

Es gilt wie im diskreten Fall der Satz vom iterierten Erwartungswert (Satz 3.7.6):

**Satz 4.7.6** (Satz vom iterierten Erwartungswert). *Sei  $\psi(X) = E(Y|X)$  eine bedingte Erwartung nach Definition 4.7.5. Dann gilt*

$$E(\psi(X)) = E(Y).$$

*Dies ist äquivalent zu den Schreibweisen*

$$\begin{aligned} E(E(Y|X)) &= E(Y), \\ \int_{-\infty}^{\infty} E(Y|X = x) f_X(x) dx &= E(Y). \end{aligned}$$

Ebenso gilt der Varianzzerlegungssatz analog zu Satz 3.7.11:

**Satz 4.7.7** (Varianzzerlegungssatz). *Sei  $Y|X$  eine stetige bedingte Zufallsvariable. Dann gilt:*

$$\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X)).$$

Setzen wir mit diesem Wissen Beispiel 4.7.3 fort:

Die Berechnung von  $E(Y)$  und  $\text{Var}(Y)$  direkt über  $f_Y(y) = \log\left(\frac{1}{y}\right)$ ,  $0 < y < 1$ , war recht umständlich (Maple). Nun ist aber bekannt, dass  $Y|\{X = x\} \sim U(0, x)$  und  $X \sim U(0, 1)$ . Mit den Sätzen 4.7.6 und 4.7.7 ergibt sich wesentlich einfacher:

$$E(Y) = E(E(Y|X)) = E\left(\frac{1}{2}X\right) = \frac{1}{2}E(X) = \frac{1}{4}$$

und

$$\begin{aligned}
 \text{Var}(Y) &= \overline{\text{E}(\text{Var}(Y|X))} + \overline{\text{Var}(\text{E}(Y|X))} \\
 &= \overline{\text{E}\left(\frac{1}{12}X^2\right)} + \overline{\text{Var}\left(\frac{1}{2}X\right)} \\
 &= \frac{1}{12}\overline{\text{E}(X^2)} + \frac{1}{4}\overline{\text{Var}(X)} \\
 &= \frac{1}{12}(\overline{\text{Var}(X)} + \overline{\text{E}(X)^2}) + \frac{1}{4}\overline{\text{Var}(X)} \\
 &= \frac{1}{12}\left(\frac{1}{12} + \frac{1}{4}\right) + \frac{1}{4} \cdot \frac{1}{12} \\
 &= \frac{1}{12} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{1}{12} \\
 &= \frac{7}{144}.
 \end{aligned}$$

Auch die Berechnung von  $\text{E}(X \cdot Y)$  kann über bedingte Erwartungen erfolgen:

$$\begin{aligned}
 \text{E}(X \cdot Y) &= \overline{\text{E}(\text{E}(X \cdot Y|X))} \\
 &= \overline{\text{E}(X \text{E}(Y|X))} \\
 &= \overline{\text{E}\left(X \frac{X}{2}\right)} \\
 &= \frac{1}{2}\overline{\text{E}(X^2)} \\
 &= \frac{1}{2}(\overline{\text{Var}(X)} + \overline{\text{E}(X)^2}) \\
 &= \frac{1}{2}\left(\frac{1}{12} + \frac{1}{4}\right) \\
 &= \frac{1}{6}.
 \end{aligned}$$

**Beispiel 4.7.8** (Bivariate Standardnormalverteilung). Für einen bivariat standardnormalverteilten Zufallsvektor  $(X, Y)$  gilt:

$$\begin{aligned}
 f_{Y|X}(y|x) &= \frac{f_{X,Y}(x, y)}{f_X(x)} \\
 &\stackrel{\text{Lemma 4.6.10}}{=} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} \frac{(y-\rho x)^2}{1-\rho^2}\right).
 \end{aligned}$$

Folglich ist

$$Y|\{X = x\} \sim N(\rho \cdot x, 1 - \rho^2),$$

und somit ergibt sich

$$E(Y|X = x) = \rho \cdot x \quad \text{und} \quad E(Y|X) = \rho \cdot X.$$

Analog gelten

$$X|\{Y = y\} \sim N(\rho \cdot y, 1 - \rho^2) \quad \text{und} \quad E(X|Y) = \rho \cdot Y.$$

Für die bedingten Verteilungen bei der *allgemeinen* bivariaten Normalverteilung (vgl. Definition 4.6.11) ergibt sich:

$$\begin{aligned} Y|\{X = x\} &\sim N\left(\nu + \rho \cdot \frac{\tau}{\sigma}(x - \mu), \tau^2(1 - \rho^2)\right) \\ X|\{Y = y\} &\sim N\left(\mu + \rho \cdot \frac{\sigma}{\tau}(y - \nu), \sigma^2(1 - \rho^2)\right). \end{aligned}$$

**Bemerkungen:**

1. Die Faktorisierung

$$\begin{aligned} f_{X,Y}(x, y) &= f_X(x) \cdot f_{Y|X}(y|x) \\ f_{X,Y}(x, y) &= f_Y(y) \cdot f_{X|Y}(x|y) \end{aligned}$$

wird häufig zur Charakterisierung von gemeinsamen Dichten durch univariate Dichten benutzt. Sie wird insbesondere zur Simulation von bivariaten Verteilungen verwendet. Allgemein gilt:

$$f(x_1, \dots, x_n) = f(x_1) \cdot f(x_2|x_1) \cdot f(x_3|x_1, x_2) \cdot \dots \cdot f(x_n|x_1, \dots, x_{n-1}). \quad (25)$$

Aus Gründen der Übersichtlichkeit wurden die Indizes der Dichte weggelassen. Es ist jedoch selbstverständlich, dass es sich um jeweils verschiedene Dichtefunktionen handelt. Die Faktorisierung behält auch bei jeder anderen Reihenfolge der Indizes  $1, 2, \dots, n$  ihre Gültigkeit, beispielsweise gilt

$$f(x_1, \dots, x_n) = f(x_n) \cdot f(x_{n-1}|x_n) \cdot f(x_{n-2}|x_n, x_{n-1}) \cdot \dots \cdot f(x_1|x_n, x_{n-1}, \dots, x_2).$$

Hat  $X_1, \dots, X_n$  die *Markov-Eigenschaft*, so lassen sich die bedingten Dichten in (25) vereinfachen:

$$f(x_1, \dots, x_n) = f(x_1) \cdot f(x_2|x_1) \cdot f(x_3|x_2) \cdot \dots \cdot f(x_n|x_{n-1}).$$

Die Markov-Eigenschaft impliziert somit

$$f(x_i|x_1, \dots, x_{i-1}) = f(x_i|x_{i-1})$$

für alle  $i = 2, \dots, n$ , d.h. gegeben die "Vergangenheit"  $X_1 = x_1, \dots, X_{i-1} = x_{i-1}$  hängt die Verteilung von  $X_i$  nur vom letzten Wert  $X_{i-1} = x_{i-1}$  ab.

2. Die Technik des Bedingens, der Berechnung von bedingten Erwartungen etc. lässt sich analog auch auf Zufallsvektoren anwenden, bei denen eine Komponente stetig, die andere hingegen diskret ist. Alle Rechenregeln behalten ihre Gültigkeit.

**Beispiel 4.7.9** (Beta-Binomialverteilung). Die *Beta-Binomialverteilung* findet häufig Anwendung in der Modellierung von *Überdispersion* (engl. *overdispersion*). Das bedeutet, dass die Varianz der Daten größer ist als man unter Annahme einer Binomialverteilung erwarten würde.

Die Konstruktion dieser Verteilung ergibt sich wie folgt:

$$\begin{aligned} X &\sim \mathcal{Be}(a, b) \\ Y|\{X = x\} &\sim B(n, x), \end{aligned}$$

wobei  $n$  als fest vorausgesetzt wird.

Dies definiert die gemeinsame Verteilung mit der Dichte- bzw. Wahrscheinlichkeitsfunktion

$$\begin{aligned} f_{X,Y}(x, y) &= f_X(x) \cdot f_{Y|X}(y|x) \\ &= \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} \cdot \binom{n}{y} x^y (1-x)^{n-y} \\ &= \binom{n}{y} \frac{1}{B(a, b)} x^{a+y-1} (1-x)^{b+n-y-1} \end{aligned}$$

für  $0 < x < 1$  und  $y \in \{0, 1, \dots, n\}$ . Daraus berechnet man leicht durch Rausintegrieren von  $x$  die Randverteilung von  $Y$ :

$$\begin{aligned} f_Y(y) &= \int_0^1 f_{X,Y}(x, y) dx \\ &= \binom{n}{y} \frac{1}{B(a, b)} \int_0^1 x^{a+y-1} (1-x)^{b+n-y-1} dx \\ &= \binom{n}{y} \frac{1}{B(a, b)} B(a+y, b+n-y) \\ &= \binom{n}{y} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+y)\Gamma(b+n-y)}{\Gamma(a+b+n)} \\ &= \frac{\Gamma(a+b)}{\Gamma(a+b+n)\Gamma(a)\Gamma(b)} \cdot \binom{n}{y} \Gamma(a+y)\Gamma(b+n-y) \end{aligned}$$

für  $y \in \{0, 1, \dots, n\}$ .

**Definition 4.7.10** (Beta-Binomialverteilung). Sei  $X$  eine Zufallsgröße mit der Wahrscheinlichkeitsfunktion

$$f_X(x) = \frac{\Gamma(a+b)}{\Gamma(a+b+n)\Gamma(a)\Gamma(b)} \cdot \binom{n}{x} \Gamma(a+x)\Gamma(b+n-x)$$

für  $x \in \{0, 1, \dots, n\}$ . Dann heißt  $X$  *Beta-binomialverteilt*, und man schreibt

$$X \sim \text{BeB}(n, a, b).$$

**Fortsetzung von Beispiel 4.7.9** Die Beta-Binomialverteilung besitzt folgende Spezialfälle:

1. Für  $a = b = 1$  ergibt sich:

$$\begin{aligned} f_Y(y) &= \frac{\Gamma(2)}{\Gamma(2+n)\Gamma(1)\Gamma(1)} \binom{n}{y} \cdot \Gamma(y+1)\Gamma(1+n-y) \\ &= \frac{1}{(n+1)!} \cdot \frac{n!}{y! \cdot (n-y)!} \cdot y! \cdot (n-y)! \\ &= \frac{1}{n+1} \end{aligned}$$

für  $y \in \{0, 1, \dots, n\}$ . Also ist  $\text{BeB}(n, 1, 1)$  eine Gleichverteilung auf  $\{0, 1, \dots, n\}$ .

2. Für  $n = 1$  ergibt sich:

$$\begin{aligned}
 f_Y(y) &= \frac{\Gamma(a+b)}{\Gamma(a+b+1)\Gamma(a)\Gamma(b)} \cdot \binom{1}{y} \Gamma(a+y)\Gamma(b+1-y) \\
 &= \left\{ \begin{array}{l} \frac{a}{a+b} \quad \text{für } y = 1 \\ \frac{b}{a+b} \quad \text{für } y = 0 \end{array} \right\} \\
 &= \left( \frac{a}{a+b} \right)^y \left( 1 - \frac{a}{a+b} \right)^{1-y}
 \end{aligned}$$

Also ist  $BeB(1, a, b)$  eine Bernoulliverteilung mit dem Parameter  $\pi = \frac{a}{a+b}$ . In diesem Falle ist keine Überdispersion möglich.

Erwartungswert und Varianz der Beta-Binomialverteilung lassen sich wie folgt bestimmen:

$$E(Y) = E(E(Y|X)) = E(n \cdot X) = n \cdot E(X) = n \frac{a}{a+b}$$

und

$$\begin{aligned}
 \text{Var}(Y) &= E(\text{Var}(Y|X)) + \text{Var}(E(Y|X)) \\
 &= E(n \cdot X(1-X)) + \text{Var}(n \cdot X) \\
 &= n[E(X) - E(X^2)] + n^2 \text{Var}(X) \\
 &= n[E(X) - E(X^2)] + n^2[E(X^2) - (E(X))^2] \\
 &= n[E(X^2)(n-1) + E(X)[1 - n \cdot E(X)]].
 \end{aligned}$$

Aus Abschnitt 4.4.7 ist bekannt, dass

$$\begin{aligned}
 E(X) &= \frac{a}{a+b} \\
 E(X^2) &= \frac{a}{a+b} \cdot \frac{a+1}{a+b+1} = E(X) \frac{a+1}{a+b+1}.
 \end{aligned}$$

Somit ergibt sich:

$$\begin{aligned}
 \text{Var}(Y) &= n \left[ E(X) \frac{a+1}{a+b+1} (n-1) + E(X) [1 - n \cdot E(X)] \right] \\
 &= n E(X) \left[ \frac{(a+1)(n-1)}{a+b+1} + 1 - \frac{n \cdot a}{a+b} \right] \\
 &= n E(X) \left[ \frac{b^2 + nb + ab}{(a+b+1)(a+b)} \right] \\
 &= n E(X) \left[ \frac{b(a+b+n)}{(a+b+1)(a+b)} \right] \\
 &= n E(X) \left[ \frac{b}{a+b} \cdot \frac{a+b+n}{a+b+1} \right] \\
 &= n E(X) \left[ \left( 1 - \frac{a}{a+b} \right) \frac{a+b+n}{a+b+1} \right] \\
 &= \underbrace{n E(X)(1 - E(X))}_I \cdot \underbrace{\frac{a+b+n}{a+b+1}}_{II}.
 \end{aligned}$$

I) Varianz der Binomialverteilung

II) Multiplikation mit einem Faktor größer 1 für den Fall  $n > 1$ , also den Fall, dass Dispersion auftritt. Dieser Faktor  $C$  heißt auch *Dispersionsfaktor*.

Eine häufige Parametrisierung der Beta-Binomialverteilung lautet

$$\begin{aligned}\mu &= \frac{a}{a+b} \\ \rho &= \frac{1}{a+b+1}.\end{aligned}$$

Damit ergibt sich nun:

$$\begin{aligned}\mathbb{E}(Y) &= n\mu \\ \text{Var}(Y) &= n\mu(1-\mu)(1+(n-1)\rho) \\ &= n\mu(1-\mu)C.\end{aligned}$$

Man beachte:

1. Für  $n = 1$  ist  $C = 1$ . Es gibt also keine Überdispersion bei der Bernoulliverteilung. Siehe hierzu auch den Spezialfall 2 in diesem Beispiel.
2. Es gilt allgemein für  $n > 1$ :

$$C = 1 + \frac{n-1}{a+b+1} = 1 + (n-1)\rho < 1 + (n-1) = n.$$

Das heißt, dass für  $n > 1$  somit  $C < n$  für alle  $a, b \in \mathbb{R}$  gilt.

#### 4.8 Funktionen von Zufallsvariablen

Sei  $X$  eine stetige Zufallsvariable mit Dichte  $f_X$  und  $g : \mathbb{R} \rightarrow \mathbb{R}$  eine Funktion. Dann ist (unter Regularitätsbedingungen) auch  $Y = g(X)$  eine Zufallsvariable. Die Verteilungsfunktion von  $Y$  ergibt sich zu

$$\begin{aligned}F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(g(X) \in (-\infty, y]) \\ &= P(X \in g^{-1}(-\infty, y]) \\ &= \int_{g^{-1}(-\infty, y]} f_X(x) dx,\end{aligned}$$

wobei  $g^{-1}(A) = \{x \in \mathbb{R} : g(x) \in A\}$ . Durch Differenzieren kann man nun häufig  $f_Y(y)$  bestimmen.

**Beispiel 4.8.1** ( $\chi^2$ -Verteilung). Sei  $X \sim N(0, 1)$  und  $g(x) = x^2$ . Dann hat  $Y = g(X)$  die Verteilungsfunktion

$$\begin{aligned}P(Y \leq y) &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \\ &= 2\Phi(\sqrt{y}) - 1 \quad \text{für } y > 0.\end{aligned}$$

Differenzieren liefert

$$\begin{aligned} f_Y(y) &= 2\varphi(\sqrt{y}) \cdot \frac{1}{2}y^{-\frac{1}{2}} \\ &= y^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y\right). \end{aligned}$$

Man nennt  $Y$   $\chi^2$ -verteilt mit einem Freiheitsgrad, kurz  $Y \sim \chi^2(1)$ . Vergleicht man die Dichte von  $Y$  mit der Dichte der Gammaverteilung (vgl. Abschnitt 4.4.4), so sieht man, dass die  $\chi^2$ -Verteilung mit einem Freiheitsgrad offensichtlich ein Spezialfall der Gamma-verteilung ist:  $Y \sim G(\frac{1}{2}, \frac{1}{2})$ .

**Beispiel 4.8.2** (Lineare Transformationen). Die Funktion  $g(x) = ax + b$  sei eine lineare Transformation mit  $a, b \in \mathbb{R}$  und  $a \neq 0$ . Dann hat  $Y = g(X) = aX + b$  die Verteilungsfunktion

$$P(Y \leq y) = P(aX + b \leq y) = \begin{cases} P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right) & \text{für } a > 0 \\ P\left(X \geq \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right) & \text{für } a < 0. \end{cases}$$

Differenzieren liefert

$$f_Y(y) = |a|^{-1} f_X\left(\frac{y-b}{a}\right).$$

Insbesondere gilt für lineare Transformationen der Standardnormalverteilung:

$$X \sim N(0, 1) \implies Y = aX + b \sim N(b, a^2).$$

Bei streng monotonen Funktionen  $g(x)$  gilt allgemein:

**Satz 4.8.3** (Transformationsatz für univariate Dichtefunktionen). *Sei  $X$  eine stetige Zufallsvariable und  $g(x)$  streng monoton mit Umkehrfunktion  $g^{-1}(y)$  und stetiger Ableitung  $\frac{\partial g^{-1}(y)}{\partial y}$ . Dann hat die Zufallsvariable  $Y = g(X)$  die Dichtefunktion*

$$f_Y(y) = \left| \frac{\partial g^{-1}(y)}{\partial y} \right| \cdot f_X(g^{-1}(y)).$$

*Beweis.* Sei zunächst  $g(x)$  streng monoton wachsend. Dann ist auch  $g^{-1}(y)$  streng monoton wachsend, also  $\frac{\partial g^{-1}(y)}{\partial y} > 0$  und

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \leq g^{-1}(y)) \\ &= F_X(g^{-1}(y)). \end{aligned}$$

Differenzieren liefert

$$f_Y(y) = \frac{\partial g^{-1}(y)}{\partial y} \cdot f_X(g^{-1}(y)).$$

Falls  $g(x)$  streng monoton fallend ist, ist  $\frac{\partial g^{-1}(y)}{\partial y} < 0$  und somit

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \geq g^{-1}(y)) \\ &= 1 - F_X(g^{-1}(y)). \end{aligned}$$

Differenzieren liefert

$$\begin{aligned} f_Y(y) &= \frac{\partial g^{-1}(y)}{\partial y} \cdot (-f_X(g^{-1}(y))) \\ &= \left| \frac{\partial g^{-1}(y)}{\partial y} \right| \cdot f_X(g^{-1}(y)) \end{aligned}$$

□

Satz 4.8.3 ist äußerst wichtig. Viele Dichtefunktionen von stetigen Verteilungen können mit seiner Hilfe berechnet werden, wie folgende Beispiele zeigen:

**Beispiel 4.8.4** (Log-Normalverteilung). Sei  $X \sim N(\mu, \sigma^2)$ . Dann ist  $Y = \exp(X)$  *log-normalverteilt*. Hier ergibt sich:

$$g(x) = \exp(x) \implies g^{-1}(y) = \log(y) \implies \frac{\partial g^{-1}(y)}{\partial y} = \frac{1}{y}.$$

Damit folgt:

$$f_Y(y) = \frac{1}{y} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(\log(y) - \mu)^2}{\sigma^2}\right) \quad \text{für } y > 0.$$

Man kann zeigen, dass

$$\begin{aligned} E(Y) &= \exp\left(\mu + \frac{1}{2}\sigma^2\right), \\ \text{Var}(Y) &= \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1). \end{aligned}$$

**Beispiel 4.8.5** (Inverse Gammaverteilung). Sei  $X \sim G(a, b)$  mit  $a, b \in \mathbb{R}^+$ . Dann ist  $Y = X^{-1} \sim \text{IG}(a, b)$  *invers gammaverteilt*. Hier ergibt sich:

$$y = g(x) = \frac{1}{x} \implies g^{-1}(y) = \frac{1}{y} \implies \frac{\partial g^{-1}(y)}{\partial y} = -\frac{1}{y^2}.$$

Somit folgt

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \cdot \left| \frac{\partial g^{-1}(y)}{\partial y} \right| \\ &= \frac{b^a \cdot \left(\frac{1}{y}\right)^{a-1} \cdot \exp\left(-\frac{b}{y}\right)}{\Gamma(a)} \cdot \frac{1}{y^2} \\ &= \frac{b^a}{\Gamma(a)} \left(\frac{1}{y}\right)^{(a+1)} \cdot \exp\left(-\frac{b}{y}\right) \\ &= \frac{b^a}{\Gamma(a)} y^{-(a+1)} \cdot \exp\left(-\frac{b}{y}\right). \end{aligned}$$

Es gilt:

$$\mathbb{E}(Y) = \frac{b}{a-1} \quad \text{für } a > 1, \quad (26)$$

$$\text{Var}(Y) = \frac{b^2}{(a-1)^2(a-2)} \quad \text{für } a > 2. \quad (27)$$

- Zu (26):

$$\begin{aligned} \mathbb{E}(Y) &= \int_0^\infty y \cdot f_Y(y) dy \\ &= \frac{b^a}{\Gamma(a)} \int_0^\infty y \cdot y^{-(a+1)} \cdot \exp\left(-\frac{b}{y}\right) dy \\ &= \frac{b^a}{\Gamma(a)} \int_0^\infty y^{-a} \cdot \exp\left(-\frac{b}{y}\right) dy \\ &= \frac{b^a}{\Gamma(a)} \frac{\Gamma(a-1)}{b^{a-1}} \\ &= \frac{b}{a-1} \quad \text{für } a > 1, b > 0. \end{aligned}$$

- Zu (27): Bekanntermaßen gilt  $\text{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2$ , und mit

$$\begin{aligned} \mathbb{E}(Y^2) &= \int_0^\infty y^2 \cdot f_Y(y) dy \\ &= \frac{b^a}{\Gamma(a)} \int_0^\infty y^2 \cdot y^{-(a+1)} \cdot \exp\left(-\frac{b}{y}\right) dy \\ &= \frac{b^a}{\Gamma(a)} \int_0^\infty y^{-a-1} \cdot \exp\left(-\frac{b}{y}\right) dy \\ &= \frac{b^a}{\Gamma(a)} \frac{\Gamma(a-2)}{b^{a-2}} \\ &= \frac{b^2}{(a-1)(a-2)} \quad \text{für } a > 2, b > 0 \end{aligned}$$

ergibt sich:

$$\begin{aligned} \text{Var}(Y) &= \frac{b^2}{(a-1)^2(a-2)} - \frac{b^2}{(a-1)^2} \\ &= \frac{(a-1)b^2 - b^2(a-2)}{(a-1)^2(a-2)} \\ &= \frac{b^2}{(a-1)^2(a-2)} \end{aligned}$$

für  $a > 2, b > 0$ .

**Beispiel 4.8.6** (Wichtige Eigenschaft der Standard-Cauchy-Verteilung). Sei  $X$  standard-Cauchy-verteilt. Dann ist  $Y = \frac{1}{X}$  ebenfalls standard-Cauchy-verteilt, denn

$$y = g(x) = \frac{1}{x} \implies g^{-1}(y) = \frac{1}{y} \implies \frac{\partial g^{-1}(y)}{\partial y} = -\frac{1}{y^2}.$$

Somit folgt:

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \cdot \left| \frac{\partial g^{-1}(y)}{\partial y} \right| \\ &= \frac{1}{\pi} \cdot \frac{1}{1 + \left(\frac{1}{y}\right)^2} \cdot \left| -\frac{1}{y^2} \right| \\ &= \frac{1}{\pi} \cdot \frac{1}{1 + y^2}. \end{aligned}$$

Allgemeiner kann man auch folgendes Szenario betrachten: Sei  $\mathbf{X} = (X_1, X_2)^T$  ein bivariater Zufallsvektor mit gemeinsamer Dichtefunktion  $f_{X_1, X_2}(x_1, x_2)$ . Seien  $g$  und  $h$  Funktionen von  $\mathbb{R}^2 \rightarrow \mathbb{R}$ . Wie lautet die gemeinsame Dichte des Zufallsvektors  $\mathbf{Y} = (Y_1, Y_2)^T = (g(X_1, X_2), h(X_1, X_2))^T$ ?

Zur Lösung dieses Problems benötigt man die *Jacobi Determinante* einer eindeutigen Abbildung

$$T : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

mit einer Umkehrfunktion  $T^{-1}$ . Hierbei sind

$$\begin{aligned} x_1 &= x_1(y_1, y_2) \\ x_2 &= x_2(y_1, y_2) \end{aligned}$$

die Komponenten der Umkehrfunktion  $T^{-1}$ . Existieren die stetigen partiellen Ableitungen  $\frac{\partial x_i}{\partial y_j}$  für  $i, j \in \{1, 2\}$ , so ist die Jacobi Determinante definiert als

$$J = J(y_1, y_2) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \frac{\partial x_1}{\partial y_1} \frac{\partial x_2}{\partial y_2} - \frac{\partial x_2}{\partial y_1} \frac{\partial x_1}{\partial y_2}.$$

Mit der Jacobi (Determinante) lässt sich nun der Transformationssatz für bivariate Dichtefunktionen formulieren:

**Satz 4.8.7** (Transformationssatz für bivariate Dichtefunktionen). *Sei  $\mathbf{X} = (X_1, X_2)^T$  ein bivariater Zufallsvektor mit gemeinsamer Dichtefunktion  $f_{X_1, X_2}(x_1, x_2)$ . Dann hat  $\mathbf{Y} = (Y_1, Y_2)^T = T(X_1, X_2)$  die gemeinsame Dichtefunktion*

$$f_{Y_1, Y_2}(y_1, y_2) = \begin{cases} |J(y_1, y_2)| \cdot f_{X_1, X_2}(x_1(y_1, y_2), x_2(y_1, y_2)) & \text{für } (x_1(y_1, y_2), x_2(y_1, y_2)) \in \mathcal{T}_{X_1, X_2} \\ 0 & \text{sonst.} \end{cases}$$

Ähnliches erhält man für Transformationen  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ . Hierbei handelt es sich dann um den *Transformationssatz für multivariate Dichtefunktionen*. Hierfür sei auf entsprechende Literatur verwiesen.

**Beispiel 4.8.8.** Sei  $\mathbf{X} = (X_1, X_2)^T$  ein bivariater Zufallsvektor mit gemeinsamer Dichtefunktion  $f_{X_1, X_2}(x_1, x_2)$ . Weiter sei  $T^{-1}$  definiert durch

$$\begin{aligned} X_1 &= aY_1 + bY_2 \\ X_2 &= cY_1 + dY_2, \end{aligned}$$

wobei  $ad - bc \neq 0$ . Dann ist

$$f_{Y_1, Y_2}(y_1, y_2) = |ad - bc| \cdot f_{X_1, X_2}(ay_1 + by_2, cy_1 + dy_2).$$

**Beispiel 4.8.9.** Seien  $X, Y$  Zufallsvariablen mit gemeinsamer Dichtefunktion  $f_{X,Y}(x, y)$ . Dann hat  $U = X \cdot Y$  die Dichtefunktion

$$f_U(u) = \int_{-\infty}^{\infty} f_{X,Y}\left(x, \frac{u}{x}\right) |x|^{-1} dx.$$

*Beweis.* Sei  $T(x, y) = (xy, x)^T = (u, v)^T$ . Dann ist  $T^{-1}(u, v) = (v, \frac{u}{v})$  mit der Jacobi

$$J(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & \frac{1}{v} \\ 1 & -\frac{u}{v^2} \end{vmatrix} = -\frac{1}{v}.$$

Daher ist

$$f_{U,V}(u, v) = |v|^{-1} \cdot f_{X,Y}\left(v, \frac{u}{v}\right).$$

Durch Rausintegrieren von  $v$  erhält man das obige Ergebnis.  $\square$

**Beispiel 4.8.10.** Seien  $X_1, X_2$  unabhängige exponentialverteilte Zufallsvariablen mit Parameter  $\lambda$ . Dann sind

$$\begin{aligned} Y_1 &= X_1 + X_2 \\ Y_2 &= \frac{X_1}{X_2} \end{aligned}$$

unabhängig. Denn: Sei

$$T : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

mit

$$\begin{aligned} y_1 &= x_1 + x_2 \\ y_2 &= \frac{x_1}{x_2}. \end{aligned}$$

Die Komponenten der Umkehrfunktion  $T^{-1}$  ergeben sich zu

$$\begin{aligned} x_1 &= \frac{y_1 y_2}{1 + y_2} \\ x_2 &= \frac{y_1}{1 + y_2}, \end{aligned}$$

und man errechnet leicht folgende Jacobi:

$$J(y_1, y_2) = -\frac{y_1}{(1 + y_2)^2}.$$

Daher ist

$$\begin{aligned}
 f_{Y_1, Y_2}(y_1, y_2) &= \frac{y_1}{(1+y_2)^2} f_{X_1, X_2}\left(\frac{y_1 y_2}{1+y_2}, \frac{y_1}{1+y_2}\right) \\
 &\stackrel{X_1, X_2 \text{ unabh.}}{=} \frac{y_1}{(1+y_2)^2} \cdot f_{X_1}\left(\frac{y_1 y_2}{1+y_2}\right) \cdot f_{X_2}\left(\frac{y_1}{1+y_2}\right) \\
 &= \frac{y_1}{(1+y_2)^2} \cdot \lambda^2 \exp\left(-\lambda \left[\frac{y_1}{1+y_2}(y_2+1)\right]\right) \\
 &= \underbrace{\lambda^2 y_1 \exp(-\lambda y_1)}_{f_{Y_1}(y_1)} \cdot \underbrace{\frac{1}{(1+y_2)^2}}_{f_{Y_2}(y_2)}.
 \end{aligned}$$

$Y_1$  und  $Y_2$  sind unabhängig, da  $f_{Y_1, Y_2}(y_1, y_2) = f_{Y_1}(y_1) \cdot f_{Y_2}(y_2)$  gilt. In der Tat ist  $Y_1 \sim G(2, \lambda)$  und  $Y_2 \sim F(2, 2)$  eine  $F$ -Verteilung mit jeweils zwei Freiheitsgraden.

#### 4.9 Summen von Zufallsvariablen

**Satz 4.9.1** (Faltungssatz). *Falls  $X$  und  $Y$  gemeinsame Dichtefunktion  $f_{X, Y}(x, y) = f(x, y)$  besitzen, so hat  $Z = X + Y$  die Dichtefunktion*

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X, Y}(x, z-x) dx = \int_{-\infty}^{\infty} f_{X, Y}(z-y, y) dy.$$

*Beweis.* Sei  $A = \{(x, y) : x + y \leq z\}$ . Dann ist

$$\begin{aligned}
 F_Z(z) &= P(X + Y \leq z) \\
 &= \iint_A f(u, v) du dv \\
 &= \int_{u=-\infty}^{\infty} \int_{v=-\infty}^{z-u} f(u, v) dv du \\
 &\stackrel{\text{Subst.: } \underline{x=u}, \underline{y=v+u}}{=} \int_{x=-\infty}^{\infty} \int_{y=-\infty}^z f(x, y-x) dy dx \\
 &= \int_{y=-\infty}^z \int_{x=-\infty}^{\infty} f(x, y-x) dx dy.
 \end{aligned}$$

Somit ist  $\int_{x=-\infty}^{\infty} f(x, z-x) dx$  die zugehörige Dichtefunktion  $f_Z(z)$  von  $Z$ . □

Sind  $X$  und  $Y$  zusätzlich unabhängig, erhält man

$$\begin{aligned}
 f_{X+Y}(z) &= \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x) dx \\
 &= \int_{-\infty}^{\infty} f_X(z-y) \cdot f_Y(y) dy.
 \end{aligned}$$

In Analogie zum diskreten Fall nennt man  $f_{X+Y}$  die *Faltung* von  $f_X$  und  $f_Y$ .

**Beispiel 4.9.2.** Seien  $X \sim N(0, 1)$  und  $Y \sim N(0, 1)$  unabhängig. Dann hat  $Z = X + Y$  die Dichte

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(z-x)^2\right) dx \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + (z-x)^2)\right) dx. \end{aligned}$$

Für die weitere Berechnung ist folgende Überlegung zu  $(x^2 + (z-x)^2)$  hilfreich:

$$\begin{aligned} x^2 + (z-x)^2 &= x^2 + z^2 - 2zx + x^2 \\ &= 2x^2 + z^2 - 2zx \\ &= 2\left(x^2 - zx + \frac{1}{2}z^2\right) \\ &= 2\left(x - \frac{1}{2}z\right)^2 + \frac{1}{2}z^2 \\ &\stackrel{v=\sqrt{2}\left(x-\frac{1}{2}z\right)}{=} v^2 + \frac{1}{2}z^2 \end{aligned}$$

Somit ergibt sich mit der Substitution  $v = g(x) = \sqrt{2}\left(x - \frac{1}{2}z\right)$ , also  $g'(x) = \sqrt{2}$ ,

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + (z-x)^2)\right) dx \\ &= \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + (z-x)^2)\right) \sqrt{2} dx \\ &= \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{1}{4}z^2\right) \cdot \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v^2\right) dv}_{=1} \\ &= \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{1}{4}z^2\right). \end{aligned}$$

Vergleicht man dies mit der Dichtefunktion der Normalverteilung, so erkennt man, dass

$$Z \sim N(0, 2).$$

Allgemein gilt für zwei unabhängige normalverteilte Zufallsgrößen:

**Satz 4.9.3** (Summe von Normalverteilungen). *Seien  $X_i \sim N(\mu_i, \sigma_i^2)$ ,  $i = 1, 2, \dots, n$ , unabhängige normalverteilte Zufallsvariablen. Dann gilt für  $Z = \sum_{i=1}^n X_i$ :*

$$Z \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

**Satz 4.9.4** (Summe von Gammaverteilungen). *Seien  $X_i \sim G(\alpha_i, \beta)$ ,  $i = 1, 2, \dots, n$ , unabhängige gammaverteilte Zufallsvariablen. Dann gilt für  $Z = \sum_{i=1}^n X_i$ :*

$$Z \sim G\left(\sum_{i=1}^n \alpha_i, \beta\right).$$

**Beispiel 4.9.5.** Die Anwendung von Satz 4.9.1 wird deutlich komplizierter, wenn der Träger von  $f(x, y)$  von  $x$  und  $y$  abhängt. Sei nun

$$f(x, y) = \begin{cases} \frac{1}{x} & \text{für } 0 \leq y \leq x \leq 1 \\ 0 & \text{sonst.} \end{cases}$$

Dann ist  $f_{X+Y}(z) = \int_A \frac{1}{x} dx$  für  $0 \leq z \leq 2$ , und für

$$A = \{x : 0 \leq z - x \leq x \leq 1\} = \left[ \frac{1}{2}z, \min(z, 1) \right].$$

Somit ergibt sich für  $0 \leq z \leq 1$

$$\begin{aligned} f_Z(z) &= \int_{\frac{1}{2}z}^z \frac{1}{x} dx \\ &= [\log(x)]_{\frac{1}{2}z}^z \\ &= \log(2), \end{aligned}$$

und für  $1 \leq z \leq 2$

$$\begin{aligned} f_Z(z) &= \int_{\frac{1}{2}z}^1 \frac{1}{x} dx \\ &= [\log(x)]_{\frac{1}{2}z}^1 \\ &= \log\left(\frac{2}{z}\right). \end{aligned}$$

**Beispiel 4.9.6** (Entfaltung, engl. deconvolution). Seien  $X \sim N(\mu, \sigma^2)$  und  $Y \sim N(\nu, \tau^2)$  unabhängig. Nach Beobachtung der Summe  $Z = X + Y$  möchte man nun Rückschlüsse auf  $X$  (bzw.  $Y$ ) ziehen. Da  $Z \sim N(\mu + \nu, \sigma^2 + \tau^2)$  und  $\text{Cov}(X, Z) = \sigma^2$ , lässt sich über Definition 4.6.11 zeigen, dass

$$X|\{Z = z\} \sim N\left(\mu + \frac{\sigma^2}{\sigma^2 + \tau^2}(z - (\mu + \nu)), \sigma^2 \left(1 - \frac{\sigma^2}{\sigma^2 + \tau^2}\right)\right).$$

## 4.10 Multivariate Normalverteilung

Der Kern der Dichte der *univariaten Standardnormalverteilung* ist  $\exp(-\frac{1}{2}x^2)$ , also eine exponentierte quadratische Form in  $x$ . Der Kern der Dichte der *bivariaten Standardnormalverteilung* ist  $\exp(-\frac{1}{2}(x^2 - 2\rho xy + y^2))$ , also eine exponentierte quadratische Form in  $x$  und  $y$  (vgl. Definition 4.6.11).

Die Verallgemeinerung zur *n-dimensionalen Normalverteilung* von Zufallsvariablen  $X_1, X_2, \dots, X_n$  gelingt daher über die Einführung einer *quadratischen Form*

$$Q(\mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = \mathbf{x}^T \mathbf{A} \mathbf{x},$$

wobei  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$  und die  $(n \times n)$ -Matrix  $\mathbf{A} = (a_{ij})$  reell und symmetrisch sei mit  $\det(\mathbf{A}) = |\mathbf{A}| \neq 0$ .  $\mathbf{A}$  heißt *positiv definit* (kurz:  $\mathbf{A} > 0$ ), falls  $Q(\mathbf{x}) > 0$  für alle  $\mathbf{x} \neq \mathbf{0}$ .

**Satz 4.10.1.** *Es gilt:*

$$\mathbf{A} > 0 \iff \lambda_i > 0 \text{ für alle } i = 1, \dots, n,$$

wobei  $\lambda_i$  die Eigenwerte der Matrix  $\mathbf{A}$  sind.

*Beweis.* Zu  $\mathbf{A}$  gibt es eine Zerlegung

$$\begin{matrix} \mathbf{A} \\ (n \times n) \end{matrix} = \begin{matrix} \mathbf{B} \\ (n \times n) \end{matrix} \begin{matrix} \mathbf{\Lambda} \\ (n \times n) \end{matrix} \begin{matrix} \mathbf{B}^T \\ (n \times n) \end{matrix},$$

wobei  $\mathbf{B}$  orthogonal ist und  $\mathbf{\Lambda}$  diagonal mit Elementen  $\lambda_i, i = 1, 2, \dots, n$ , auf der Diagonale. Daher:

$$\begin{aligned} Q(\mathbf{x}) &= \mathbf{x}^T \mathbf{A} \mathbf{x} \\ &= \underbrace{\mathbf{x}^T \mathbf{B}}_{\mathbf{y}^T} \mathbf{\Lambda} \underbrace{\mathbf{B}^T \mathbf{x}}_{\mathbf{y}} \\ &\quad \begin{matrix} (1 \times n) & & (n \times 1) \end{matrix} \\ &= \sum_{i=1}^n \lambda_i y_i^2 \end{aligned}$$

und  $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$  für  $\mathbf{x} \neq \mathbf{0}$  genau dann, wenn  $\lambda_i > 0$  für alle  $i = 1, 2, \dots, n$ . □

Hieraus ergibt sich folgende Frage: Wann ist

$$f(\mathbf{x}) = K \cdot \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x}\right)$$

für  $\mathbf{x} \in \mathbb{R}^n$  eine legitime Dichtefunktion für den Zufallsvektor  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ ?  
Wir überprüfen Positivität und Normiertheit:

1.  $f(\mathbf{x}) \geq 0$  für alle  $\mathbf{x} \in \mathbb{R}^n \Rightarrow K > 0$
- 2.

$$\begin{aligned} \int_{\mathbb{R}^n} f(\mathbf{x}) d\mathbf{x} &= K \cdot \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x}\right) d\mathbf{x} \\ &= K \cdot \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2} \sum_{i=1}^n \lambda_i y_i^2\right) d\mathbf{y} \\ &= K \cdot \prod_{i=1}^n \underbrace{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \lambda_i y_i^2\right) dy_i}_{\text{Kern der } N(0, \lambda_i^{-1})} \\ &= K \cdot \prod_{i=1}^n \frac{\sqrt{2\pi}}{\sqrt{\lambda_i}} \\ \stackrel{\det(\mathbf{A})=|\mathbf{A}|=\prod_{i=1}^n \lambda_i}{=} & K \cdot \frac{(2\pi)^{\frac{n}{2}}}{\sqrt{|\mathbf{A}|}} \\ \stackrel{!}{=} & 1 \iff K = \sqrt{|\mathbf{A}|} \cdot \frac{1}{(2\pi)^{\frac{n}{2}}}. \end{aligned}$$

Aus 1. und 2. folgt:

$f(\mathbf{x})$  ist Dichte  $\iff \mathbf{A}$  ist positiv definit.

Somit ergibt sich:

$$f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\mathbf{A}|^{\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x}\right).$$

Man zeigt leicht, dass  $E(\mathbf{X}) = \mathbf{0}$ , d.h.  $E(X_i) = 0$  für alle  $i = 1, \dots, n$ .

**Definition 4.10.2** (Multivariate Normalverteilung). Ein Zufallsvektor  $\mathbf{X} = (X_1, \dots, X_n)^T$  hat eine *multivariate Normalverteilung* mit Parametern  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$  und  $\boldsymbol{\Sigma} = \mathbf{A}^{-1}$ , falls die Dichtefunktion von  $\mathbf{X}$  die Form

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

hat. Notation:  $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

**Satz 4.10.3.** Sei  $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Dann gilt

1.  $E(\mathbf{X}) = \boldsymbol{\mu}$ ,
2.  $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$ .

Somit können wir die Parameter  $\boldsymbol{\mu}$  als *Erwartungswertvektor* und  $\boldsymbol{\Sigma}$  als *Kovarianzmatrix* von  $\mathbf{X}$  identifizieren. Alternativ kann man auch die *Präzisionsmatrix*  $\mathbf{A} = \boldsymbol{\Sigma}^{-1}$  an Stelle der Kovarianzmatrix zur Parametrisierung der Normalverteilung verwenden.

**Satz 4.10.4** (Lineare Transformationen von multivariaten Normalverteilungen). Sei  $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ,  $\mathbf{D}$  eine  $(m \times n)$ -Matrix vom Rang  $m \leq n$  und  $\mathbf{b} \in \mathbb{R}^m$ . Dann gilt

$$\mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{b} \sim N_m(\mathbf{D}\boldsymbol{\mu} + \mathbf{b}, \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^T).$$

Es folgen einige Anwendungen und Spezialfälle des Satzes 4.10.4. Dazu benötigen wir die Zerlegung der Kovarianzmatrix  $\boldsymbol{\Sigma}$  in  $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T$  für eine  $(n \times n)$ -Matrix  $\mathbf{L}$ . Beispielsweise liefert die *Cholesky-Zerlegung* eine (eindeutige) untere Dreiecksmatrix  $\mathbf{L}$  mit dieser Eigenschaft. Aus  $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T$  folgt sofort

$$\boldsymbol{\Sigma}^{-1} = (\mathbf{L}^T)^{-1} \mathbf{L}^{-1} = (\mathbf{L}^{-1})^T \mathbf{L}^{-1},$$

d.h. es folgt  $\boldsymbol{\Sigma}^{-1} = \mathbf{M}\mathbf{M}^T$  mit  $\mathbf{M} = (\mathbf{L}^{-1})^T$ .

1. **Simulation einer multivariaten Normalverteilung aus einer multivariaten Standardnormalverteilung.**

Sei  $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{I})$  multivariat standardnormalverteilt ( $\mathbf{I}$  ist die Einheitsmatrix) und  $\mathbf{Y} = \mathbf{L}\mathbf{X} + \boldsymbol{\mu}$ . Dann ist  $\mathbf{Y} \sim N_n(\mathbf{L} \cdot \mathbf{0} + \boldsymbol{\mu}, \mathbf{L}\mathbf{I}\mathbf{L}^T) = N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ -verteilt.

2. **Standardisierung von multivariaten Normalverteilungen.**

Sei  $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  und  $\mathbf{Y} = \mathbf{M}^T(\mathbf{X} - \boldsymbol{\mu})$ . Dann ist  $\mathbf{Y} \sim N_n(\mathbf{M}^T(\boldsymbol{\mu} - \boldsymbol{\mu}), \mathbf{M}^T \boldsymbol{\Sigma} \mathbf{M}) = N_n(\mathbf{0}, (\mathbf{L}^{-1}) \mathbf{L} \mathbf{L}^T (\mathbf{L}^{-1})^T) = N_n(\mathbf{0}, \mathbf{I})$ -verteilt.

### 3. Linearkombination von Normalverteilungen.

Sei  $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  mit  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$  und  $\boldsymbol{\Sigma} = (\sigma_{ij})_{ij}$ , wobei  $\sigma_{ii} = \sigma_i^2$  ist. Dann ist  $Y = \mathbf{a}^T \mathbf{X} + b = \sum_{i=1}^n a_i X_i + b$  wieder normalverteilt mit

$$\begin{aligned} E(Y) &= \mathbf{a}^T E(\mathbf{X}) + b \\ &= \mathbf{a}^T \boldsymbol{\mu} + b \\ &= \sum_{i=1}^n a_i \mu_i + b \end{aligned}$$

und

$$\begin{aligned} \text{Var}(Y) &= \mathbf{a}^T \boldsymbol{\Sigma} \mathbf{a} \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \sigma_{ij} \\ &= \sum_{i=1}^n a_i^2 \sigma_i^2 + 2 \sum_{i < j} a_i a_j \sigma_{ij}. \end{aligned}$$

Abschließend folgt ein Satz, der als Verallgemeinerung von Beispiel 4.7.8 angesehen werden kann.

**Satz 4.10.5.** Sei  $\mathbf{X} \sim N_{m+n}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  multivariat normalverteilt, wobei  $\mathbf{X} = (\mathbf{X}_1^T, \mathbf{X}_2^T)^T$  in die  $m$ - bzw.  $n$ -dimensionalen Vektoren  $\mathbf{X}_1$  und  $\mathbf{X}_2$  partitioniert sei. Sei ferner  $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T)^T$  und

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

analog partitioniert. Dann gilt für die Randverteilung von  $\mathbf{X}_1$

$$\mathbf{X}_1 \sim N_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$$

und für die bedingte Verteilung von  $\mathbf{X}_1$  gegeben  $\mathbf{X}_2 = \mathbf{x}_2$

$$\mathbf{X}_1 | \{\mathbf{X}_2 = \mathbf{x}_2\} \sim N_m(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2})$$

mit

$$\begin{aligned} \boldsymbol{\mu}_{1|2} &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ \text{und } \boldsymbol{\Sigma}_{1|2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}. \end{aligned}$$

### 4.11 Von der Normalverteilung abgeleitete Verteilungen

**Satz 4.11.1.** Seien  $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$  unabhängig. Dann gilt

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{1}{n} \sigma^2\right) \\ S^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \sim G\left(\frac{n-1}{2}, \frac{n-1}{2\sigma^2}\right). \end{aligned}$$

Alternativ findet man auch folgende äquivalente Aussage für  $S^2$ :

$$\frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2$$

bzw.

$$S^2 \sim \frac{\sigma^2}{n-1} \chi_{n-1}^2.$$

Es fällt auf, dass

$$E(\bar{X}) = \mu$$

und

$$E(S^2) = \frac{\frac{n-1}{2}}{\frac{n-1}{2\sigma^2}} = \sigma^2$$

gilt. Somit sind die Schätzer  $\bar{X}$  und  $S^2$  erwartungstreu für  $\mu$  und  $\sigma^2$ . Weiterhin gilt:

**Satz 4.11.2.** Seien  $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$  unabhängig. Dann sind  $\bar{X}$  und  $S^2$  unabhängig.

*Beweis.* Siehe Buch von Grimmett und Stirzaker auf Seite 120. □

Im Folgenden wollen wir noch anmerken, dass eine bestimmte Funktion von  $\bar{X}$   $t$ -verteilt ist. Dazu betrachten wir die beiden Zufallsvariablen

$$U = \frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2$$

und

$$V = \frac{\sqrt{n}}{\sigma} (\bar{X} - \mu) \sim N(0, 1)$$

und bilden

$$T = \frac{V}{\sqrt{\frac{U}{n-1}}}.$$

Dann ist  $T$   $t$ -verteilt mit  $n-1$  Freiheitsgraden:  $T \sim t(n-1)$ .

Abschließend betrachten wir eine häufige Konstruktion der  $F$ -Verteilung: Seien  $U \sim \chi_r^2$  und  $V \sim \chi_s^2$  unabhängige Zufallsvariablen. Dann heißt

$$F = \frac{\frac{U}{r}}{\frac{V}{s}}$$

$F$ -verteilt mit den Freiheitsgraden  $r$  und  $s$ . Notation:  $F \sim F(r, s)$ .  $F$  ist also der Quotient aus einer  $G\left(\frac{r}{2}, \frac{r}{2}\right)$ -verteilten und einer davon unabhängigen  $G\left(\frac{s}{2}, \frac{s}{2}\right)$ -verteilten Zufallsvariable. Die  $F$ -Verteilung wird häufig im Rahmen des linearen Modells und der Varianzanalyse zum Testen verwendet. Auf folgende zwei Eigenschaften sei hingewiesen:

1.  $F^{-1} \sim F(s, r)$ .
2. Falls  $T \sim t(r)$ , dann  $T^2 \sim F(1, r)$ .

## 4.12 Simulation aus stetigen Verteilungen

**Ziel** ist die Erzeugung von "unabhängigen" (Pseudo-)Zufallszahlen am Computer aus einer Verteilung mit der Dichtefunktion  $f(x)$  bzw. mit der Verteilungsfunktion  $F(x)$ . Als **Basis** dienen "unabhängige" gleichverteilte (Pseudo-)Zufallsvariablen  $U_i \sim U(0, 1)$ ,  $i \in \mathbb{N}$ .

Das Inversionsverfahren, welches wir bereits in Satz 3.3.16 für diskrete Zufallsvariablen kennengelernt haben, ist für stetige Zufallsvariablen auf Grund der im Allgemeinen stetigen und streng monotonen Verteilungsfunktion wesentlich einfacher zu beweisen.

**Satz 4.12.1** (Inversionsverfahren). *Sei die Verteilungsfunktion  $F_X(x)$  stetig und streng monoton auf dem Träger. Dann ist die Quantilsfunktion  $F_X^-(u)$  eindeutig definiert, und es gilt für  $U \sim U(0, 1)$ :*

$$F_X^-(U) \sim F_X,$$

d.h.  $X$  hat Verteilungsfunktion  $F_X$ .

*Beweis.* Entweder über den Transformationssatz für Dichten oder direkt:

$$\begin{aligned} P(F_X^-(U) \leq x) &= P(U \leq F_X(x)) \\ &= F_X(x), \end{aligned}$$

da  $U \sim U(0, 1)$ . □

**Beispiel 4.12.2** (Exponentialverteilung). Sei  $F_X(x) = 1 - \exp(-\lambda x)$  mit  $\lambda > 0$ , d.h.  $F_X(x)$  ist stetig und streng monoton. Dann berechnet man die Quantilsfunktion von  $F_X(x)$  über

$$\begin{aligned} u &= 1 - \exp(-\lambda x) \\ \Leftrightarrow \exp(-\lambda x) &= 1 - u \\ \Leftrightarrow -\lambda x &= \log(1 - u) \\ \Leftrightarrow x &= -\frac{1}{\lambda} \log(1 - u), \end{aligned}$$

also  $F_X^-(u) = -\frac{1}{\lambda} \log(1 - u)$ . Unter Beachtung von  $U \sim U(0, 1) \Rightarrow 1 - U \sim U(0, 1)$  folgt nach dem Inversionsverfahren:

$$X = -\frac{1}{\lambda} \log(U)$$

hat Verteilung  $F_X$ , d.h. ist exponentialverteilt mit Parameter  $\lambda$ .

**Beispiel 4.12.3** (Standard-Cauchy-Verteilung). Gegeben sei die Dichtefunktion

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

Die zugehörige Verteilungsfunktion und deren Quantilsfunktion ergeben sich als

$$\begin{aligned} F(x) &= \int_{-\infty}^x \frac{1}{\pi(1+u^2)} du \\ &= \frac{1}{\pi} [\arctan(u)]_{-\infty}^x \\ &= \frac{1}{\pi} \arctan(x) + \frac{1}{2} \end{aligned}$$

und  $F^-(u) = \tan\left(\pi\left(u - \frac{1}{2}\right)\right).$

Also ist  $X = \tan(\pi(U - \frac{1}{2}))$ , der Tangens einer  $U(-\pi/2, \pi/2)$ -verteilten Zufallsvariable, Standard-Cauchy-verteilt.

Leider ist eine Verteilungsfunktion bzw. deren Inverse häufig nicht analytisch zugänglich. Beispiele hierfür sind:

- Normalverteilung,
- Gammaverteilung,
- Betaverteilung.

Zur Simulation dieser Verteilungen finden andere Verfahren Anwendung, die auf der Dichtefunktion  $f_X(x)$  an Stelle von  $F_X(x)$  basieren. Am wichtigsten ist das *Rejection Sampling* (*Verwerfungsmethode*) zum Erzeugen von Zufallszahlen aus einer stetigen Verteilung mit Dichte  $f_X(x)$ .

Hierzu verwendet man zwei unabhängige Zufallsvariablen  $U$  und  $Z$ , wobei  $U \sim U(0, 1)$  und  $Z \sim f_Z$ . Die Dichte  $f_Z$  ist beliebig zu wählen, es muss jedoch ein  $a \geq 1$  existieren mit

$$f_X(z) \leq a \cdot f_Z(z) \quad \text{für alle } z \in \mathbb{R}.$$

Nun gilt:

$$\begin{aligned} & P(Z \leq x | a \cdot U \cdot f_Z(Z) \leq f_X(Z)) \\ = & \frac{P(Z \leq x, a \cdot U \cdot f_Z(Z) \leq f_X(Z))}{P(a \cdot U \cdot f_Z(Z) \leq f_X(Z))} \\ = & \frac{\int_{-\infty}^x P(a \cdot U \cdot f_Z(Z) \leq f_X(Z) | Z = z) f_Z(z) dz}{\int_{-\infty}^{+\infty} P(a \cdot U \cdot f_Z(Z) \leq f_X(Z) | Z = z) f_Z(z) dz}. \end{aligned} \quad (28)$$

Nun ist

$$P(a \cdot U \cdot f_Z(Z) \leq f_X(Z) | Z = z) = P\left(U \leq \frac{f_X(z)}{a \cdot f_Z(z)}\right) = \frac{f_X(z)}{a \cdot f_Z(z)},$$

da nach Annahme  $f_X(z)/(a \cdot f_Z(z)) \leq 1$  und  $U \sim U(0, 1)$ . Somit ist

$$\begin{aligned} (28) &= \frac{\int_{-\infty}^x \frac{f_X(z)}{a \cdot f_Z(z)} f_Z(z) dz}{\int_{-\infty}^{+\infty} \frac{f_X(z)}{a \cdot f_Z(z)} f_Z(z) dz} \\ &= \frac{\int_{-\infty}^x f_X(z) dz}{\int_{-\infty}^{+\infty} f_X(z) dz} \\ &= \int_{-\infty}^x f_X(z) dz \\ &= F_X(x), \end{aligned}$$

d.h. bedingt auf das Ereignis  $E = \{a \cdot U \cdot f_Z(Z) \leq f_X(Z)\}$  hat die Zufallsvariable  $Z$  die gewünschte Verteilungsfunktion  $F_X$  mit zugehöriger Dichtefunktion  $f_X$ .

Der *Rejection Sampling Algorithmus* zur Erzeugung von  $X \sim f_X$  lautet also:

1. Erzeuge unabhängige Zufallsvariablen  $Z \sim f_Z$  und  $U \sim U(0, 1)$ .
2. Falls  $U \leq \frac{f_X(Z)}{a \cdot f_Z(Z)}$ , setze  $X = Z$  (*acceptance step*).
3. Ansonsten gehe zurück zu 1. (*rejection step*).

**Bemerkungen:**

1. Die Größe  $\alpha(z) = f_X(z)/(a \cdot f_Z(z))$  nennt man *Akzeptanzwahrscheinlichkeit*, da die Realisation  $Z = z$  aus  $f_Z$  mit genau dieser Wahrscheinlichkeit als Zufallszahl aus  $f_X$  akzeptiert wird.
2. Jedes Paar  $(U, Z)$  erfüllt die Bedingung  $E$  mit Wahrscheinlichkeit  $a^{-1}$ :

$$\begin{aligned}
 P(a \cdot U \cdot f_Z(Z) \leq f_X(Z)) &= \int_{-\infty}^{\infty} \frac{f_X(z)}{a \cdot f_Z(z)} f_Z(z) dz \\
 &= \int_{-\infty}^{\infty} \frac{f_X(z)}{a} dz \\
 &= a^{-1}.
 \end{aligned}$$

Da die einzelnen Versuche unabhängig sind, folgt, dass die Anzahl der Versuche bis zum ersten Erfolg geometrisch verteilt ist. Also ist die erwartete Anzahl von Versuchen bis zum ersten Erfolg gleich  $a$ .

**Beispiel 4.12.4** (Rejection Sampling aus der Normalverteilung). Im Folgenden wollen wir mit Hilfe der Standard-Cauchy-Verteilung Zufallszahlen aus der Standardnormalverteilung erzeugen. Somit ist

$$\begin{aligned}
 f_X(x) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) && \text{(Standardnormalverteilung),} \\
 f_Z(z) &= \frac{1}{\pi} \cdot \frac{1}{1+z^2} && \text{(Standard-Cauchy-Verteilung).}
 \end{aligned}$$

Man erkennt leicht (zum Beispiel durch Ableiten), dass

$$a = \sup_{x \in \mathbb{R}} \frac{f_X(x)}{f_Z(x)} = \sqrt{\frac{\pi}{2}} \cdot \sup_{x \in \mathbb{R}} \left( (1+x^2) \exp\left(-\frac{1}{2}x^2\right) \right) = \sqrt{\frac{\pi}{2}} \cdot \frac{2}{\sqrt{e}} = \sqrt{\frac{2\pi}{e}} \approx 1.52$$

eine geeignete Konstante ist. Die Akzeptanzwahrscheinlichkeit ergibt sich zu

$$\alpha(z) = \frac{f_X(z)}{a \cdot f_Z(z)} = \frac{\sqrt{e}}{2} (1+z^2) \exp\left(-\frac{1}{2}z^2\right) \cdot \cdot$$

## 5 Erzeugende Funktionen und deren Anwendungen

### 5.1 Wahrscheinlichkeitserzeugende Funktionen

Die Idee von *erzeugenden Funktionen* (engl. *generating functions*) ist die Speicherung der Information einer Folge  $a = \{a_i : i = 0, 1, \dots\}$  von reellen Zahlen in einer Funktion, zum Beispiel in der (*gewöhnlichen*) *erzeugenden Funktion*  $G_a$ , welche durch

$$G_a(s) = \sum_{i=0}^{\infty} a_i s^i,$$

für alle  $s \in \mathbb{R}$  definiert wird, für die die Summe konvergiert.

Bezeichne mit  $G_a^{(j)}(s)$  die  $j$ -te Ableitung von  $G_a(s)$ . Dann ergibt sich

$$\begin{aligned} G_a^{(1)}(s) &= \sum_{i=1}^{\infty} i \cdot a_i \cdot s^{i-1} \\ G_a^{(2)}(s) &= \sum_{i=2}^{\infty} i \cdot (i-1) \cdot a_i \cdot s^{i-2} \\ &\vdots \\ G_a^{(j)}(s) &= \sum_{i=j}^{\infty} i \cdot (i-1) \cdot \dots \cdot (i-j+1) \cdot a_i \cdot s^{i-j}. \end{aligned}$$

Insbesondere ist offensichtlich

$$G_a^{(j)}(0) = j! a_j \quad \implies \quad a_j = \frac{1}{j!} G_a^{(j)}(0).$$

Man kann also die Folge  $a = \{a_i : i = 0, 1, \dots\}$  über die Ableitungen der erzeugenden Funktion rekonstruieren.

Ein weiteres Beispiel für eine erzeugende Funktion ist die *exponential-erzeugende Funktion*

$$E_a(s) = \sum_{i=0}^{\infty} a_i \frac{s^i}{i!}, \tag{29}$$

für alle  $s \in \mathbb{R}$ , für die die Summe konvergiert. Hier gilt offensichtlich:

$$E_a^{(j)}(0) = a_j.$$

In der Stochastik wird häufig die erzeugende Funktion der Wahrscheinlichkeitsfunktion  $f(i) = P(X = i)$  einer diskreten Zufallsvariable  $X$  betrachtet. Unter anderem liefert dies eine einfache Art zur Berechnung von Faltungen.

**Definition 5.1.1** (Faltung). Die *Faltung* von zwei reellen Folgen  $a = \{a_i : i = 0, 1, \dots\}$  und  $b = \{b_i : i = 0, 1, \dots\}$  ist die Folge  $c = \{c_i : i = 0, 1, \dots\}$  mit

$$c_n = a_0 b_n + a_1 b_{n-1} + \dots + a_n b_0 = \sum_{i=0}^n a_i b_{n-i}.$$

Man schreibt  $c = a * b$ .

Für die gewöhnliche erzeugende Funktion von  $c = a * b$  gilt

$$G_c(s) = G_a(s) \cdot G_b(s), \quad (30)$$

denn

$$\begin{aligned} G_c(s) &= \sum_{n=0}^{\infty} c_n s^n \\ &= \sum_{n=0}^{\infty} \sum_{i=0}^n a_i b_{n-i} s^i s^{n-i} \\ &= \sum_{i=0}^{\infty} \sum_{n=i}^{\infty} a_i b_{n-i} s^i s^{n-i} \\ &= \sum_{i=0}^{\infty} a_i s^i \sum_{n=i}^{\infty} b_{n-i} s^{n-i} \\ &= \sum_{i=0}^{\infty} a_i s^i \sum_{j=0}^{\infty} b_j s^j \\ &= G_a(s) \cdot G_b(s). \end{aligned}$$

**Beispiel 5.1.2.** Seien  $X$  und  $Y$  unabhängige Poissonverteilte Zufallsvariablen mit Parametern  $\lambda > 0$  bzw.  $\mu > 0$ . Wie lautet die Verteilung von  $Z = X + Y$ ?

Bisher haben wir eine solche Aufgabe mit dem Faltungssatz (Satz 3.8.2) gelöst:

$$f_{X+Y}(z) = \sum_x f_X(x) f_Y(z-x).$$

Die Wahrscheinlichkeitsfunktion  $f_{X+Y}$  ist also die Faltung der Wahrscheinlichkeitsfunktionen  $f_X$  und  $f_Y$ . Daher liegt es nahe, die erzeugenden Funktionen der Wahrscheinlichkeitsfunktionen  $f_X(i)$ ,  $f_Y(i)$  und  $f_Z(i)$  zu betrachten. Die von  $X$  lautet:

$$\begin{aligned} G_X(s) &= \sum_{i=0}^{\infty} f_X(i) \cdot s^i \\ \implies G_X(s) &= \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \exp(-\lambda) \cdot s^i \\ &= \exp(-\lambda) \underbrace{\sum_{i=0}^{\infty} \frac{(\lambda s)^i}{i!}}_{\exp(\lambda s)} \\ &= \exp(\lambda(s-1)), \end{aligned}$$

und analog ergibt sich für  $G_Y(s)$

$$G_Y(s) = \exp(\mu(s-1)).$$

Somit erhält man nach (30)

$$\begin{aligned} G_Z(s) &= G_X(s) \cdot G_Y(s) \\ &= \exp((\lambda + \mu)(s - 1)) \end{aligned}$$

und daher offensichtlich

$$Z = X + Y \sim \text{Po}(\lambda + \mu).$$

Die Berechnung der Verteilung von  $Z = X + Y$  ist also über die erzeugenden Funktionen der jeweiligen Wahrscheinlichkeitsfunktionen deutlich einfacher als über den Faltungssatz 3.8.2 (vgl. Beispiel 3.8.3). Man nennt die erzeugende Funktion einer Wahrscheinlichkeitsfunktion die *wahrscheinlichkeitserzeugende Funktion*.

**Definition 5.1.3** (Wahrscheinlichkeitserzeugende Funktion). Sei  $X$  eine diskrete Zufallsvariable mit Träger  $\mathcal{T} \subseteq \mathbb{N}_0 = \{0, 1, 2, \dots\}$  und Wahrscheinlichkeitsfunktion  $f(i) = P(X = i)$ . Die *wahrscheinlichkeitserzeugende Funktion* (WeF) (engl. *probability generating function; pgf*) der Zufallsvariable  $X$  ist die erzeugende Funktion der Wahrscheinlichkeitsfunktion  $f_X(i)$  von  $X$ :

$$G_X(s) = \sum_{i=0}^{\infty} f_X(i) s^i \stackrel{\text{Lemma 3.3.3}}{=} \mathbb{E}(s^X).$$

Wir werden gleich sehen, dass  $G_X(s)$  zumindest für  $|s| \leq 1$  konvergiert. Dafür benötigen wir einige Eigenschaften von Potenzreihen. Sei  $G_a(s) = \sum_{i=0}^{\infty} a_i s^i$  eine Potenzreihe, wobei  $a = \{a_i : i = 0, 1, 2, \dots\}$  eine Folge reeller Zahlen sei.

**Satz 5.1.4.** *Es gibt einen Konvergenzradius  $R \geq 0$ , so dass  $G_a(s)$  für*

$$\begin{aligned} |s| < R & \text{ absolut konvergiert,} \\ |s| > R & \text{ divergiert.} \end{aligned}$$

$G_a(s)$  ist gleichmäßig konvergent auf allen Mengen  $\{s : |s| \leq R'\}$  mit  $R' < R$ .

**Satz 5.1.5.**  $G_a(s)$  kann beliebig oft im Punkte  $s$  differenziert werden, falls  $|s| < R$ .

**Satz 5.1.6.** Falls  $G_a(s) = G_b(s)$  für alle  $|s| < R'$ , wobei  $0 < R' \leq R$ , dann gilt  $a_n = b_n$  für alle  $n$ . Weiterhin gilt:

$$a_n = \frac{1}{n!} G_a^{(n)}(0).$$

**Satz 5.1.7** (Satz von Abel). Falls  $a_i \geq 0$  und  $G_a(s)$  für  $|s| < 1$  konvergiert, dann gilt

$$\lim_{s \uparrow 1} G_a(s) = G_a(1) = \sum_{i=0}^{\infty} a_i,$$

egal, ob die Summe auf der rechten Seite konvergiert oder nicht.

Für die WeF

$$G_X(s) = \sum_{i=0}^{\infty} s^i P(X = i)$$

einer Zufallsvariable  $X$  gilt

$$G_X(0) = P(X = 0) \quad \text{und} \quad G_X(1) = 1,$$

d.h. der Konvergenzradius von  $G_X(s)$  ist mindestens gleich 1.

**Beispiel 5.1.8.**

1. Die konstante Zufallsvariable  $X$  mit  $P(X = c) = 1$  für ein  $c \in \mathbb{N}_0$  hat die WeF  $G_X(s) = s^c$ .

2. Für die Bernoulliverteilung gilt bekanntermaßen

$$P(X = 1) = p \quad \text{und} \quad P(X = 0) = 1 - p = q,$$

und somit ergibt sich für sie die WeF

$$G_X(s) = q + p \cdot s.$$

3. Für die geometrische Verteilung gilt bekanntermaßen

$$P(X = k) = p(1 - p)^{k-1} \quad \text{mit } k = 1, 2, \dots,$$

und somit ergibt sich die WeF zu

$$\begin{aligned} G_X(s) &= \sum_{k=1}^{\infty} p(1-p)^{k-1} s^k \\ &= p \cdot s \sum_{k=1}^{\infty} [s(1-p)]^{k-1} \\ &= p \cdot s \sum_{k'=0}^{\infty} [s(1-p)]^{k'} \\ &\stackrel{\text{Geom. Reihe}}{=} \frac{ps}{1 - s(1-p)}. \end{aligned}$$

4. Für eine Poissonverteilte Zufallsvariable  $X$  gilt gemäß Beispiel 5.1.2:

$$G_X(s) = \exp(\lambda(s - 1)).$$

**Satz 5.1.9.** *Besitzt die Zufallsvariable  $X$  die WeF  $G_X(s)$ , so gilt:*

$$E(X \cdot (X - 1) \cdot \dots \cdot (X - k + 1)) = G_X^{(k)}(1).$$

Man nennt  $E(X \cdot (X - 1) \cdot \dots \cdot (X - k + 1))$  das *k-te faktorielle Moment*. Insbesondere gilt somit

$$E(X) = G_X'(1) = G_X^{(1)}(1).$$

**Bemerkung:** Falls der Konvergenzradius von  $G_X(s)$  gleich 1 ist, so ist  $G_X^{(k)}(1)$  eine Abkürzung für

$$G_X^{(k)}(1) = \lim_{s \uparrow 1} G_X^{(k)}(s).$$

*Beweis zu Satz 5.1.9.* Für  $s < 1$  gilt:

$$\begin{aligned}
 G_a(s) &= \sum_{i=0}^{\infty} s^i \cdot P(X = i) \\
 G_a^{(1)}(s) &= \sum_{i=0}^{\infty} i \cdot s^{i-1} P(X = i) \\
 G_a^{(2)}(s) &= \sum_{i=0}^{\infty} i \cdot (i-1) \cdot s^{i-2} \cdot P(X = i) \\
 &\vdots \\
 G_a^{(k)}(s) &= \sum_{i=0}^{\infty} i \cdot (i-1) \cdot \dots \cdot (i-k+1) \cdot s^{i-k} \cdot P(X = i) \\
 &= E(X \cdot (X-1) \cdot \dots \cdot (X-k+1) s^{X-k})
 \end{aligned}$$

Für  $s \uparrow 1$  folgt mit Satz 5.1.7:

$$\begin{aligned}
 \lim_{s \uparrow 1} G_X^{(k)}(s) &= \sum_{i=0}^{\infty} i \cdot (i-1) \cdot \dots \cdot (i-k+1) \cdot P(X = i) \\
 &= E(X \cdot (X-1) \cdot \dots \cdot (X-k+1)).
 \end{aligned}$$

□

Häufig benötigt man Satz 5.1.9 zur Berechnung der Varianz:

**Korollar 5.1.10** (Berechnung der Varianz). *Es gilt:*

$$\begin{aligned}
 \text{Var}(X) &= E(X^2) - [E(X)]^2 \\
 &= \underbrace{E(X(X-1))}_{\substack{2. \text{ fakt. Moment} \\ 2. \text{ fakt. Moment}}} + E(X) - [E(X)]^2 \\
 &= G_X''(1) + G_X'(1) - [G_X'(1)]^2.
 \end{aligned}$$

**Beispiel 5.1.11** (Poissonverteilung). Für  $X \sim \text{Po}(\lambda)$  gilt wegen

$$\begin{aligned}
 G_X(s) &= \exp(\lambda(s-1)) \\
 G_X'(s) &= \lambda \exp(\lambda(s-1)) \\
 G_X''(s) &= \lambda^2 \exp(\lambda(s-1))
 \end{aligned}$$

offensichtlich

$$\begin{aligned}
 E(X) &= G_X'(1) = \lambda \\
 E(X(X-1)) &= G_X''(1) = \lambda^2
 \end{aligned}$$

und somit

$$\text{Var}(X) = G_X''(1) + G_X'(1) - [G_X'(1)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

**Beispiel 5.1.12** (Geometrische Verteilung). Für die WeF einer geometrisch verteilten Zufallsvariable  $X$  gilt

$$\begin{aligned}
 G_X(s) &\stackrel{\text{Bsp. 5.1.8}}{=} \frac{ps}{1-s(1-p)} \\
 G'_X(s) &= \frac{p(1-s(1-p)) + ps(1-p)}{(1-s(1-p))^2} \\
 &= \frac{p}{(1-s(1-p))^2} \\
 G''_X(s) &= \frac{-p \cdot 2 \cdot (1-s(1-p)) \cdot -(1-p)}{(1-s(1-p))^4} \\
 &= \frac{2 \cdot p(1-p)}{(1-s(1-p))^3}.
 \end{aligned}$$

Somit ergibt sich

$$\begin{aligned}
 E(X) = G'_X(1) &= \frac{p}{p^2} = \frac{1}{p}, \\
 E(X(X-1)) = G''_X(1) &= \frac{2 \cdot p(1-p)}{p^3} = \frac{2(1-p)}{p^2}
 \end{aligned}$$

und schließlich

$$\begin{aligned}
 \text{Var}(X) &= G''_X(1) + G'_X(1) - [G'_X(1)]^2 \\
 &= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} \\
 &= \frac{2-2p+p-1}{p^2} \\
 &= \frac{1-p}{p^2}.
 \end{aligned}$$

Dieses Resultat ist direkt (ohne WeF) nur recht aufwändig herzuleiten.

Nun beschäftigen wir uns wieder mit Summen von unabhängigen Zufallsvariablen:

**Satz 5.1.13.** Sind  $X$  und  $Y$  unabhängig, so gilt:

$$G_{X+Y}(s) = G_X(s) \cdot G_Y(s).$$

Dies gilt analog auch für  $n$  unabhängige Zufallsvariablen  $X_i$ ,  $i = 1, \dots, n$ :

$$G_{\sum_{i=1}^n X_i}(s) = \prod_{i=1}^n G_{X_i}(s).$$

*Beweis.* Es gibt hier zwei Beweismethoden:

1. Direkt über den Faltungssatz und (30).
2.  $X, Y$  unabhängig  $\implies s^X$  und  $s^Y$  unabhängig, insbesondere

$$E(s^X \cdot s^Y) = E(s^X) \cdot E(s^Y).$$

Damit folgt

$$\begin{aligned}
 G_{X+Y}(s) &= E(s^{X+Y}) \\
 &= E(s^X \cdot s^Y) \\
 &= E(s^X) \cdot E(s^Y) \\
 &= G_X(s) \cdot G_Y(s).
 \end{aligned}$$

□

**Beispiel 5.1.14** (Binomialverteilung). Seien  $X_i \sim B(p)$ ,  $i = 1, \dots, n$ , unabhängig, so ist  $S = X_1 + \dots + X_n \sim B(n, p)$ . Da

$$G_{X_i}(s) = q + p \cdot s$$

mit  $q = 1 - p$  (vgl. Beispiel 5.1.8), folgt

$$G_S(s) = (q + p \cdot s)^n.$$

Interessant ist der Fall, wenn der Stichprobenumfang  $N = n$  zufällig ist:

**Satz 5.1.15** (Vermengung von Zufallsvariablen). Seien  $X_1, X_2, \dots$  unabhängig und identisch verteilt (iid) mit WeF  $G_X(s)$  und  $N$  eine von  $X_1, X_2, \dots$  unabhängige diskrete Zufallsvariable mit dem Träger  $\mathcal{T} \subset \mathbb{N}_0$  und WeF  $G_N(s)$ . Dann gilt für  $S = X_1 + X_2 + \dots + X_N$ :

$$G_S(s) = G_N(G_X(s)).$$

Man spricht von der Vermengung der Zufallsvariablen  $X$  und  $N$  (engl. compounding).

Dieser Satz ist gewissermassen eine Verallgemeinerung von Satz 5.1.13. Im Spezialfall einer konstanten Zufallsvariable  $N$  ergibt sich nämlich  $P(N = n) = 1$ , mit Beispiel 5.1.8, Teil 1, folgt  $G_N(s) = s^n$  und somit  $G_S(s) = [G_X(s)]^n$ .

**Bemerkung:** Die Abkürzung *iid* für "unabhängig und identisch verteilt" ergibt sich aus dem Englischen und steht für "independent and identically distributed".

*Beweis zu Satz 5.1.15.*

$$\begin{aligned}
 G_S(s) &= E(s^S) \\
 &= E(E(s^S|N)) \\
 &= \sum_n E(s^S|N = n) \cdot P(N = n) \\
 &= \sum_n E(s^{X_1 + \dots + X_n}) \cdot P(N = n) \\
 &= \sum_n E(s^{X_1} \cdot s^{X_2} \cdot \dots \cdot s^{X_n}) \cdot P(N = n) \\
 &\stackrel{\text{Unabh.}}{=} \sum_n E(s^{X_1}) \cdot E(s^{X_2}) \cdot \dots \cdot E(s^{X_n}) \cdot P(N = n) \\
 &= \sum_n [G_X(s)]^n \cdot P(N = n) \\
 &= G_N(G_X(s)).
 \end{aligned}$$

□

**Beispiel 5.1.16** (Fortsetzung von Beispiel 3.7.7). Eine Henne legt  $N \sim \text{Po}(\lambda)$  Eier. Jedes Ei brütet sie (unabhängig von den anderen) mit einer Wahrscheinlichkeit  $p$  aus.  $K = X_1 + X_2 + \dots + X_N$  ist die Anzahl der Küken. Also:  $X_i \sim B(p)$  iid mit identischer WeF  $G_X(s)$ . Wir wissen:

$$\begin{aligned} G_N(s) &= \exp(\lambda(s-1)), \\ G_X(s) &= 1 - p + ps. \end{aligned}$$

Nach Satz 5.1.15 folgt:

$$\begin{aligned} G_K(s) &= G_N(G_X(s)) \\ &= \exp(\lambda((1-p+ps)-1)) \\ &= \exp(\lambda p(s-1)). \end{aligned}$$

Also  $K \sim \text{Po}(\lambda p)$ .

## 5.2 Anwendung: Verzweigungsprozesse

Verzweigungsprozesse beschreiben die zeitliche Entwicklung von Populationen (z.B. von Zellen), von Epidemien, etc. Sei  $Z_n$  die Anzahl von Mitgliedern in der  $n$ -ten Generation, z.B. die Anzahl von Infizierten bei Beobachtung einer Epidemie. Jedes Mitglied der  $n$ -ten Generation erzeugt eine Familie von Mitgliedern in der  $(n+1)$ -ten Generation. Die Größe dieser Familie, d.h. die Anzahl der Nachkommen, sei zufällig mit Träger  $\{0, 1, 2, \dots\}$ .

Wir nehmen im Folgenden an:

1. Die Familiengrößen sind unabhängige Zufallsvariablen.
2. Die Familiengrößen haben identische Verteilung mit Wahrscheinlichkeitsfunktion  $f(i)$ ,  $i = 0, 1, 2, \dots$ , und WeF  $G(s)$ .
3.  $Z_0 = 1$ .

Gesucht ist nun  $G_n(s) = E(s^{Z_n})$ , die WeF von  $Z_n$ . Zunächst gilt  $G_1(s) = G(s)$ , und wir können  $G_n(s)$  über folgenden Satz berechnen:

**Satz 5.2.1** (Iterierte von  $G(s)$ ). *Es gilt*

$$\begin{aligned} G_{m+n}(s) &= G_m(G_n(s)) \\ &= G_n(G_m(s)) \end{aligned}$$

und daher

$$G_n(s) = \underbrace{G(G(\dots(G(s))\dots))}_{n\text{-mal}},$$

die  $n$ -fach Iterierte von  $G(s)$ .

*Beweis.* Jedes Mitglied der  $(m+n)$ -ten Generation hat einen eindeutigen Vorfahren in der  $m$ -ten Generation:

$$Z_{m+n} = X_1 + X_2 + \dots + X_{Z_m},$$

wobei  $X_i$  die Anzahl der Mitglieder der  $(m+n)$ -ten Generation ist, die das  $i$ -te Mitglied der  $m$ -ten Generation zum Vorfahren haben. Aus den Annahmen 1. und 2. folgt, dass auch die Zufallsvariablen  $X_i$  iid mit WeF  $G_X(s)$  sind. Mit Satz 5.1.15 folgt:

$$G_{m+n}(s) = G_m(G_X(s)),$$

wobei  $G_X(s) = G_n(s)$ , und daher

$$\begin{aligned} G_n(s) &= G(G_{n-1}(s)) \\ &= G(G(G_{n-2}(s))) \\ &\vdots \\ &= G(G(\dots(G(s))\dots)). \end{aligned}$$

□

Mit diesem Satz lassen sich die Momente von  $Z_n$  leicht berechnen:

**Lemma 5.2.2.** Sei  $\mu = E(Z_1)$  und  $\sigma^2 = \text{Var}(Z_1)$ . Dann gilt:

$$\begin{aligned} E(Z_n) &= \mu^n \\ \text{Var}(Z_n) &= \begin{cases} n \cdot \sigma^2 & , \text{ falls } \mu = 1, \\ \frac{\sigma^2(\mu^n - 1)\mu^{n-1}}{\mu - 1} & , \text{ falls } \mu \neq 1. \end{cases} \end{aligned}$$

*Beweis.* Aus Satz 5.2.1 weiß man, dass

$$G_n(s) = G(G_{n-1}(s)),$$

und es folgt nach der Kettenregel:

$$G'_n(s) = G'(G_{n-1}(s)) \cdot G'_{n-1}(s).$$

Insbesondere ist für  $s = 1$  mit Satz 5.1.9:

$$\underbrace{G'_n(1)}_{=E(Z_n)} = \underbrace{G'(G_{n-1}(1))}_{=1} \cdot \underbrace{G'_{n-1}(1)}_{=E(Z_{n-1})}$$

$\underbrace{\hspace{10em}}_{=E(Z_1)}$

und somit

$$E(Z_n) = \mu \cdot E(Z_{n-1}) = \mu^n.$$

Das zweite Resultat für  $\text{Var}(Z_n)$  folgt über  $G''_n(1)$  und Korollar 5.1.10. □

Diesen Beweis haben wir bereits in Beispiel 3.7.14 für  $\mu = \sigma^2 = \lambda > 0$  geführt.

Zu beachten ist:

Für  $\mu < 1$  gilt:

$$\begin{aligned} E(Z_n) &\rightarrow 0, \\ \text{Var}(Z_n) &\rightarrow 0. \end{aligned}$$

Für  $\mu > 1$  gilt:

$$\begin{aligned} E(Z_n) &\rightarrow \infty, \\ \text{Var}(Z_n) &\rightarrow \infty. \end{aligned}$$

Im zweiten Fall *explodiert* der Prozess mit positiver Wahrscheinlichkeit, daher divergieren  $E(Z_n)$  und  $\text{Var}(Z_n)$ .

### 5.2.1 Verzweigungsprozesse mit Migration

In jeder Generation  $n$  werde nun eine zufällige Anzahl  $I_n$  von Mitgliedern hinzugefügt, die analog zu den bisherigen Mitgliedern (also denen, die tatsächlich Nachkommen sind) Nachkommen zeugen können.

Seien  $I_n$  iid Zufallsvariablen mit WeF  $H(s)$ , die vom Verzweigungsprozess bis zu dem Zeitpunkt unabhängig sind, an dem sie dazu kommen und dann den Verzweigungsprozess beeinflussen. Insgesamt hat also die Gesamtzahl  $Z_n$  in der  $n$ -ten Generation WeF  $G_n$ , wobei

$$G_{n+1}(s) = G_n(G(s)) \cdot H(s).$$

Hierbei ist  $G(s)$  die WeF der Anzahl der Nachkommen mit dem Erwartungswert  $\mu$ , und  $H(s)$  ist die WeF der Immigration mit dem Erwartungswert  $\lambda$ .

Durch Anwendung der Sätze 5.1.13 und 5.1.15 ergibt sich Folgendes: Sei  $\nu_n = E(Z_n)$ . Dann gilt

$$G'_{n+1}(1) = G'_n(G(1)) \cdot G'(1) \cdot H(1) + G_n(G(1)) \cdot H'(1),$$

d.h.

$$\nu_{n+1} = \nu_n \cdot \mu + \lambda.$$

Man kann zeigen, dass der Prozess  $\{Z_n\}$  für  $\mu < 1$  gegen eine *stationäre Verteilung* (Verteilung mit konstantem Erwartungswert und konstanter Varianz) konvergiert. Der Erwartungswert  $\nu$  erfüllt dann:

$$\nu = \nu \cdot \mu + \lambda \Rightarrow \nu = \frac{\lambda}{1 - \mu}.$$

## 5.3 Allgemeiner Erwartungswertbegriff

**Ziel** dieses Kapitel ist die Verallgemeinerung des Begriffs der WeF auf allgemeine Typen von Zufallsvariablen. Dazu nötig ist ein allgemeinerer Erwartungswertbegriff, der unabhängig vom Typ der Zufallsvariable (stetig, diskret, etc.) ist.

### 5.3.1 Notation

Bisher gilt folgende Notation:

- Sei  $X$  eine *diskrete* Zufallsvariable mit der Wahrscheinlichkeitsfunktion  $f(x)$ . Dann gilt:

$$E(X) = \sum_x x \cdot f(x). \quad (31)$$

- Sei  $X$  eine *stetige* Zufallsvariable mit der Dichtefunktion  $f(x)$ , dann gilt:

$$E(X) = \int x \cdot f(x) dx. \quad (32)$$

Gleichung (31) kann umgeschrieben werden zu

$$E(X) = \sum x dF(x),$$

wobei  $dF(x) = F(x) - \lim_{y \uparrow x} F(y) = f(x)$ . Analog dazu lässt sich (32) umschreiben zu

$$E(X) = \int x dF(x),$$

wobei  $dF(x) = \frac{dF(x)}{dx} dx = f(x) dx$ . Dies legt folgende allgemeine Notation nahe:

$$E(X) = \int x dF \quad \text{bzw.} \quad E(X) = \int x dF(x).$$

Sämtliche Resultate für Erwartungswerte können nun in einer vereinheitlichten Notation geschrieben werden, z.B.:

Für  $g : \mathbb{R} \rightarrow \mathbb{R}$  gilt

$$E(g(X)) = \int g(x) dF.$$

### 5.3.2 Abstrakte Integration

Der Erwartungswert einer Zufallsvariable  $X$  ist eine Funktion der Verteilungsfunktion  $F$ .  $F$  kann wiederum als Funktion der Abbildung  $X$  und dem zugrundeliegenden Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$  betrachtet werden. Also kann auch  $E(X)$  über  $(\Omega, \mathcal{F}, P)$  und  $X$  beschrieben werden. Dazu benötigt man einen Integrationsbegriff auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$ .

**Definition 5.3.1** (Einfache Zufallsvariable). Eine Zufallsvariable  $X$  heißt *einfach*, wenn sie nur endlich viele unterschiedliche Werte annimmt; sie kann also in der Form

$$X = \sum_{i=1}^n x_i I_{A_i}$$

für eine bestimmte Partition  $A_1, A_2, \dots, A_n$  von  $\Omega$  (vgl. Definition 1.4.3) und bestimmte reelle Zahlen  $x_1, x_2, \dots, x_n$  dargestellt werden. Man definiert nun das *Integral* von  $X$ ,  $E(X)$ , als

$$E(X) = \sum_{i=1}^n x_i \cdot P(A_i).$$

**Satz 5.3.2.** Jede nicht-negative Zufallsvariable  $X : \Omega \rightarrow [0, \infty)$  kann als Grenzwert einer aufsteigenden Folge  $\{X_n\}$  von einfachen Zufallsvariablen geschrieben werden:

$$X_n(\omega) \uparrow X(\omega)$$

für alle  $n \rightarrow \infty$  und alle  $\omega \in \Omega$ . Man definiert nun das Integral von  $X$  wie folgt:

$$E(X) = \lim_{n \rightarrow \infty} E(X_n).$$

**Satz 5.3.3.** Jede Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  kann als Differenz von zwei nicht-negativen Zufallsvariablen dargestellt werden:

$$X = X^+ - X^-,$$

wobei

$$X^+ = X^+(\omega) = \max\{X(\omega), 0\}$$

und

$$X^- = X^-(\omega) = -\min\{X(\omega), 0\}.$$

Nun definiert man das Integral von  $X$  wie folgt:

$$E(X) = E(X^+) - E(X^-),$$

wobei mindestens  $E(X^+)$  oder  $E(X^-)$  endlich sein muss.

**Korollar 5.3.4.** Nach Satz 5.3.3 ist  $E(X)$  nun zumindest für jede Zufallsvariable  $X$  mit  $E(|X|) = E(X^+ + X^-) < \infty$  definiert.

In der Sprache der Maßtheorie schreibt man

$$E(X) = \int_{\Omega} X(\omega) dP \quad \text{bzw.} \quad E(X) = \int_{\Omega} X(\omega) P(d\omega).$$

Der Erwartungswert-Operator  $E$  hat speziell für diskrete bzw. stetige Zufallsvariablen alle uns bekannten Eigenschaften. Weiterhin gilt:

**Satz 5.3.5** (Stetigkeit von  $E$ ). Sei  $\{X_n\}$  eine Folge von Zufallsvariablen mit  $X_n(\omega) \rightarrow X(\omega)$  für alle  $\omega \in \Omega$ .

- Monotone Konvergenz: Falls  $X_n(\omega) \geq 0$  und  $X_n(\omega) \leq X_{n+1}(\omega)$  für alle  $n$  und  $\omega$ , dann gilt

$$E(X_n) \rightarrow E(X)$$

für  $n \rightarrow \infty$ .

- Dominierte Konvergenz: Falls  $|X_n(\omega)| \leq Y(\omega)$  für alle  $n$  und  $\omega$  und  $E(|Y|) < \infty$ , dann gilt:

$$E(X_n) \rightarrow E(X)$$

für  $n \rightarrow \infty$ . Man sagt hierzu: Die Zufallsvariable  $X$  wird von der Zufallsvariable  $Y$  dominiert.

- *Speziell: Falls  $|X_n(\omega)| \leq c$  mit konstantem  $c \in \mathbb{R}^+$  für alle  $n$  und  $\omega$ , so gilt:*

$$E(X_n) \rightarrow E(X)$$

für  $n \rightarrow \infty$ .

*Desweiteren gilt: Ereignisse mit Wahrscheinlichkeit null (Nullereignisse) haben keinen Beitrag zum Erwartungswert. Es genügt vorauszusetzen, dass  $X_n(\omega) \rightarrow X(\omega)$  fast sicher (f.s.) gilt, d.h. für alle  $\omega \in \Omega$  außer möglichen Nullereignissen:*

$$X_n(\omega) \xrightarrow{f.s.} X(\omega).$$

Eine Folgerung der monotonen Konvergenz ist die (bereits bekannte) Additivität von Erwartungswerten:

**Korollar 5.3.6.** *Sei  $Z_1, Z_2, \dots$  eine Folge von Zufallsvariablen mit  $Z_i \geq 0$  und  $E(Z_i) < \infty$  und sei  $X = \sum_{i=1}^{\infty} Z_i$ . Dann gilt:*

$$E(X) = \sum_{i=1}^{\infty} E(Z_i).$$

**Lemma 5.3.7** (Lemma von Fatou; FATOU, 1878-1929). *Sei  $\{X_n\}$  eine Folge von Zufallsvariablen mit  $X_n \geq Y$  fast sicher für alle  $n \in \mathbb{N}$  und ein  $Y$  mit  $E(|Y|) < \infty$ . Dann gilt:*

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

**Definition 5.3.8** (Lebesgue-Stieltjes-Integral; LEBESGUE, 1875-1941, STIELTJES, 1856-94). Sei  $X$  eine Zufallsvariable mit Verteilungsfunktion  $F$ .  $F$  definiert ein Wahrscheinlichkeitsmaß  $\mu_F$  auf den Borel-Mengen von  $\mathbb{R}$  wie folgt:

- Setze  $\mu_F((a, b]) = F(b) - F(a)$ .
- $\mu_F$  kann auf die Borelsche  $\sigma$ -Algebra  $\mathcal{B}$  erweitert werden, die kleinste  $\sigma$ -Algebra, die alle halboffenen Intervalle  $(a, b]$  enthält. Also ist  $(\mathbb{R}, \mathcal{B}, \mu_F)$  ein Wahrscheinlichkeitsraum, der zu einem vollständigen Wahrscheinlichkeitsraum  $(\mathbb{R}, \mathcal{L}_F, \mu_F)$  vervollständigt werden kann (vgl. Kapitel 1.6); hier ist also  $\mathcal{L}_F$  die kleinste  $\sigma$ -Algebra, die  $\mathcal{B}$  und alle Teilmengen von  $\mu_F$ -Nullmengen enthält. Vergleiche hierzu die Diskussion in und vor Definition 4.1.2.

Sei nun also  $g : \mathbb{R} \rightarrow \mathbb{R}$   $\mathcal{L}_F$ -messbar. Dann nennt man das abstrakte Integral

$$\begin{aligned} E(g(X)) &= \int g(x) dF \\ &= \int g(x) dF(x) \end{aligned}$$

das *Lebesgue-Stieltjes-Integral* von  $g$  bezüglich  $\mu_F$ .

Dieses stellt man sich am besten wie das abstrakte Integral

$$E(X) = \int_{\Omega} X(\omega) dP,$$

vor.

## 5.4 Momentenerzeugende Funktion und Kumulanten

**Definition 5.4.1** (Momentenerzeugende Funktion). Die *momentenerzeugende Funktion* (*MeF*) einer Zufallsvariablen  $X$  ist die Funktion  $M : \mathbb{R} \rightarrow \mathbb{R}_0^+$  mit

$$M(t) = E(e^{tX}) = E(\exp(tX)).$$

Offensichtlich gilt  $M(0) = 1$ .

**Satz 5.4.2.** Falls  $M(t)$  auf einem offenen Intervall, das die Null enthält, endlich ist, so gilt:

1.  $m_k = E(X^k) = M^{(k)}(0)$ , insbesondere  $E(X) = M'(0)$ .
2.  $M(t)$  kann durch Taylorentwicklung dargestellt werden als

$$M(t) = \sum_{k=0}^{\infty} \frac{m_k}{k!} t^k,$$

d.h.  $M(t)$  ist die exponentialerzeugende Funktion der Folge der Momente von  $X$ .

3. Sind  $X$  und  $Y$  unabhängig, so gilt:

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t).$$

4. Falls  $Y = a \cdot X + b$  so gilt

$$M_Y(t) = \exp(bt)M_X(at).$$

5. Falls  $X$  eine wahrscheinlichkeitserzeugende Funktion  $G_X(s)$  besitzt, so gilt  $M_X(t) = G_X(\exp(t))$ .

**Beispiel 5.4.3** (Poisson-Verteilung). Da für eine Poisson-verteilte Zufallsvariable  $X$  bekanntermaßen  $G_X(s) = \exp(\lambda(s - 1))$  ist, folgt:

$$\begin{aligned} M_X(s) &= G_X(\exp(s)) \\ &= \exp(\lambda(\exp(s) - 1)), \\ M_X'(s) &= \exp(\lambda(\exp(s) - 1)) \cdot \lambda \exp(s) \\ &= M_X(s) \cdot \lambda \exp(s). \end{aligned}$$

Für den Erwartungswert gilt:

$$E(X) = M_X'(0) = \lambda.$$

Für die Berechnung der Varianz mit MeF benötigt man die zweite Ableitung:

$$\begin{aligned} M_X''(s) &= M_X'(s) \cdot \lambda \exp(s) + M_X(s) \cdot \lambda \exp(s) \\ &= \lambda \exp(s) \cdot (M_X(s) + M_X'(s)). \end{aligned}$$

Somit ergibt sich mit

$$E(X^2) = M_X''(0) = \lambda(1 + \lambda) = \lambda^2 + \lambda$$

die Varianz zu

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \lambda.$$

**Definition 5.4.4** (Kumulantenerzeugende Funktion). Die *kumulantenerzeugende Funktion* (*KeF*) einer Zufallsvariablen  $X$  ist die Funktion

$$K(t) = \log M(t).$$

**Definition 5.4.5** (Kumulante). Die *r-te Kumulante* einer Zufallsvariablen  $X$  ist

$$\kappa_r = K^{(r)}(0),$$

so dass sich  $K(t)$  in der Form

$$K(t) = \sum_{r=1}^{\infty} \frac{\kappa_r}{r!} t^r$$

darstellen lässt.

Man beachte, dass die 0-te Kumulante verschwindet, da  $\kappa_0 = K(0) = \log M(0) = \log 1 = 0$ . Für die nächsten zwei Kumulanten einer Zufallsvariable  $X$  gilt:

$$\begin{aligned} \kappa_1 = K'(0) &= \frac{M'(0)}{M(0)} = m_1 = E(X), \\ \kappa_2 = K''(0) &= \frac{M''(0)M(0) - [M'(0)]^2}{[M(0)]^2} \\ &= M''(0) - [M'(0)]^2 = m_2 - m_1^2 = \text{Var}(X). \end{aligned}$$

Die ersten beiden Kumulanten sind also gleich dem Erwartungswert und der Varianz von  $X$ .

**Beispiel 5.4.6** (Momentenerzeugende Funktion und Kumulanten der Normalverteilung). Sei  $X \sim N(0, 1)$ . Dann ist

$$\begin{aligned} M_X(t) &= E(\exp(t \cdot X)) \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(tx - \frac{1}{2}x^2\right) dx \\ &= \exp\left(\frac{1}{2}t^2\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-t)^2\right) dx \\ &= \exp\left(\frac{1}{2}t^2\right). \end{aligned}$$

Mit Satz 5.4.2 folgt für  $Y = \sigma \cdot X + \mu$

$$\begin{aligned} M_Y(t) &= \exp(\mu t) \exp\left(\frac{1}{2}(\sigma t)^2\right) \\ &= \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right). \end{aligned}$$

Folglich ist

$$K_Y(t) = \mu t + \frac{1}{2}\sigma^2 t^2,$$

$\kappa_1 = \mu$ ,  $\kappa_2 = \sigma^2$ , und alle höheren Kumulanten sind null.

**Beispiel 5.4.7** (Kumulanten der Poissonverteilung). Da  $M_X(t) = \exp(\lambda(\exp(t) - 1))$  für  $X \sim \text{Po}(\lambda)$ , folgt  $K_X(t) = \lambda(\exp(t) - 1)$ ,  $K_X^{(r)} = \lambda \exp(t)$  für alle  $r = 1, 2, \dots$  und somit  $\kappa_r = \lambda$  für alle  $r = 1, 2, \dots$ . Bei der Poissonverteilung sind also alle Kumulanten gleich  $\lambda$ .

**Satz 5.4.8.** *Sind  $X$  und  $Y$  unabhängig, so gilt*

$$K_{X+Y}(t) = K_X(t) + K_Y(t).$$

*Es folgt, dass die  $r$ -te Kumulante einer Summe  $X_1 + \dots + X_n$  von  $n$  unabhängigen Zufallsvariablen gleich der Summe der jeweiligen  $r$ -ten Kumulanten der  $X_i$  ist.*

Die dritte und vierte Kumulante  $\kappa_3$  und  $\kappa_4$  nennt man *Schiefe* und *Kurtosis* (manchmal auch *Kurtosis Exzess*). Bei der Normalverteilung sind beide gleich null. Häufig verwendet man auch die *standardisierte Schiefe*  $\kappa_3/\kappa_2^{3/2}$  und *standardisierte Kurtosis*  $\kappa_4/\kappa_2^2$ .

## 5.5 Charakteristische Funktionen

Ein Problem von momentenerzeugenden Funktionen ist, dass  $M(t)$  nicht immer endlich ist. Immer lässt sich hingegen mit *charakteristischen Funktionen* arbeiten.

**Definition 5.5.1** (Charakteristische Funktion). Die *charakteristische Funktion (CF)* einer Zufallsvariablen  $X$  ist die Funktion  $\phi: \mathbb{R} \rightarrow \mathbb{C}$ , die durch

$$\phi(t) = \mathbb{E}(e^{itX}) = \mathbb{E}(\exp(itX))$$

mit  $i = \sqrt{-1}$  definiert ist.

Die Anwendung der komplexen Analysis führt zu:

$$\begin{aligned} \phi(t) = \mathbb{E}(\exp(itX)) &= \mathbb{E}(\cos(tX) + i \sin(tX)) \\ &= \mathbb{E}(\cos(tX)) + i \mathbb{E}(\sin(tX)). \end{aligned}$$

Man erkennt die Analogie zur Fourier-Transformation. Eine Wiederholung der entsprechenden Kapitel der Analysis erleichtert das Verständnis im Weiteren.

**Satz 5.5.2.** *Die CF  $\phi$  einer Zufallsvariablen  $X$  erfüllt*

- (a)  $\phi(0) = 1$  und  $|\phi(t)| \leq 1$  für alle  $t \in \mathbb{R}$ .
- (b)  $\phi(t)$  ist gleichmäßig stetig auf  $\mathbb{R}$ .
- (c)  $\phi(t)$  ist nicht-negativ definit, d.h.

$$\sum_{j=1}^n \sum_{k=1}^n \phi(t_j - t_k) z_j \bar{z}_k \geq 0$$

für  $t_1, t_2, \dots, t_n \in \mathbb{R}$  und  $z_1, z_2, \dots, z_n \in \mathbb{C}$ .  $\bar{z}_k$  ist die zu  $z_k$  konjugiert komplexe Zahl.

Die konjugiert komplexe Zahl zu  $a = \alpha + i\beta$  ist

$$\bar{a} = \alpha - i\beta.$$

Der Betrag  $|a|$  einer komplexen Zahl  $a = \alpha + i\beta$  ist definiert als

$$|a| = \sqrt{a \cdot \bar{a}} = \sqrt{\alpha^2 + \beta^2}.$$

*Beweis zu Satz 5.5.2.* Zu (a): Zunächst ist  $\phi(0) = E(\cos(0)) + iE(\sin(0)) = 1 + i \cdot 0 = 1$ . Weiterhin gilt:

$$\begin{aligned} |\phi(t)| &= |E(\exp(itX))| \\ &\stackrel{\text{Satz 4.5.1}}{\leq} E(|\exp(itX)|) \\ &= E(|\cos(tX) + i \cdot \sin(tX)|) \\ &= E\left(\sqrt{\cos^2(tX) + \sin^2(tX)}\right) \\ &= E(1) \\ &= 1. \end{aligned}$$

Für den Beweis von (b) und (c) siehe Buch von Grimmett und Stirzaker oder andere Literatur. □

**Bemerkung:** Es gilt sogar der Umkehrschluss: Funktionen, die (a)-(c) erfüllen, sind charakteristische Funktionen (*Satz von Bochner*).

Eine der wichtigsten Eigenschaften von CF ist, dass man aus ihr wieder die Verteilung von  $X$  berechnen kann. Zunächst betrachten wir nur die Momente der Verteilung.

**Satz 5.5.3.** *Ist  $X$  symmetrisch um Null, d.h.  $f(x) = f(-x)$  für alle  $x \in \mathbb{R}$ , so ist  $\phi_X(t)$  reellwertig.*

*Beweis.* Da  $\phi(t) = E(\cos(tX)) + iE(\sin(tX))$ , ist  $E(\sin(t)) = 0$  zu zeigen:

$$\begin{aligned} E(\sin(t)) &= \int_{-\infty}^{\infty} \sin(tx)f(x)dx \\ &= \int_{-\infty}^0 \sin(tx)f(x)dx + \int_0^{\infty} \sin(tx)f(x)dx \\ &\stackrel{\sin(x) = -\sin(-x)}{=} \int_{-\infty}^0 -\sin(-tx)f(-x)dx + \int_0^{\infty} \sin(tx)f(x)dx \\ &= -\int_0^{\infty} \sin(tx)f(x)dx + \int_0^{\infty} \sin(tx)f(x)dx \\ &= 0. \end{aligned}$$

□

**Satz 5.5.4.**

(a) Falls  $\phi^{(k)}(0)$  existiert, so gilt:

$$\begin{aligned} \mathbb{E}(|X^k|) &< \infty \text{ für } k \text{ gerade,} \\ \mathbb{E}(|X^{k-1}|) &< \infty \text{ für } k \text{ ungerade.} \end{aligned}$$

(b) Für  $\mathbb{E}(|X^k|) < \infty$  gilt (Satz von TAYLOR für Funktionen von komplexen Variablen, 1685-1731):

$$\phi(t) = \sum_{j=0}^k \frac{\mathbb{E}(X^j)}{j!} (it)^j + o(t^k).$$

Insbesondere ist  $\phi^{(k)}(0) = i^k \mathbb{E}(X^k)$ .

Die Notation  $o(t^k)$  ist hier wie folgt zu verstehen: Für das Restglied  $f(t)$  gilt  $\frac{f(t)}{t^k} \rightarrow 0$  für  $t \rightarrow 0$ .

**Satz 5.5.5.** Sind  $X$  und  $Y$  unabhängige Zufallsvariablen, so gilt:

$$\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t).$$

*Beweis.* Es ist

$$\begin{aligned} &\phi_{X+Y}(t) \\ &= \mathbb{E}(\exp(it(X+Y))) \\ &= \mathbb{E}(\exp(itX) \exp(itY)) \\ &= \mathbb{E}([\cos(tX) + i \cdot \sin(tX)][\cos(tY) + i \cdot \sin(tY)]) \\ &= \mathbb{E}(\cos(tX) \cos(tY) - \sin(tX) \sin(tY) + i[\cos(tX) \sin(tY) + \sin(tX) \cos(tY)]) \\ &\stackrel{\text{unabh.}}{=} \mathbb{E}(\cos(tX)) \cdot \mathbb{E}(\cos(tY)) - \mathbb{E}(\sin(tX)) \cdot \mathbb{E}(\sin(tY)) \\ &\quad + i[\mathbb{E}(\cos(tX)) \cdot \mathbb{E}(\sin(tY)) + \mathbb{E}(\sin(tX)) \cdot \mathbb{E}(\cos(tY))] \\ &= [\mathbb{E}(\cos(tX)) + i \cdot \mathbb{E}(\sin(tX))] \cdot [\mathbb{E}(\cos(tY)) + i \cdot \mathbb{E}(\sin(tY))] \\ &= \mathbb{E}(\exp(itX)) \cdot \mathbb{E}(\exp(itY)) \\ &= \phi_X(t) \cdot \phi_Y(t). \end{aligned}$$

□

**Satz 5.5.6.** Sei  $Y = aX + b$  mit  $a, b \in \mathbb{R}$ . Dann gilt:

$$\phi_Y(t) = \exp(itb) \phi_X(at).$$

Dieser Satz beinhaltet folgenden Spezialfall:  $Y = aX \Rightarrow \phi_Y(t) = \phi_{aX}(t) = \phi_X(at)$ .

*Beweis zu Satz 5.5.6.* Es gilt:

$$\begin{aligned} \phi_X(t) &= \mathbb{E}(\exp(it(aX + b))) \\ &= \mathbb{E}(\exp(itb) \cdot \exp(itaX)) \\ &= \exp(itb) \mathbb{E}(\exp(itaX)) \\ &= \exp(itb) \phi_X(at). \end{aligned}$$

□

**Satz 5.5.7.** *Unter bestimmten Regularitätsvoraussetzungen (Details im Buch von Grimmett und Stirzaker) gilt:*

$$\phi(t) = M(it).$$

*Dies macht die Berechnung von CFen bei Kenntnis der MeF sehr einfach.*

## 5.6 Beispiele für charakteristische Funktionen

Die Herleitung der folgenden Ergebnisse findet man im Buch von Grimmett und Stirzaker.

1. Konstante Variable. Die konstante Variable  $X = c$  hat CF

$$\phi(t) = \exp(itc).$$

2. Bernoulliverteilung:

$$\phi(t) = q + p \cdot \exp(it) \quad \text{mit } q = 1 - p.$$

3. Binomialverteilung:

$$\phi(t) = [q + p \cdot \exp(it)]^n \quad \text{mit } q = 1 - p.$$

4. Exponentialverteilung:

$$\phi(t) = \frac{\lambda}{\lambda - it}.$$

5. Standard-Cauchy-Verteilung:

$$\phi(t) = \exp(-|t|).$$

6. Normalverteilung:

$$\phi(t) = \exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right).$$

Hieraus folgt für die Standardnormalverteilung:

$$\phi(t) = \exp\left(-\frac{1}{2}t^2\right).$$

7. Gammaverteilung:  $X \sim G(a, b)$ , dann

$$\phi(t) = \left(\frac{b}{b - it}\right)^a.$$

## 5.7 Umkehrformel und Stetigkeitssatz

**Satz 5.7.1** (Inverse FOURIER-Transformation, 1768-1830). *Sei  $X$  eine stetige Zufallsvariable mit Dichtefunktion  $f(x)$  und CF  $\phi(t)$ . Dann gilt für alle  $x \in \mathbb{R}$ , an denen  $f(x)$  differenzierbar ist:*

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(itx) \cdot \phi(t) dt.$$

Eine hinreichende (aber nicht notwendige) Bedingung dafür, dass  $\phi(t)$  die CF einer stetigen Zufallsvariable ist, ist  $\int_{-\infty}^{\infty} |\phi(t)| dt < \infty$ .

Der allgemeine Fall (d.h.  $X$  nicht stetig) ist deutlich komplizierter, aber auch hier lässt sich die Verteilungsfunktion einer beliebigen Zufallsvariablen  $X$  aus deren CF berechnen:

**Satz 5.7.2.** *Sei  $X$  eine Zufallsvariable mit Verteilungsfunktion  $F(x)$  und CF  $\phi(t) = E(\exp(itX)) = \int_{-\infty}^{\infty} \exp(itx) dF(x)$ . Sei ferner  $\bar{F}(x) = \frac{1}{2}(F(x) + \lim_{y \uparrow x} F(y))$ . Dann gilt:*

$$\bar{F}(b) - \bar{F}(a) = \frac{1}{2\pi} \lim_{N \rightarrow \infty} \int_{-N}^N \frac{\exp(-iat) - \exp(-ibt)}{it} \phi(t) dt.$$

Für den Fall der stetigen Verteilungsfunktion ist  $\bar{F}(x) = F(x)$ .

**Korollar 5.7.3** (Eindeutigkeitssatz von charakteristischen Funktionen). *Zwei Zufallsvariablen  $X$  und  $Y$  haben genau dann die gleiche CF, wenn ihre Verteilungsfunktionen identisch sind. Das heißt, die Verteilung einer Zufallsvariable ist durch ihre CF vollständig determiniert.*

Zur Formulierung der Stetigkeitssatzes (Satz 5.7.5) benötigen wir:

**Definition 5.7.4.** Eine Folge von Verteilungsfunktionen  $F_1, F_2, \dots$  konvergiert zu einer Verteilungsfunktion  $F$  (Schreibweise:  $F_n \rightarrow F$ ), falls  $\lim_{n \rightarrow \infty} F_n(x) = F(x)$  an allen Punkten  $x \in \mathbb{R}$ , an denen  $F(x)$  stetig ist.

**Bemerkung:** Wieso fordert man  $F_n \rightarrow F$  nur an allen Stetigkeitsstellen von  $F(x)$ ? Betrachte

$$F_n(x) = \begin{cases} 0 & \text{für } x < \frac{1}{n} \\ 1 & \text{für } x \geq \frac{1}{n} \end{cases} \quad \text{und} \quad G_n(x) = \begin{cases} 0 & \text{für } x < -\frac{1}{n} \\ 1 & \text{für } x \geq -\frac{1}{n} \end{cases}.$$

Sei

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 & \text{für } x \geq 0 \end{cases}$$

die Verteilungsfunktion der konstanten Zufallsvariable  $X = 0$ . Dann gilt

$$G_n(x) \longrightarrow F(x) \quad \text{für alle } x \in \mathbb{R},$$

aber

$$F_n(x) \longrightarrow F(x) \quad \text{für alle } x \in \mathbb{R} \setminus \{0\},$$

da  $F_n(0) \rightarrow 0$ . Also ist  $\lim_{n \rightarrow \infty} F_n(x)$  nicht rechtsstetig und somit keine Verteilungsfunktion. Trotzdem ist es naheliegend, auch  $F(x)$  als Grenzwert von  $F_n(x)$ ,  $n \rightarrow \infty$ , zuzulassen.

**Satz 5.7.5** (Stetigkeitssatz). Sei  $F_1, F_2, \dots$  eine Folge von Verteilungsfunktionen mit zugehörigen CF  $\phi_1, \phi_2, \dots$ .

- (a) Falls  $\lim_{n \rightarrow \infty} F_n(x) = F(x)$  mit zugehöriger CF  $\phi$ , dann gilt auch  $\lim_{n \rightarrow \infty} \phi_n(t) = \phi(t)$  für alle  $t \in \mathbb{R}$ .
- (b) Falls  $\phi(t) = \lim_{n \rightarrow \infty} \phi_n(t)$  existiert und im Punkt  $t = 0$  stetig ist, dann ist  $\phi$  die CF einer Verteilungsfunktion  $F$ , und es gilt  $F_n \rightarrow F$ .

## 5.8 Zwei Grenzwertsätze

Dieser Abschnitt beschäftigt sich mit den beiden wohl bekanntesten Sätzen der Statistik:

- Gesetz der großen Zahlen (GDGZ), engl. Law of Large Numbers (LLN),
- Zentraler Grenzwertsatz (ZGWS), engl. Central Limit Theorem (CLT).

Beide betreffen die Konvergenz einer Folge von Zufallsvariablen, wir benötigen also

**Definition 5.8.1** (Konvergenz in Verteilung). Sei  $X_1, X_2, \dots$  eine Folge von Zufallsvariablen mit zugehörigen Verteilungsfunktionen  $F_1, F_2, \dots$ . Man sagt,  $X_n$  konvergiert in Verteilung gegen  $X$  (Schreibweise:  $X_n \xrightarrow{D} X$ ), falls  $F_n \rightarrow F$  für  $n \rightarrow \infty$ , wobei  $F$  die Verteilungsfunktion von  $X$  ist.

**Bemerkung:** Die Schreibweise der Konvergenz in Verteilung,  $X_n \xrightarrow{D} X$ , ergibt sich aus der englischen Bezeichnung *convergence in distribution*.

**Satz 5.8.2** (Gesetz der großen Zahlen, GDGZ). Sei  $X_1, X_2, \dots$  eine Folge von unabhängigen und identisch verteilten (iid) Zufallsvariablen mit endlichem Erwartungswert  $\mu$ . Definiere die Folge der Partialsummen  $S_1, S_2, \dots$  durch

$$S_n = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i.$$

Dann gilt:

$$\frac{1}{n} S_n \xrightarrow{D} \mu.$$

**Bemerkungen:**

1. Bei  $T_n = \frac{1}{n} S_n$  spricht man auch von der *mittleren Partialsumme*.
2. Bei  $\mu$  handelt es sich in diesem Fall um die deterministische Zufallsvariable mit Punktmasse im Punkt  $\mu$  und der CF  $\phi(t) = \exp(it\mu)$ .
3. Das GDGZ funktioniert nicht immer! Ein prominentes Beispiel ist die Cauchy-Verteilung, also die  $t$ -Verteilung mit einem Freiheitsgrad. Hier ist auch die Voraussetzung des existierenden Erwartungswertes nicht erfüllt.
4. Das GDGZ existiert in zahlreichen Modifikationen und Erweiterungen. Das *starke* Gesetz der großen Zahlen werden wir im nächsten Kapitel kennenlernen.

*Beweis zu Satz 5.8.2.* Zu zeigen ist, dass

$$F_{\frac{1}{n}S_n}(x) = P\left(\frac{1}{n}S_n \leq x\right) \longrightarrow \begin{cases} 0, & \text{falls } x < \mu \\ 1, & \text{falls } x > \mu. \end{cases}$$

Sei  $\phi_X(t)$  die CF der  $X_i$ 's und  $\phi_n(t)$  die CF von  $T_n = \frac{1}{n}S_n$ . Wegen Satz 5.5.5 (sind  $X, Y$  unabhängig, dann  $\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t)$ ) und Satz 5.5.6 ( $\phi_{aX}(t) = \phi_X(at)$ ) gilt:  $S_n = X_1 + X_2 + \dots + X_n$  hat CF

$$\phi_{S_n}(t) \stackrel{\text{Satz 5.5.5}}{=} [\phi_X(t)]^n,$$

und  $\frac{1}{n}S_n$  hat CF

$$\phi_n(t) \stackrel{\text{Satz 5.5.6}}{=} \left[ \phi_X\left(\frac{t}{n}\right) \right]^n. \quad (33)$$

Nun lässt sich Satz 5.5.4 (Taylor-Reihe) auf  $\phi_X(t)$  mit  $k = 1$  anwenden:

$$\begin{aligned} \phi_X(t) &= \sum_{j=0}^1 \frac{E(X^j)}{j!} (it)^j + o(t) \\ &= 1 + i\mu t + o(t). \end{aligned}$$

In (33) eingesetzt ergibt dies:

$$\phi_n(t) = \left[ 1 + i\mu \frac{t}{n} + o\left(\frac{t}{n}\right) \right]^n \longrightarrow \exp(it\mu) \quad \text{für } n \rightarrow \infty,$$

da bekanntermaßen  $\lim_{n \rightarrow \infty} (1 + \frac{x}{n})^n = \exp(x)$ . Die Funktion  $\exp(it\mu)$  ist die CF der Einpunktverteilung im Punkt  $\mu$ . Mit dem Eindeutigkeitssatz (Korollar 5.7.3) und dem Stetigkeitssatz (Satz 5.7.5) folgt die Behauptung.  $\square$

**Bemerkung:** Sei nun zusätzlich  $\sigma^2 = \text{Var}(X_i) < \infty$  und  $\sigma^2 > 0$ . Aus Satz 5.8.2 folgt:

$$\frac{1}{n}(S_n - n\mu) \xrightarrow{D} (0, \frac{\sigma^2}{n}) \hat{=} (\text{Erwartungswert, Varianz}).$$

Ferner gilt offensichtlich  $(S_n - n\mu) \sim (0, n\sigma^2)$ , woraus sofort folgt, dass  $\frac{1}{\sqrt{n}}(S_n - n\mu)$  Erwartungswert 0 und Varianz  $\sigma^2$  besitzt. Der Zentrale Grenzwertsatz macht eine Aussage über die asymptotische Verteilung von  $\frac{1}{\sqrt{n}}(S_n - n\mu)$ :

$$\frac{1}{\sqrt{n}}(S_n - n\mu) \xrightarrow{D} N(0, \sigma^2)$$

bzw.

$$\frac{1}{\sqrt{n\sigma^2}}(S_n - n\mu) \xrightarrow{D} N(0, 1).$$

**Satz 5.8.3** (Zentraler Grenzwertsatz, ZGWS). *Seien  $X_1, X_2, \dots$  iid Zufallsvariablen mit  $\mu = E(X_i) < \infty$  und  $0 < \text{Var}(X_i) = \sigma^2 < \infty$ . Dann gilt für  $S_n = X_1 + X_2 + \dots + X_n$ :*

$$U_n = \frac{1}{\sqrt{n\sigma^2}}(S_n - n\mu) \xrightarrow{D} N(0, 1)$$

für  $n \rightarrow \infty$ .

*Beweis.* Sei  $Y_i = \frac{X_i - \mu}{\sigma}$  und  $\phi_Y$  die CF von den  $Y_i$ 's. Dann gilt:

$$\mathbf{E}(Y) = \mathbf{E}(Y_i) = 0 \quad (34)$$

$$\mathbf{Var}(Y) = \mathbf{Var}(Y_i) = 1, \quad (35)$$

da es sich bei  $Y_i$  um die Standardisierung von  $X_i$  handelt. Ferner ist

$$U_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i.$$

Mit Satz 5.5.4,  $k = 2$ , folgt:

$$\begin{aligned} \phi_Y(t) &= \sum_{j=0}^2 \frac{\mathbf{E}(Y^j)}{j!} (it)^j + o(t^2) \\ &= 1 + \mathbf{E}(Y)it + \frac{\mathbf{E}(Y^2)}{2} i^2 t^2 + o(t^2) \\ &\stackrel{(34)(35)}{=} 1 - \frac{1}{2} t^2 + o(t^2). \end{aligned}$$

Dann folgt für die CF von  $\sum_{i=1}^n Y_i$ :

$$\phi_{\sum_{i=1}^n Y_i}(t) = \left[ 1 - \frac{1}{2} t^2 + o(t^2) \right]^n$$

und für die CF  $\psi_n(t)$  von  $U_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$

$$\psi_n(t) = \left[ 1 + \frac{1}{2} \cdot \frac{t^2}{n} + o\left(\frac{t^2}{n}\right) \right]^n.$$

Da nun aber  $o\left(\frac{t^2}{n}\right) \rightarrow 0$  für  $n \rightarrow \infty$ , gilt

$$\psi_n(t) \longrightarrow \exp\left(-\frac{1}{2} t^2\right)$$

für  $n \rightarrow \infty$ . Da  $\exp\left(-\frac{1}{2} t^2\right)$  die CF der Standardnormalverteilung ist, folgt mit dem Stetigkeitssatz (Satz 5.7.5) und dem Eindeutigkeitsatz (Korollar 5.7.3) die Behauptung.  $\square$

**Bemerkung:** Das Kürzel  $\xrightarrow{D}$  wird häufig ersetzt durch  $\overset{a}{\sim}$ , was für "ist asymptotisch verteilt wie" steht. Somit sagt Satz 5.8.3:

$$U_n = \frac{1}{\sqrt{n\sigma^2}} (S_n - n\mu) \overset{a}{\sim} N(0, 1).$$

Die folgenden alternativen Schreibweisen für den ZGWS lassen sich daraus leicht ableiten:

$$\frac{\sqrt{n}}{\sigma} (T_n - \mu) \overset{a}{\sim} N(0, 1),$$

und

$$\frac{1}{\sqrt{n}}(S_n - n\mu) = \sqrt{n}(T_n - \mu) \stackrel{a}{\sim} N(0, \sigma^2).$$

Etwas vereinfachend findet man auch häufig die Schreibweisen

$$S_n \stackrel{a}{\sim} N(n\mu, n\sigma^2)$$

und

$$T_n \stackrel{a}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right),$$

bei denen die Parameter der asymptotischen Normalverteilung vom Stichprobenumfang  $n$  abhängen.

**Beispiel 5.8.4.** Sei  $S_n = \sum_{i=1}^n X_i$  und  $X_i \sim B(p)$ . Dann gilt für große  $n$ :

$$\begin{aligned} \frac{1}{\sqrt{np(1-p)}}(S_n - np) &\stackrel{a}{\sim} N(0, 1) \\ \text{bzw. } S_n &\stackrel{a}{\sim} N(np, np(1-p)) \\ \text{bzw. } T_n &\stackrel{a}{\sim} N\left(p, \frac{p(1-p)}{n}\right). \end{aligned}$$

Dies ist die älteste Version des ZGWS. Mit  $p = \frac{1}{2}$  wurde sie um 1733 von MOIVRE (1667-1754) und mit  $p \in (0, 1)$  später von DE LAPLACE (1749-1827) gezeigt.

Der ZGWS existiert in zahlreichen Modifikationen und Erweiterungen. Unter Anderem kann eine multivariate Version formuliert werden:

**Satz 5.8.5.** Sei  $(\mathbf{X}_i)_{i \in \mathbb{N}}$  eine Folge von iid  $p$ -dimensionalen Zufallsvektoren mit endlichem Erwartungswertvektor  $\boldsymbol{\mu} = \mathbf{E}(\mathbf{X}_i)$  und endlicher Kovarianzmatrix  $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}_i) > \mathbf{0}$ . Dann gilt:

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^n \mathbf{X}_i - n\boldsymbol{\mu}\right) \xrightarrow{D} N_p(\mathbf{0}, \boldsymbol{\Sigma}),$$

wobei  $N_p(\mathbf{0}, \boldsymbol{\Sigma})$  eine  $p$ -dimensionale Normalverteilung mit Erwartungswertvektor  $\mathbf{0}$  und Kovarianzmatrix  $\boldsymbol{\Sigma}$  ist.

**Beispiel 5.8.6.** Sei  $\mathbf{X}_i \sim M_p(1, \boldsymbol{\pi})$ , d.h.  $\boldsymbol{\mu} = \boldsymbol{\pi}$  und  $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\pi}) + \boldsymbol{\pi}\boldsymbol{\pi}^T$ . Dann gilt für  $\mathbf{S}_n = \sum_{i=1}^n \mathbf{X}_i$ :

$$\frac{1}{\sqrt{n}}(\mathbf{S}_n - n\boldsymbol{\mu}) \xrightarrow{D} N_p(\mathbf{0}, \boldsymbol{\Sigma}),$$

was gewissermaßen die multivariate Verallgemeinerung von Beispiel 5.8.4 ist.

Die Delta-Methode (kurz:  $\Delta$ -Methode) trifft nun eine Aussage über die Konvergenz von  $g(T_n)$ , wobei  $g$  eine (bis auf Regularitätsbedingungen) beliebige Funktion sei.

**Satz 5.8.7** (Delta-Methode /  $\Delta$ -Methode). Sei  $T_n = \frac{1}{n}(X_1 + \dots + X_n)$ , wobei  $X_i$  iid mit Erwartungswert  $\mu$  und Varianz  $\sigma^2$ . Sei ferner  $g$  eine (zumindest in einer Umgebung von  $\mu$ ) stetig differenzierbare Funktion mit Ableitung  $g'$  und  $g'(\mu) \neq 0$ . Dann gilt:

$$\sqrt{n}(g(T_n) - g(\mu)) \xrightarrow{D} N(0, [g'(\mu)]^2 \cdot \sigma^2)$$

für  $n \rightarrow \infty$ .

Vereinfachend kann man sich merken, dass aus  $Z \stackrel{a}{\sim} N(\nu, \tau^2)$

$$g(Z) \stackrel{a}{\sim} N(g(\nu), [g'(\nu)]^2 \cdot \tau^2)$$

folgt.

**Beispiel 5.8.8.** Sei  $X_i \stackrel{iid}{\sim} B(p)$ . Dann ist bekanntermaßen

$$\begin{aligned} E(X_i) &= \mu = p, \\ \text{Var}(X_i) &= \sigma^2 = p(1-p) \end{aligned}$$

und

$$\begin{aligned} \sum_{i=1}^n X_i &\sim B(n, p), \\ \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i &\stackrel{a}{\sim} N\left(p, \frac{p(1-p)}{n}\right). \end{aligned}$$

Häufig interessiert man sich für die *Chancen* (engl. *odds*)

$$g_1(p) = \frac{p}{1-p}$$

oder für die *log-Odds*

$$g_2(p) = \log \frac{p}{1-p}.$$

Nun ist aber

$$g'_1(p) = \frac{1}{(1-p)^2}$$

und

$$g'_2(p) = \frac{1}{p(1-p)}.$$

Die Anwendung der  $\Delta$ -Methode liefert:

$$\underbrace{\frac{\bar{X}_n}{1 - \bar{X}_n}}_{\text{emp. Chance}} \stackrel{a}{\sim} N\left(\frac{p}{1-p}, \underbrace{\left[\frac{1}{(1-p)^2}\right]^2 \cdot \frac{p(1-p)}{n}}_{=\frac{p}{(1-p)^3 \cdot n}}\right)$$

und

$$\underbrace{\log \frac{\bar{X}_n}{1 - \bar{X}_n}}_{\text{emp. log-Chance}} \stackrel{a}{\sim} N\left(\log \frac{p}{1-p}, \frac{1}{p(1-p) \cdot n}\right).$$

Mit der asymptotischen Verteilung von  $\log(\bar{X}_n/(1 - \bar{X}_n))$  lassen sich somit Konfidenzintervalle für die Log-Odds  $\log(p/(1 - p))$  konstruieren, die durch Rücktransformation mit der inversen *Logit-Funktion*  $1/(1 + \exp(-x))$  zu Konfidenzintervallen für  $p$  umgerechnet werden können. Diese Intervalle haben den Vorteil, dass die Intervallgrenzen immer im Intervall  $(0, 1)$  liegen.

Eine multivariate Verallgemeinerung der  $\Delta$ -Methode liefert

**Satz 5.8.9** (Multivariate Delta-Methode). *Sei  $\mathbf{T}_n = \frac{1}{n}(\mathbf{X}_1 + \dots + \mathbf{X}_n)$ , wobei die  $p$ -dimensionalen Zufallsvektoren  $\mathbf{X}_i$  iid mit Erwartungswert  $\boldsymbol{\mu}$  und Kovarianzmatrix  $\boldsymbol{\Sigma}$  seien. Sei ferner  $g : \mathbb{R}^p \rightarrow \mathbb{R}^q$  ( $q \leq p$ ) eine in einer Umgebung von  $\boldsymbol{\mu}$  stetig differenzierbare Abbildung, wobei die  $q \times p$ -Matrix*

$$\mathbf{D} = \left( \frac{\partial g_i(\mathbf{x})}{\partial \mathbf{x}_j} \right)_{ij}$$

vollen Rang  $q$  besitze. Dann gilt:

$$\sqrt{n}(g(\mathbf{T}_n) - g(\boldsymbol{\mu})) \xrightarrow{D} N_q(\mathbf{0}, \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^T)$$

für  $n \rightarrow \infty$ .

## 6 Konvergenz von Zufallsvariablen

### 6.1 Einführung

Bei der Konvergenz von Zufallsvariablen handelt es sich um die Konvergenz von Funktionen auf Wahrscheinlichkeitsräumen.

Im Weiteren sei  $f_1, f_2, \dots$  eine Folge von Funktionen, die das Intervall  $[0, 1]$  nach  $\mathbb{R}$  abbilden. Man unterscheidet:

**Definition 6.1.1** (Punktweise Konvergenz). Die Folge  $f_1, f_2, \dots$  *konvergiert punktweise* gegen die Funktion  $f$ , wenn für alle  $x \in [0, 1]$

$$f_n(x) \longrightarrow f(x) \quad \text{für } n \rightarrow \infty$$

gilt.

**Definition 6.1.2** (Konvergenz bzgl. einer Norm). Die Folge  $f_1, f_2, \dots$  *konvergiert bzgl. der Norm  $\|\cdot\|$*  gegen die Funktion  $f$ , wenn

$$\|f_n - f\| \longrightarrow 0 \quad \text{für } n \rightarrow \infty.$$

Beispielsweise lautet die  $L_p$ -Norm:

$$\|g\|_p = \left( \int_0^1 |g(x)|^p dx \right)^{\frac{1}{p}}$$

für  $p \geq 1$  und  $\|g\|_p < \infty$ .

**Definition 6.1.3** (Konvergenz bzgl. eines Maßes). Definiere eine Distanz zwischen zwei Funktionen  $g, h : [0, 1] \rightarrow \mathbb{R}$  durch

$$d_\varepsilon(g, h) = \int_E dx$$

mit festem  $\varepsilon > 0$  und

$$E = \{u \in [0, 1] : |g(u) - h(u)| > \varepsilon\}.$$

Die Folge  $f_1, f_2, \dots$  *konvergiert im Maß* gegen die Funktion  $f$ , falls

$$d_\varepsilon(f_n, f) \longrightarrow 0 \quad \text{für } n \rightarrow \infty$$

und alle  $\varepsilon > 0$ .

### 6.2 Konvergenzarten von Zufallsvariablen

**Definition 6.2.1.** Sei  $X_1, X_2, \dots$  eine Folge von Zufallsvariablen auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$ . Man sagt,

1.  $X_n \rightarrow X$  *fast sicher (f.s., engl. almost surely, a.s.)*, falls

$$P\{\omega \in \Omega : \underbrace{X_n(\omega) \rightarrow X(\omega)}_{\text{punktweise Konvergenz}} \text{ für } n \rightarrow \infty\} = 1.$$

punktweise Konvergenz  
gemäß Definition 6.1.1

Die fast sichere Konvergenz ist schwächer als die punktweise Konvergenz.

2.  $X_n \rightarrow X$  im  $r$ -ten Mittel,  $r \geq 1$ , falls  $E(|X_n^r|) < \infty$  für alle  $n$  und

$$E(|X_n - X|^r) \rightarrow 0 \text{ für } n \rightarrow \infty.$$

Zu beachten ist, dass  $r = \text{const.}$  Siehe auch Definition 6.1.2.

3.  $X_n \rightarrow X$  in Wahrscheinlichkeit, falls

$$P(|X_n - X| > \varepsilon) \rightarrow 0 \text{ für } n \rightarrow \infty \text{ und } \varepsilon > 0.$$

4.  $X_n \rightarrow X$  in Verteilung, falls

$$P(X_n \leq x) \rightarrow P(X \leq x) \text{ für } n \rightarrow \infty$$

an allen Punkten  $x \in \mathbb{R}$ , an denen  $F_X(x) = P(X \leq x)$  stetig ist.

Folgende Schreibweisen haben sich für die in Definition 6.2.1 beschriebenen Konvergenzarten durchgesetzt:

1.  $X_n$  konvergiert fast sicher (almost surely) gegen  $X$ :

$$\begin{aligned} X_n &\xrightarrow{\text{f.s.}} X, \\ X_n &\xrightarrow{\text{a.s.}} X. \end{aligned}$$

Man sagt auch,  $X_n$  konvergiert *mit Wahrscheinlichkeit 1* gegen  $X$ .

2.  $X_n$  konvergiert im  $r$ -Mittel gegen  $X$ :

$$X_n \xrightarrow{r} X.$$

3.  $X_n$  konvergiert in Wahrscheinlichkeit (in probability) gegen  $X$ :

$$X_n \xrightarrow{P} X.$$

4.  $X_n$  konvergiert in Verteilung (in distribution) gegen  $X$ :

$$X_n \xrightarrow{D} X.$$

**Bemerkung:** Man spricht bei

- $X_n \xrightarrow{1} X$  von *Konvergenz im Mittel*,
- $X_n \xrightarrow{2} X$  von *Konvergenz im quadratischen Mittel*.

**Beispiel 6.2.2.** Sei  $X \sim B(\frac{1}{2})$  Bernoulli verteilt und  $X_1, X_2, \dots$  Zufallsvariablen mit  $X_n = X$  für alle  $n \in \mathbb{N}$ . Dann  $X_n \rightarrow X$  in allen vier Konvergenzarten, da die Variablen dieselben sind. Betrachte nun  $Y = 1 - X$ . Dann ist  $F_X(x) = F_Y(x)$ , da  $Y \sim B(\frac{1}{2})$ . Daher  $X_n \xrightarrow{D} Y$ . Aber  $X_n$  konvergiert auf keine andere Art gegen  $Y$ , da  $|X_n - Y| = 1$  immer gilt.

**Satz 6.2.3.** *Es gelten folgende Zusammenhänge zwischen den Konvergenzarten:*

$$\begin{aligned} X_n \xrightarrow{f.s.} X &\implies X_n \xrightarrow{P} X \\ X_n \xrightarrow{r} X &\implies X_n \xrightarrow{P} X \\ X_n \xrightarrow{P} X &\implies X_n \xrightarrow{D} X \end{aligned}$$

und somit auch

$$\begin{aligned} X_n \xrightarrow{f.s.} X &\implies X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X \\ X_n \xrightarrow{r} X &\implies X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X \end{aligned}$$

für alle  $r \geq 1$ . Ferner gilt für  $r \geq s \geq 1$ :

$$X_n \xrightarrow{r} X \implies X_n \xrightarrow{s} X.$$

Alle weitere Implikationen sind im Allgemeinen falsch! Es gibt jedoch weitere Spezialfälle.

**Satz 6.2.4.** *Es gilt:*

1. Falls  $X_n \xrightarrow{D} c$ , wobei  $c$  eine Konstante ist, dann gilt auch  $X_n \xrightarrow{P} c$ , also

$$X_n \xrightarrow{D} c \iff X_n \xrightarrow{P} c.$$

2. Falls  $X_n \xrightarrow{P} X$  und  $P(|X_n| \leq k) = 1$  für alle  $n \in \mathbb{N}$  und  $k$  beliebig aber konstant, dann folgt auch  $X_n \xrightarrow{r} X$  für alle  $r \geq 1$ , also

$$X_n \xrightarrow{P} X \iff X_n \xrightarrow{r} X.$$

3. Falls für  $P_n(\varepsilon) = P(|X_n - X| > \varepsilon)$  gilt:  $\sum_n P_n(\varepsilon) < \infty$  für alle  $\varepsilon > 0$ , so gilt

$$X_n \xrightarrow{f.s.} X.$$

*Beweise zu den Sätzen 6.2.3 und 6.2.4.* Siehe Buch von Grimmett und Stirzaker. Im Folgenden sei nur ein Beispiel zu  $X_n \xrightarrow{P} X \not\Rightarrow X_n \xrightarrow{1} X$  genannt.  $\square$

**Beispiel 6.2.5.** Sei  $\{X_n\}_{n \in \mathbb{N}}$  unabhängig mit

$$X_n = \begin{cases} n^3 & \text{mit Wahrscheinlichkeit } n^{-2}, \\ 0 & \text{mit Wahrscheinlichkeit } 1 - n^{-2}. \end{cases}$$

Dann ist  $P(|X_n - 0| > \varepsilon) = n^{-2}$ , für große  $n$  gilt also  $X_n \xrightarrow{P} 0$ . Aber

$$E(|X_n|) = n^3 \cdot n^{-2} = n \longrightarrow \infty$$

für  $n \rightarrow \infty$ , und somit gilt *nicht*  $X_n \xrightarrow{1} 0$ .

Satz 6.2.3 impliziert, dass Konvergenz in Verteilung die schwächste Form der Konvergenz ist. Interessant ist aber der

**Satz 6.2.6** (Repräsentationssatz von Skorokhod). *Seien  $\{X_n\}$  und  $X$  Zufallsvariablen mit Verteilungsfunktionen  $\{F_n\}$  und  $F$  und  $X_n \xrightarrow{D} X$  (bzw.  $F_n \rightarrow F$ ) für  $n \rightarrow \infty$ . Dann gibt es einen Wahrscheinlichkeitsraum  $(\Omega', \mathcal{F}', P')$  und Zufallsvariablen  $\{Y_n\}$  und  $Y$  von  $\Omega'$  nach  $\mathbb{R}$  mit*

- $\{Y_n\}$  und  $Y$  haben Verteilungsfunktion  $\{F_n\}$  und  $F$ ,
- $Y_n \xrightarrow{f.s.} Y$  für  $n \rightarrow \infty$ .

*Es gibt also eine Folge  $\{Y_n\}$  von Zufallsvariablen, identisch verteilt wie  $\{X_n\}$ , die gegen eine Kopie von  $X$  konvergiert.*

**Satz 6.2.7.** *Sei  $X_n \xrightarrow{D} X$  und  $g: \mathbb{R} \rightarrow \mathbb{R}$  stetig. Dann gilt:*

$$g(X_n) \xrightarrow{D} g(X).$$

*Beweis.* Definiere  $\{Y_n\}$  und  $Y$  wie in Satz 6.2.6. Dann  $Y_n \xrightarrow{f.s.} Y$ . Wegen der Stetigkeit von  $g$  gilt:

$$\begin{aligned} \{\omega : g(Y_n(\omega)) \rightarrow g(Y(\omega))\} &\supseteq \{\omega : Y_n(\omega) \rightarrow Y(\omega)\} \\ \implies g(Y_n) &\xrightarrow{f.s.} g(Y) \\ \implies g(Y_n) &\xrightarrow{D} g(Y). \end{aligned}$$

Da aber  $g(Y_n)$  und  $g(Y)$  die gleiche Verteilung wie  $g(X_n)$  und  $g(X)$  besitzen, folgt:

$$g(X_n) \xrightarrow{D} g(X).$$

□

Wir bemerken an dieser Stelle dass für  $g(x)$  stetig offensichtlich auch

$$X_n \xrightarrow{f.s.} X \implies g(X_n) \xrightarrow{f.s.} g(X)$$

gilt, dies ist ja ein Teil des Beweises. Die analoge Aussage gilt auch für Konvergenz in Wahrscheinlichkeit (siehe Übungsaufgabe):

$$X_n \xrightarrow{P} X \implies g(X_n) \xrightarrow{P} g(X).$$

### 6.3 Einige zusätzliche Resultate

**Satz 6.3.1.** *Sei  $h: \mathbb{R} \rightarrow [0, \infty)$  eine nicht-negative Funktion. Dann gilt:*

$$P(h(X) \geq a) \leq \frac{E(h(X))}{a} \quad \text{für alle } a > 0.$$

Spezialfälle sind die folgenden zwei Ungleichungen:

**Satz 6.3.2** (MARKOVsche Ungleichung, 1856-1922). *Es gilt:*

$$P(|X| \geq a) \leq \frac{E(|X|)}{a} \quad \text{für alle } a > 0.$$

**Satz 6.3.3** (CHEBYSHEV'sche Ungleichung, 1821-1894). *Es gilt:*

$$P(X^2 \geq a) \leq \frac{E(X^2)}{a} \quad \text{für alle } a > 0$$

und somit auch

$$P((X - E(X))^2 \geq a) \leq \frac{\text{Var}(X)}{a} \quad \text{für alle } a > 0.$$

*Beweis zu Satz 6.3.1.* Sei  $A$  das Ereignis  $\{h(X) \geq a\}$ . Dann gilt:

$$h(X) \geq a \cdot I_A,$$

wobei  $I_A$  die Indikatorfunktion des Ereignisses  $A$  ist, und daher:

$$E(h(X)) \geq a \cdot E(I_A) = a \cdot P(h(X) \geq a),$$

also:

$$P(h(X) \geq a) \leq \frac{E(h(X))}{a} \quad \text{für alle } a > 0.$$

□

Die Markov'sche Ungleichung wird gerne zum Beweis von

$$X_n \xrightarrow{1} X \implies X_n \xrightarrow{P} X$$

verwendet: Es gelte  $X_n \xrightarrow{1} X$ , d.h.

$$E(|X_n - X|) \rightarrow 0 \quad \text{für } n \rightarrow \infty.$$

Bleibt also zu zeigen, dass  $X_n \xrightarrow{P} X$ , d.h.

$$P(|X_n - X| > \varepsilon) \rightarrow 0 \quad \text{für alle } \varepsilon > 0 \text{ und } n \rightarrow \infty.$$

Nun ist

$$P(|X_n - X| > \varepsilon) \leq \frac{E(|X_n - X|)}{\varepsilon} \rightarrow 0 \quad \text{für alle } \varepsilon > 0,$$

da  $E(|X_n - X|) \rightarrow 0$  für  $n \rightarrow \infty$ .

**Satz 6.3.4.** *Sei  $X_n \xrightarrow{f.s.} X$  und  $Y_n \xrightarrow{f.s.} Y$ . Dann gilt für die Summe  $X_n + Y_n$ :*

$$X_n + Y_n \xrightarrow{f.s.} X + Y.$$

*Ein analoges Resultat gilt auch für die Konvergenzarten  $\xrightarrow{r}$  und  $\xrightarrow{P}$ . Aus  $X_n \xrightarrow{D} X$  und  $Y_n \xrightarrow{D} Y$  folgt im Allgemeinen aber nicht  $X_n + Y_n \xrightarrow{D} X + Y$ .*

Der Fall  $\xrightarrow{D}$  wird nun separat unter leicht verschärften Bedingungen abgehandelt:

**Satz 6.3.5** (SLUTSKY's Theorem, 1880-1948). *Falls  $X_n \xrightarrow{D} X$  und  $Y_n \xrightarrow{D} a$  mit  $a \in \mathbb{R}$ , dann gilt:*

1.  $X_n + Y_n \xrightarrow{D} X + a$ .

2.  $X_n \cdot Y_n \xrightarrow{D} a \cdot X$ .

**Beispiel 6.3.6.** Seien  $X_1, \dots, X_n$  unabhängige und identisch verteilte Zufallsvariablen mit endlichem Erwartungswert  $\mu$  und endlicher Varianz  $\sigma^2$ . Das GDGZ stellt sicher, dass

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{D} \mu$$

für  $n \rightarrow \infty$ , aber was kann man zur Konvergenz der empirischen Varianz

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \left( \sum_{i=1}^n X_i^2 \right) - \bar{X}_n^2$$

sagen? Mit Satz 6.2.7 folgt zunächst, dass

$$\bar{X}_n^2 \xrightarrow{D} \mu^2.$$

Da  $E(X_i^2) = \sigma^2 + \mu^2$ , gilt weiterhin mit dem GDGZ

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{D} \sigma^2 + \mu^2,$$

und mit Satz 6.3.5 folgt somit

$$S_n^2 \xrightarrow{D} \sigma^2.$$

Eine analoge Aussage erhält man, wenn man in  $S_n^2$  den Faktor  $1/n$  durch  $1/(n-1)$  ersetzt. Dann gilt sogar  $E(S_n) = \sigma^2$  für alle  $n \in \mathbb{N}$ .

## 6.4 Gesetze der großen Zahlen

**Satz 6.4.1.** Seien  $X_1, X_2, \dots$  iid mit  $E(X_i^2) < \infty$  und  $E(X_i) = \mu$ . Dann gilt

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{2} \mu \quad \text{für } n \rightarrow \infty.$$

**Bemerkung:** Wir haben hier stärkere Voraussetzung aber auch eine stärkere Konvergenzaussage als in Satz 5.8.2, bei dem nur

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{D} \mu \quad \text{für } n \rightarrow \infty.$$

unter der Annahme  $E(|X_i|) < \infty$  gezeigt wurde. Mit Satz 6.2.4, Teil 1, folgt aber aus Satz 5.8.2 zumindest Konvergenz nach Wahrscheinlichkeit, d.h.

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mu \quad \text{für } n \rightarrow \infty.$$

Deswegen wird Satz 5.8.2 auch das *schwache Gesetz der großen Zahlen* genannt.

*Beweis zu Satz 6.4.1.* Sei  $S_n = \sum_{i=1}^n X_i$ . Dann konvergiert  $\frac{1}{n}S_n$  im Quadratmittel gegen  $\mu$  da

$$\begin{aligned} \mathbb{E}\left(\left(\frac{1}{n}S_n - \mu\right)^2\right) &= \mathbb{E}\left(\frac{1}{n^2}(S_n - \mathbb{E}(S_n))^2\right) \\ &= \frac{1}{n^2} \text{Var}(S_n) \\ &\stackrel{\text{unabh.}}{=} \frac{1}{n^2} (n \cdot \text{Var}(X_i)) \\ &= \frac{1}{n} \text{Var}(X_i) \longrightarrow 0 \end{aligned}$$

für  $n \rightarrow \infty$ . □

Für die Konvergenz im Quadratmittel ist die Bedingung  $\mathbb{E}(X_i^2) < \infty$  notwendig und hinreichend. Für fast sichere Konvergenz genügt aber  $\mathbb{E}(|X_i|) < \infty$ :

**Satz 6.4.2** (Starkes Gesetz der großen Zahlen, engl. Strong Law of Large Numbers, SLLN). *Seien  $X_1, X_2, \dots$  iid Zufallsvariablen. Dann gilt:*

$$\frac{1}{n} S_n \xrightarrow{f.s.} \mu \quad \text{für } n \rightarrow \infty \text{ und } \mu = \text{const.}$$

*genau dann, wenn  $\mathbb{E}(|X_i|) < \infty$ . Dann ist  $\mu = \mathbb{E}(X_i)$ .*

Notwendig und hinreichend für die schwache Konvergenz ist folgende Eigenschaft:

**Satz 6.4.3.** *Sei  $\{X_n\}$  eine Folge von iid Zufallsvariablen mit gemeinsamer Verteilungsfunktion  $F$ . Für ein  $\mu$  gilt:*

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mu$$

*genau dann, wenn die CF  $\phi(t)$  der  $X_i$ 's im Punkte  $t = 0$  differenzierbar ist und  $\phi'(0) = i\mu$  gilt.*

**Beispiel 6.4.4.** Da bekanntermaßen bei der Standard-Cauchy-Verteilung die CF  $\phi(t) = \exp(-|t|)$  lautet, kann schwache Konvergenz nicht gelten, da die Differenzierbarkeit in  $t = 0$  nicht gegeben ist. Es gilt sogar:

$$\phi_{\frac{1}{n}S_n}(t) = \left[ \exp\left(-\frac{|t|}{n}\right) \right]^n = \exp(-|t|),$$

was bedeutet, dass auch  $\frac{1}{n}S_n$  Standard-Cauchy-verteilt ist, egal, wie groß  $n$  ist.

## Stichwortverzeichnis

- $\chi^2$ -Verteilung, 72
- $\sigma$ -Additivität, 5
- $\sigma$ -Algebra, 4, 15
  - von  $\mathcal{F}$  erzeugte, 16
  - Borelsche, 51
  - kleinste, 4
  - triviale, 4
- Abhängigkeit
  - lineare, 37
  - stochastische, 61
- absolut konvergent, 30
- acceptance step, 87
- Akzeptanzwahrscheinlichkeit, 87
- Algebra, 4
- antiton, 7
- asymptotisch verteilt, 110
- Axiome von Kolmogorov, 5
- Basisreproduktionszahl, 48
- Bayes, Thomas, 10
- bedingt unabhängig, 15
- bedingte(r,s)
  - Erwartung, 43, 67
  - Erwartungswert, 43, 45, 67
  - Varianz, 45
  - Wahrscheinlichkeit, 7
  - Wahrscheinlichkeitsfunktion, 42
- Bernoulli-
  - variable, 19
  - verteilung, 19, 27
    - Quantilsfunktion, 33
- Beta-
  - Binomialverteilung, 69, 70
  - funktion, 58
  - verteilung, 58
- Betrag, 104
- Binomialverteilung, 24, 27
  - Approximation, 36
  - negative, 50
- bivariate
  - Normalverteilung, 64, 68
  - Standardnormalverteilung, 80
- Borel, Emile, 51
- Borelmenge, 51
- Borelsche  $\sigma$ -Algebra, 51, 100
- Buffon's Nadel, 25
- Cauchy, Augustin Louis, 38
- Cauchy-Schwarz-Ungleichung, 38
- Cauchy-Verteilung, 57, 75
  - Standard-, 57
- Chance, 112
  - empirische, 112
- charakteristische Funktion, 103
- Chebychev, Pafnuty, 117
- Chebychevsche Ungleichung, 117
- Cholesky-Zerlegung, 82
- Dart, 25
- definit
  - nicht-negativ, 103
  - positiv, 80
  - positiv semi-, 40
- degeneriert, 59
- Delta-Methode, 112
  - multivariate, 113
- Dichtefunktion, 20, 51
  - bedingte, 66
  - gemeinsame, 25, 61
  - Rand-, 61
- Differenz, 3
  - symmetrische, 3
- disjunkt, 3
- disjunkte Zerlegung, 8
- diskret, 20, 23
  - gemeinsam, 23
- diskret-stetig, 22
- Dispersionsfaktor, 72
- dominierte Konvergenz, 99
- Eindeutigkeitssatz, 107
- Elementarereignis, 2
- Entfaltung, 80
- Epidemie, 47
- Ereignis, 2
  - Elementar-, 2
  - fast sicheres, 7
  - Gegen-, 3

- Null-, 7, 100
- sicheres, 3
- unabhängiges, 14
- unmögliches, 3
- Ergebnis, 2
- raum, 2
- Erwartung
  - bedingte, 43, 67
- erwartungstreu, 84
- Erwartungswert, 29, 31, 52, 61, 97
- vektor, 82
- bedingter, 43, 45, 67
- eines Zufallsvektors, 35
- iterierter, 44, 67
- erzeugende Funktion, 88
- exponential-, 88
- gewöhnliche, 88
- kumulanten-, 102
- wahrscheinlichkeits-, 90
- Erzeugendensystem, 16
- exponentialerzeugende Funktion, 101
- Exponentialverteilung, 54
- Expositions-Odds Ratio, 11
  
- F-Verteilung, 78
- faktorielles Moment, 91
- Fall-Kontroll-Studie, 11
- fallende Folge, 7
- Faltung, 49, 78, 88
- Faltungssatz, 49, 78
- fast sicher, 7, 19, 114
- Fatou, Pierre, 100
- Finetti, Bruno de, 2
- Fourier, Jean, 107
- Fourier-Transformation
  - inverse, 107
- Freiheitsgrade, 58
  
- Gamma-
  - funktion, 56
  - verteilung, 55
  - inverse, 74
- Gegenereignis, 3
- geometrische Verteilung, 36
- Gesetz
  - der großen Zahlen, 20, 108
  - schwaches, 119
  - starkes, 120
  - vom Durchschnitt, 20
  - von de Morgan, 5
- Gleichverteilung, 54
- Grund-
  - menge, 2
  - raum, 2
  
- harmonische Reihe, 30
- Hit or Miss, 26
  
- iid, 94
- Indikatorfunktion, 19
- Informationsungleichung, 60
- Integral, 98
- inverse
  - Fourier-Transformation, 107
  - Gammaverteilung, 74
- Inversionsverfahren, 33, 85
- isoton, 6
- Iterierte, 95
  
- Jacobi Determinante, 76
- Jensen, Johan, 59
- Jensensche Ungleichung, 59
  
- Kern, 55
- Kolmogorov, Andrei, 5
- Komplement, 3
- konkav, 59
- strikt, 59
- konstant, 19
- Konvergenz
  - bzgl. einer Norm, 114
  - bzgl. eines Maßes, 114
  - dominierte, 99
  - fast sichere, 114
  - im Mittel, 115
  - im quadratischen Mittel, 115
  - in Verteilung, 108, 115
  - in Wahrscheinlichkeit, 115
  - mit Wahrscheinlichkeit 1, 115
  - monotone, 99
  - punktweise, 114
  - von Verteilungsfunktionen, 107
- konvex, 59
- strikt, 59
- Kopie, 117

Korrelation, 38  
     lineare Transformationen, 40  
 Korrelationsmatrix, 42  
 Kovarianz, 38  
     lineare Transformationen, 40  
 Kovarianzmatrix, 40, 82  
 Krankheits-Odds Ratio, 11  
 Kullback, Solomon, 60  
 Kullback-Leibler-  
     Diskrepanz, 60  
     Distanz, 60  
 Kumulante, 102  
 kumulantenerzeugende Funktion, 102  
 Kurtosis, 103  
     Exzess, 103  
     standardisierte, 103  
  
 Laplace, Pierre-Simon de, 111  
 Law  
     of Averages, 20  
     of Large Numbers, 20  
 Lebesgue, Henri, 100  
 Lebesgue-Stieltjes-Integral, 100  
 Leibler, Richard, 60  
 Lemma von Fatou, 100  
 lineare Abhängigkeit, 37  
 Log-Normalverteilung, 60, 74  
 log-Odd, 112  
 Logit-Funktion, 113  
  
 Münzwurf, 2, 5, 18, 27, 28  
     3-seitige Münze, 24  
 Marginalverteilung, 23  
 Markov, Andrei, 117  
 Markov-Eigenschaft, 69  
 Markovsche Ungleichung, 117  
 messbar, 18  
 Mises, Richard von, 2  
 Moivre, Abraham de, 111  
 Moment, 31  
     faktorielles, 91  
     zentrales, 31  
 momentenerzeugende Funktion, 101  
 monotone Konvergenz, 99  
 Monotonie, 6, 18, 23  
 Monte Carlo  
     Integration, 26  
     Simulation, 25  
 Morgan, Augustus de, 5  
 Multinomialverteilung, 25, 42  
 multivariate Normalverteilung, 82  
  
 negative Binomialverteilung, 50  
 nicht-negativ definit, 103  
 Norm, 114  
 Normalisierungskonstante, 55  
 Normalverteilung, 55  
     bivariate, 64  
     Log-, 60  
     multivariate, 82  
     n-dimensionale, 80  
     Standard-, 55  
 Normiertheit, 5, 18, 23  
 null, 7  
 Nullereignis, 7, 100  
  
 O.J. Simpson Prozess, 12  
 Odd, 112  
 Odds Ratio, 11  
  
 paarweise unabhängig, 15  
 Partialsumme, 108  
     mittlere, 108  
 Partition, 8  
 Poincaré, Jules Henri, 6  
 Poissonverteilung, 27, 35  
     verschobene, 45  
 positiv  
     definit, 80  
     semi-definit, 40  
 Positivität, 5, 23  
 Potenzmenge, 4  
 Präzisionsmatrix, 82  
 Prisoners' paradox, 9  
 Produkt-  
     maß, 16  
     raum, 16  
 pull through property, 45  
  
 quadratische Form, 80  
 Quantil, 33  
 Quantilsfunktion, 33  
  
 R, 37  
 Rand-

dichtefunktion, 61  
 verteilung, 23, 61  
 Random Walk, 16  
 Rechtsstetigkeit, 19  
 Rejection Sampling, 86  
   Algorithmus, 86  
 rejection step, 87  
 relatives Risiko, 11  
 Satz  
   von der totalen Wahrscheinlichkeit, 8  
   vom iterierten Erwartungswert, 44, 67  
   von Abel, 90  
   von Bayes, 10  
     bei zwei bedingten Ereignissen, 12  
     Variante, 11  
   von Bochner, 104  
   von Taylor, 105  
 Schiefe, 103  
   standardisierte, 103  
 Schwarz, Hermann Amandus, 38  
 Schwellenwerttheorem, 48  
 semi-definit, 40  
 Siebformel von Sylvester-Poincaré, 6  
 Simpson, Orenthal James, 12  
 Slutsky's Theorem, 118  
 Slutsky, Eugene, 118  
 Standard-  
   abweichung, 31  
   Cauchy-Verteilung, 57  
   normalverteilung, 55  
     bivariate, 63  
 Standardnormalverteilung  
   bivariate, 68, 80  
   univariate, 80  
 stationäre Verteilung, 97  
 stetig, 20  
   gemeinsam, 25  
   von oben, 23  
   von rechts, 19  
 Stetigkeit von E, 99  
 Stetigkeitssatz, 108  
 Stieltjes, Thomas, 100  
 stochastisch unabhängig, 14  
 stochastische Abhängigkeit, 61  
 Studentverteilung, 58  
 Sylvester, James, 6  
 symmetrisch, 40  
 t-Verteilung, 58  
   Standard-, 58  
 Taylor, Brook, 105  
 totale Wahrscheinlichkeit, 8  
 tower property, 45  
 Träger, 20, 27, 51  
 Transformationssatz, 73, 76  
 Trinomialverteilung, 24, 35  
   Kovarianzmatrix, 41  
 Ueberdispersion, 69  
 unabhängig, 14, 28  
   bedingt, 15  
   paarweise, 15  
   stochastisch, 14  
 Unabhängigkeit, 14  
   bedingte, 15, 29  
   einer Familie, 14, 29  
   von Ereignissen, 14  
   von Zufallsvariablen, 28, 52, 62  
 univariate Standardnormalverteilung, 80  
 unkorreliert, 32  
 Urnen, 9  
 Varianz, 31  
   bedingte, 45  
 Varianzzerlegungssatz, 45, 67  
 Vermengung, 94  
 Verschiebungssatz, 31  
 Verteilung  
   Marginal-, 23  
   Rand-, 23, 61  
   stationäre, 97  
 Verteilungsfunktion, 18, 27  
   bedingte, 42, 66  
   gemeinsame, 22, 61  
 Verwerfungsmethode, 86  
 Verzweigungsprozesse, 47  
 Vollständigkeit, 16  
 Würfelwurf, 3, 5  
 wachsende Folge, 6  
 Wahrscheinlichkeit, 2, 5  
   bedingte, 7  
   Satz von der totalen, 8  
   totale, 8

Wahrscheinlichkeits-  
begriff  
  frequentistischer, 2, 20  
  subjektivistischer, 2  
funktion, 20, 27  
  bedingte, 42  
  gemeinsame, 23  
maß, 5  
raum, 5  
wahrscheinlichkeitserzeugende Funktion, 90

Zentraler Grenzwertsatz, 108, 109  
  multivariat, 111

Zufallsvariable, 18  
  deterministische, 19  
  diskret-stetige, 22  
  diskrete, 20  
  einfache, 98  
  konstante, 19  
  stetige, 20, 51

Zufallsvektor, 22