



**TECHNISCHE  
UNIVERSITÄT  
DRESDEN**

---

**Fakultät Mathematik** Institut für Stochastik, Professur für Angewandte Stochastik

---

# STOCHASTIK

**Prof. Dr. Anita Behme**

Sommersemester 2019

Autor : Eric Kunze  
E-Mail : [eric.kunze@mailbox.tu-dresden.de](mailto:eric.kunze@mailbox.tu-dresden.de)

# Inhaltsverzeichnis

<b>0</b>	<b>Einleitung</b>	<b>2</b>
<b>1</b>	<b>Grundbegriffe der Wahrscheinlichkeitstheorie</b>	<b>3</b>
1.1	Wahrscheinlichkeitsräume . . . . .	3
1.2	Zufallsvariablen . . . . .	8
<b>2</b>	<b>Erste Standardmodelle: Diskrete Verteilungen</b>	<b>12</b>
2.1	Diskrete Gleichverteilungen . . . . .	12
2.2	Urnenmodelle . . . . .	12
2.2.1	Urnenmodell mit Zurücklegen: Multinomial-Verteilung . . . . .	12
2.2.2	Urnenmodell ohne Zurücklegen: Hypergeometrische Verteilung . . . . .	15
2.3	Poisson-Approximation und -Verteilung . . . . .	15
<b>3</b>	<b>Bedingte Wahrscheinlichkeiten und Unabhängigkeit</b>	<b>17</b>
3.1	Bedingte Wahrscheinlichkeiten . . . . .	17
3.2	Stochastische Unabhängigkeit . . . . .	22
3.2.1	Konstruktion unabhängiger Zufallsvariablen . . . . .	27
3.3	Faltungen . . . . .	28
<b>4</b>	<b>Weitere Standardmodelle: Stetige Verteilungen</b>	<b>31</b>
4.1	Stetige Gleichverteilung . . . . .	31
4.2	Wartezeitverteilungen . . . . .	31
4.2.1	Negative Binomialverteilung . . . . .	31
4.2.2	Exponential- und Gammaverteilung . . . . .	32
<b>5</b>	<b>Erwartungswert und Varianz</b>	<b>35</b>
5.1	Der Erwartungswert . . . . .	35
5.2	Varianz und höhere Momente . . . . .	39
5.3	Wahrscheinlichkeitserzeugende Funktionen . . . . .	41
<b>6</b>	<b>Normalverteilung</b>	<b>45</b>
<b>7</b>	<b>Momenterzeugende und charakteristische Funktionen</b>	<b>47</b>
7.1	Momenterzeugende Funktionen . . . . .	47
7.2	Charakteristische Funktionen . . . . .	49

# Kapitel 0

## EINLEITUNG

### Literatur

- *Georgii* : Stochastik (5. Auflage)
- *Schilling* : Wahrscheinlichkeit
- *Bauer* : Wahrscheinlichkeitstheorie (5. Auflage)
- *Krengel* : Einführung in die W-Theorie und Statistik
- *Dehling & Haupt* : Einführung in die W-Theorie und Statistik

### Was ist Stochastik ?

Altgriechisch „Stochastikos“ ( $\sigma\tau\chi\alpha\tau\iota\kappa\acute{\omicron}$ )  $\leadsto$  „scharfsinnig im Vermuten“

Fragestellungen stammen insbesondere aus dem Glücksspiel, heute vielmehr auch aus der Versicherungs- und Finanzmathematik - überall da, wo Zufall / Risiko / Chance auftaucht.

### Was ist mathematische Stochastik ?

- Beschreibt zufällige Phänomene in einer exakten Sprache.  
Bsp.: „Beim Würfeln erscheint jedes sechste Mal (im Schnitt) die Augenzahl 6“  $\leadsto$  Gesetz der großen Zahlen
- lässt sich in zwei Teilgebiete unterteilen: Wahrscheinlichkeitstheorie & Statistik  
Die W-Theorie beschreibt und untersucht konkret gegebene Zufallssituationen. Dagegen zieht die Statistik Schlussfolgerungen aus Beobachtungen. Dabei benötigt sie die Modelle der W-Theorie - umgekehrt benötigt auch die W-Theorie die Statistik zur Bestätigung der Modelle.
- In diesem Semester konzentrieren wir uns auf die Wahrscheinlichkeitstheorie.

# Kapitel 1

## GRUNDBEGRIFFE DER WAHRSCHEINLICHKEITSTHEORIE

### 1.1 Wahrscheinlichkeitsräume

#### Ergebnisraum

Welche möglichen Ausgänge eines zufälligen Geschehens interessieren uns?

##### Beispiel

Würfeln: Augenzahl, aber nicht Lage, Fallhöhe, usw.

##### Definition 1.1 (Ergebnisraum)

Die Menge der relevanten Ergebnisse eines Zufallsgeschehens nennen wir **Ergebnisraum** und bezeichnen diesen mit  $\Omega$ .

##### Beispiel

- Würfeln:  $\Omega = \{1, 2, \dots, 6\}$
- Wartezeiten:  $\Omega = \mathbb{R}_+ = [0, \infty)$  (also überabzählbar)

#### Ereignisse

Oft interessiert man sich gar nicht für das konkrete Ergebnis des Zufallsexperiments, sondern nur für das Eintreten gewisser Ereignisse.

##### Beispiel

Würfeln: Zahl ist  $> 3$

Wartezeiten: Wartezeit ist  $\leq 5$  Minuten

Wir wollen also Teilmengen des Ergebnisraums betrachten, d.h. Elemente von  $\mathcal{P}(\Omega)$  (Potenzmenge), denen eine Wahrscheinlichkeit zugeordnet werden kann d.h. welche *messbar* sind.

**Definition 1.2 (Ereignisraum)**

Sei  $\Omega \neq \emptyset$  ein Ergebnisraum und  $\mathcal{F}$  eine  $\sigma$ -Algebra auf  $\Omega$ , d.h. eine Familie von Teilmengen von  $\Omega$ , sodass

- (1)  $\Omega \in \mathcal{F}$
- (2)  $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
- (3)  $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i \geq 1} A_i \in \mathcal{F}$

Dann heißt  $(\Omega, \mathcal{F})$  **Ereignisraum** oder messbarer Raum.

**Wahrscheinlichkeit**

Wir ordnen nun den Ereignissen Wahrscheinlichkeiten mittels einer Abbildung  $\mathbb{P}: \mathcal{F} \rightarrow [0, 1]$  zu, sodass

(N) Normierung:  $\mathbb{P}(\Omega) = 1$

(A) Additivität: Für paarweise disjunkte Ereignisse  $A_1, A_2, \dots \in \mathcal{F}$  ist  $\mathbb{P}\left(\bigcup_{i \geq 1} A_i\right) = \sum_{i \geq 1} \mathbb{P}(A_i)$ .

(N), (A) und die Nichtnegativität von  $\mathbb{P}$  werden als Kolmogorov-Axiome bezeichnet (nach Kolmogorov: Grundbegriffe der Wahrscheinlichkeitstheorie, 1933).

**Definition 1.3 (Wahrscheinlichkeit)**

Sei  $(\Omega, \mathcal{F})$  ein Ereignisraum und  $\mathbb{P}: \mathcal{F} \rightarrow [0, 1]$  eine Abbildung mit den Eigenschaften (N) und (A). Dann heißt  $\mathbb{P}$  **Wahrscheinlichkeitsmaß** oder auch **Wahrscheinlichkeitsverteilung**.

Aus der Definition folgen direkt die folgenden Eigenschaften:

**Satz 1.4 (Rechenregeln für Wahrscheinlichkeitsmaße)**

Sei  $\mathbb{P}$  ein W-Maß auf einem Ereignisraum  $(\Omega, \mathcal{F})$  und  $A, B, A_1, A_2, \dots \in \mathcal{F}$ . Dann gilt:

- (1)  $\mathbb{P}(\emptyset) = 0$
- (2) Endliche Additivität:  $\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$  und  $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$
- (3) Monotonie:  $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
- (4)  $\sigma$ -Subadditivität:  $\mathbb{P}\left(\bigcup_{i \geq 1} A_i\right) \leq \sum_{i \geq 1} \mathbb{P}(A_i)$
- (5)  $\sigma$ -Stetigkeit: Wenn  $A_n \nearrow A$  (d.h.  $A_1 \subseteq A_2 \subseteq \dots$  und  $A = \bigcup_{i=1}^{\infty} A_i$ ) oder  $A_n \searrow A$ , so gilt  $\mathbb{P}(A_n) \rightarrow \mathbb{P}(A)$  für  $n \rightarrow \infty$

*Beweis.* siehe MINT oder Schillings Lehrbuch □

**Beispiel 1.5**

Für einen beliebigen Ereignisraum  $(\Omega, \mathcal{F})$  und ein beliebiges Element  $\xi \in \Omega$  definiert

$$\delta_\xi(A) := \begin{cases} 1 & \xi \in A \\ 0 & \text{sonst} \end{cases}$$

ein (degeneriertes) W-Maß auf  $(\Omega, \mathcal{F})$ , welches wir als **Dirac-Maß** oder Dirac-Verteilung bezeichnen.

**Beispiel 1.6**

Wir betrachten das Zufallsexperiment "Würfeln mit einem fairen, 6-seitigen Würfel" mit der Ergebnismenge  $\Omega = \{1, \dots, 6\}$  und Ereignisraum  $\mathcal{F} = \mathcal{P}(\Omega)$ . Setzen wir aus Symmetriegründen

$$\mathbb{P}(A) = \frac{\#A}{6}$$

mit  $\#A = |A|$  = Kardinalität. Dies definiert ein W-Maß.

**Beispiel 1.7 (Wartezeiten an der Bushaltestelle)**

Ergebnisraum  $\Omega = \mathbb{R}_+$  und Ereignisraum Borel'sche  $\sigma$ -Algebra  $\mathcal{F} = \mathcal{B}(\mathbb{R}_+)$ . Ein mögliches W-Maß können wir durch

$$\mathbb{P}(A) := \int_A \lambda e^{-\lambda x} \, dx$$

für einen Parameter  $\lambda > 0$  festlegen. (offensichtlich gelten  $\mathbb{P}(\Omega) = 1$  und die  $\sigma$ -Additivität aufgrund der  $\sigma$ -Additivität des Integrals). Wir bezeichnen dieses Maß als **Exponentialverteilung**. (Warum gerade dieses Maß für Wartezeiten gut geeignet ist, sehen wir später.)

**Satz 1.8 (Konstruktion von WMaßen mit Dichten)**

Sei  $(\Omega, \mathcal{F})$  ein Ereignisraum.

- $\Omega$  abzählbar,  $\mathcal{F} = \mathcal{P}(\Omega)$ :

Sei  $\rho = (\rho(\omega))_{\omega \in \Omega}$  eine Folge in  $[0, 1]$  in  $\sum_{\omega \in \Omega} \rho(\omega) = 1$ , dann definiert

$$\mathbb{P}(A) = \sum_{\omega \in A} \rho(\omega), \quad A \in \mathcal{F}$$

ein (diskretes) WMaß  $\mathbb{P}$  auf  $(\Omega, \mathcal{F})$ .  $\rho$  wird als **Zähldichte** bezeichnet. Umgekehrt definiert jedes WMaß  $\mathbb{P}$  auf  $(\Omega, \mathcal{F})$  mittels  $\rho(\omega) = \mathbb{P}(\{\omega\})$ ,  $\omega \in \Omega$  eine Folge  $\rho$  mit den obigen Eigenschaften.

- $\Omega \subseteq \mathbb{R}^n$ ,  $\mathcal{F} = \mathcal{B}(\Omega)$ :

Sei  $\rho: \Omega \rightarrow [0, \infty)$  eine Funktion, sodass

- (1)  $\int_{\Omega} \rho(x) \, dx = 1$
- (2)  $\{x \in \Omega: \rho(x) \leq c\} \in \mathcal{B}(\Omega)$  für alle  $c > 0$

dann definiert  $\rho$  ein WMaß  $\mathbb{P}$  auf  $(\Omega, \mathcal{F})$  durch

$$\mathbb{P}(A) = \int_A \rho(x) \, dx = \int_A \rho d\lambda, \quad A \in \mathcal{B}(\Omega)$$

Das Integral interpretieren wir stets als Lebesgue-Integral bzgl. Lebesgue-Maß  $\lambda$ .  $\rho$  bezeichnet wir als **Dichte**, **Dichtefunktion** oder **Wahrscheinlichkeitsdichte** von  $\mathbb{P}$  und nennen ein solches  $\mathbb{P}$  (absolut) **stetig** (bzgl. dem Lebesgue-Maß).

*Beweis.* Der diskrete Fall ist klar. Im stetigen Fall folgt die Behauptung aus den bekannten Eigenschaften des Lebesgue-Integrals ( $\nearrow$  Schilling MINT, Lemma 8.9)  $\square$

**Bemerkung**

- Die eindeutige Beziehung zwischen Dichte und Wahrscheinlichkeitsmaß überträgt sich nicht auf den stetigen Fall.
  - ▷ Nicht jedes Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{B}(\Omega))$ ,  $\Omega \subset \mathbb{R}^n$  besitzt eine Dichte.
  - ▷ Zwei Dichtefunktionen definieren dasselbe Wahrscheinlichkeitsmaß, wenn sie sich nur auf einer Menge von Lebesgue-Maß 0 unterscheiden.
- Jede auf  $\Omega \subset \mathbb{R}^n$  definierte Dichtefunktion  $\rho$  lässt sich auf ganz  $\mathbb{R}^n$  fortsetzen durch  $\rho(x) = 0$ ,  $x \notin \Omega$ . Das erzeugte WMaß auf  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$  lässt mit den WMaß auf  $(\Omega, \mathcal{B}(\Omega))$  identifizieren.
- Mittels Dirac-Maß  $\delta_x$  können auch jedes diskrete WMaß auf  $\Omega \subset \mathbb{R}^n$  als WMaß auf  $\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)$  interpretieren:

$$\mathbb{P}(A) = \sum_{\omega \in A} \rho(\omega) = \int_A d \left( \sum_{\omega \in \Omega} \rho(\omega) \delta_{\omega} \right) \quad A \in \mathcal{B}(\mathbb{R}^n)$$

- stetige und diskrete WMaße lassen sich kombinieren z.B. definiert

$$\mathbb{P}(A) = \frac{1}{2} \delta_0 + \frac{1}{2} \int_A \mathbb{1}_{[0,1]}(x) \, dx, \quad A \in \mathcal{B}(\mathbb{R})$$

ein WMaß auf  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

Abschließend erinnern wir uns an:

**Satz 1.9 (Eindeutigkeitssatz für Wahrscheinlichkeitsmaße)**

Sei  $(\Omega, \mathcal{F})$  Ereignisraum und  $\mathbb{P}$  ein WMaß auf  $(\Omega, \mathcal{F})$ . Sei  $\mathcal{F} = \omega(\mathcal{G})$  für ein  $\cap$ -stabiles Erzeugendensystem  $\mathcal{G} \subset \mathcal{P}(\Omega)$ . Dann ist  $\mathbb{P}$  bereits durch seine Einschränkung  $\mathbb{P}|_{\mathcal{G}}$  eindeutig bestimmt.

*Beweis.* ↗ Schilling MINT, Satz 4.5. □

Insbesondere definiert z.B.

$$\mathbb{P}([0, a)) = \int_0^a \lambda e^{-\lambda x} dx = 1 - e^{-\lambda a}, a > 0$$

bereits die Exponentialverteilung aus Beispiel 1.7.

**Definition 1.10 (Gleichverteilung)**

Ist  $\Omega$  endlich, so heißt das WMaß mit konstanter Zähldichte

$$\rho(\omega) = \frac{1}{|\Omega|}$$

die **(diskrete) Gleichverteilung** auf  $\Omega$  und wird mit  $U(\Omega)$  notiert ( $U = uniform$ ).

Ist  $\Omega \subset \mathbb{R}^n$  eine Borelmenge mit Lebesgue-Maß  $0 < \lambda^n(\Omega) < \infty$  so heißt das WMaß auf  $(\Omega, \mathcal{B}(\Omega))$  mit konstanter Dichtefunktion

$$\rho(x) = \frac{1}{\lambda^n(\Omega)}$$

die **(stetige) Gleichverteilung** auf  $\Omega$ . Sie wird ebenso mit  $U(\Omega)$  notiert.

## Wahrscheinlichkeitsräume

**Definition 1.11 (Wahrscheinlichkeitsraum)**

Ein Tripel  $(\Omega, \mathcal{F}, \mathbb{P})$  mit  $\Omega, \mathcal{F}$  Ereignisraum und  $\mathbb{P}$  WMaß auf  $(\Omega, \mathcal{F})$ , nennen wir **Wahrscheinlichkeitsraum**.



## 1.2 Zufallsvariablen

Zufallsvariablen dienen dazu von einem gegebenen Ereignisraum  $(\Omega, \mathcal{F})$  zu einem Modellausschnitt  $(\Omega', \mathcal{F}')$  überzugehen. Es handelt sich also um Abbildungen  $X: \Omega \rightarrow \Omega'$ . Damit wir auch jedem Ereignis in  $\mathcal{F}'$  eine Wahrscheinlichkeit zuordnen können, benötigen wir

$$A' \in \mathcal{F}' \Rightarrow X^{-1}A' \in \mathcal{F}$$

d.h.  $X$  sollte **messbar** sein.

### Definition 1.12 (Zufallsvariable)

Seien  $(\Omega, \mathcal{F})$  und  $(\Omega', \mathcal{F}')$  Ereignisräume. Dann heißt jede messbare Abbildung

$$X: \Omega \rightarrow \Omega'$$

**Zufallsvariable** (von  $(\Omega, \mathcal{F})$  nach  $(\Omega', \mathcal{F}')$  / auf  $(\Omega', \mathcal{F}')$  oder **Zufallselement**.

### Beispiel 1.13

- (1) Ist  $\Omega$  abzählbar und  $\mathcal{F} = \mathcal{P}(\Omega)$ , so ist jede Abbildung  $X: \Omega \rightarrow \Omega'$  messbar und damit eine Zufallsvariable.
- (2) Ist  $\Omega \subset \mathbb{R}^n$  und  $\mathcal{F} = \mathcal{B}(\Omega)$ , so ist jede stetige Funktion  $X: \Omega \rightarrow \mathbb{R}$  messbar und damit eine Zufallsvariable.

### Satz 1.14

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $X$  eine Zufallsvariable von  $(\Omega, \mathcal{F})$  nach  $(\Omega', \mathcal{F}')$ . Dann definiert

$$\mathbb{P}'(A') := \mathbb{P}(X^{-1}(A')) = \mathbb{P}(\{X \in A'\}), \quad A' \in \mathcal{F}'$$

ein Wahrscheinlichkeitsmaß auf  $(\Omega', \mathcal{F}')$ , welches wir als **Wahrscheinlichkeitsverteilung von  $X$  unter  $\mathbb{P}$**  bezeichnen.

*Beweis.* Aufgrund der Messbarkeit von  $X$  ist die Definition sinnvoll. Zudem gelten

$$\mathbb{P}'(\Omega') = \mathbb{P}(X^{-1}(\Omega')) = \mathbb{P}(\Omega) = 1$$

und für  $A'_1, A'_2, \dots \in \mathcal{F}'$  paarweise disjunkt.

$$\begin{aligned} \mathbb{P}'\left(\bigcup_{i \geq 1} A'_i\right) &= \mathbb{P}\left(X^{-1}\left(\bigcup_{i \geq 1} A'_i\right)\right) \\ &= \mathbb{P}\left(\bigcup_{i \geq 1} X^{-1}(A'_i)\right) \\ &= \sum_{i \geq 1} \mathbb{P}(X^{-1}A'_i) \quad \text{da auch } X^{-1}A'_1, X^{-1}A'_2, \dots \text{ paarweise disjunkt sind} \\ &= \sum_{i \geq 1} \mathbb{P}'(A'_i) \end{aligned}$$

Also ist  $\mathbb{P}'$  ein Wahrscheinlichkeitsmaß. □

### Bemerkung

- Aus Gründen der Lesbarkeit schreiben wir in der Folge

$$\mathbb{P}(X \in A) = \mathbb{P}(\{\omega: X(\omega) \in A\})$$

- Ist  $X = \text{id}$ , so fallen die Begriffe Wahrscheinlichkeitsmaß und Wahrscheinlichkeitsverteilung zusammen.
- In der (weiterführenden) Literatur zur Wahrscheinlichkeitstheorie wird oft auf die Angabe eines zugrundeliegenden Wahrscheinlichkeitsraumes verzichtet und stattdessen eine “Zufallsvariable mit Verteilung  $\mathbb{P}$  auf  $\Omega$ ” eingeführt. Gemeint ist (fast) immer  $X$  als Identität auf  $(\Omega, \mathcal{F}, \mathbb{P})$  mit  $\mathcal{F} = \mathcal{P}(\Omega)$  oder  $\mathcal{F} = \mathcal{B}(\Omega)$ .
- Für die Verteilung von  $X$  unter  $\mathbb{P}$  schreibe  $\mathbb{P}_X$  und  $X \sim \mathbb{P}_X$  für die Tatsache, dass  $X$  gemäß  $\mathbb{P}_X$  verteilt ist.

### Definition 1.15

Zwei Zufallsvariablen sind **identisch verteilt**, wenn sie dieselbe Verteilung haben.

Von besonderen Interesse sind für uns die Zufallsvariablen, die nach  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  abbilden, sogenannte **reelle Zufallsvariablen**.

Da die halboffenen Intervalle  $\mathcal{B}(\mathbb{R})$  erzeugen, ist die Verteilung einer reellen Zufallsvariable durch die Werte  $(-\infty, c]$ ,  $c \in \mathbb{R}$  eindeutig festgelegt.

### Definition 1.16 (Verteilungsfunktion)

Sei  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$  ein Wahrscheinlichkeitsraum, so heißt

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

**(kumulative) Verteilungsfunktion von  $\mathbb{P}$ .** Ist  $X$  eine reelle Zufallsvariable auf beliebigen WRaum  $(\Omega, \mathcal{F}, \mathbb{P})$ , so heißt

$$F: \mathbb{R} \rightarrow [0, 1] \text{ mit } x \mapsto \mathbb{P}(X \leq x) = \mathbb{P}(X \in (-\infty, x])$$

die (kumulative) Verteilungsfunktion von  $X$ .

### Beispiel 1.17

$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$  mit  $\mathbb{P}$  Exponentialverteilung mit Parameter  $\lambda > 0$

$$\mathbb{P}(A) = \int_{A \cap [0, \infty)} \lambda e^{-\lambda x} dx \quad A \in \mathcal{B}(\mathbb{R})$$

Dann ist

$$F(x) = \mathbb{P}((-\infty, x)) = \begin{cases} 0 & x \leq 0 \\ \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x} & x > 0 \end{cases}$$

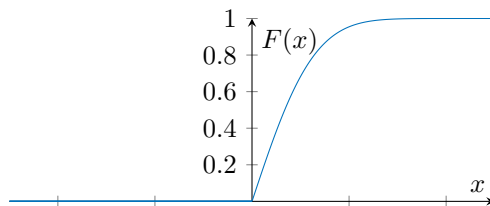


Abbildung 1.1: Verteilungsfunktion der Exponentialverteilung

**Beispiel 1.18**

Das Würfeln mit einem fairen, sechseitigen Würfel kann mittels einer reellen Zufallsvariablen

$$X: \{1, 2, \dots, 6\} \rightarrow \mathbb{R} \text{ mit } x \mapsto x$$

modelliert werden.

Es folgt als Verteilungsfunktion

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X^{-1}(-\infty, x]) = \mathbb{P}((-\infty, x]) = \frac{1}{6} \sum_{i=1}^6 \mathbb{1}_{\{i \leq x\}}$$

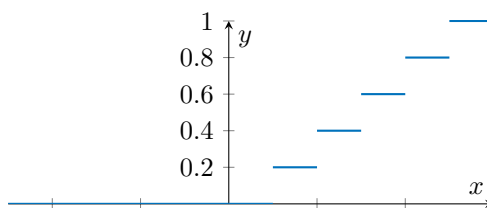


Abbildung 1.2: Verteilungsfunktion des Würfel experiments

Diese Erkenntnisse lassen sich auch verallgemeinern:

**Satz 1.19**

Ist  $\mathbb{P}$  ein Wahrscheinlichkeitsmaß auf  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  und  $F$  die zugehörige Verteilungsfunktion, so gelten

- (1)  $F$  ist monoton wachsend
- (2)  $F$  ist rechtsseitig stetig
- (3)  $\lim_{x \rightarrow -\infty} F(x) = 0$  und  $\lim_{x \rightarrow \infty} F(x) = 1$

Umgekehrt existiert zu jeder Funktion  $F: \mathbb{R} \rightarrow [0, 1]$  mit Eigenschaften (1) bis (3) eine reelle Zufallsvariable auf  $((0, 1), \mathcal{B}((0, 1)), \mathbb{U}((0, 1)))$  mit Verteilungsfunktion  $F$ .

*Beweis.* Ist  $F$  Verteilungsfunktion, so folgt mit 1.4

$$x \leq y \Rightarrow F(x) = \mathbb{P}((-\infty, x]) \stackrel{1.4.3}{\leq} \mathbb{P}((-\infty, y]) = F(y)$$

und

$$\lim_{m \searrow c} F(x) = \lim_{m \searrow c} \mathbb{P}((-\infty, x]) \stackrel{\sigma\text{-Stetigkeit}}{=} \mathbb{P}((-\infty, c]) = F(c)$$

sowie

$$\lim_{x \rightarrow -\infty} F(x) \stackrel{1.4.5}{=} \mathbb{P}(\emptyset) \stackrel{1.4.1}{=} 0 \quad \text{und} \quad \lim_{x \rightarrow \infty} F(x) \stackrel{1.4.5}{=} \mathbb{P}(\mathbb{R}) = 1$$

Umgekehrt wähle

$$X(u) := \inf \{x \in \mathbb{R} : F(x) \geq u\}, \quad u \in (0, 1)$$

Dann ist  $X$  eine “linkseitige Inverse” von  $F$  (auch **Quantilfunktion** oder **verallgemeinerte Inverse**). Wegen (3) gilt  $-\infty < X(u) < \infty$  und zudem

$$\{X \leq x\} = (0, F(x)) \cap (0, 1) \in \mathcal{B}((0, 1))$$

Da diese halboffene Mengen ein Erzeugendensystem von  $\mathcal{B}(\mathbb{R})$  bilden, folgt bereits die Messbarkeit von  $X$ , also ist  $X$  eine Zufallsvariable. Insbesondere hat die Menge  $\{X \leq x\}$  gerade LEBESGUE-Maß  $F(x)$  und damit hat  $X$  die Verteilungsfunktion  $F$ .  $\square$

### Korollar 1.20

Ist  $\mathbb{P}$  Wahrscheinlichkeitsmaß auf  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  und  $F$  die zugehörige Verteilungsfunktion. Dann besitzt  $\mathbb{P}$  genau eine Dichtefunktion  $\rho$ , wenn  $F$  stetig differenzierbar ist, denn dann gilt

$$F(x) = \int_0^x \rho(x) \, dx \quad \text{bzw.} \quad \rho(x) = F'(x)$$

*Beweis.* Folgt aus Satz 1.8, der Definition 1.16 der Verteilungsfunktion und dem Eindeutigkeitsatz 1.9.  $\square$

# Kapitel 2

## ERSTE STANDARDMODELLE: DISKRETE VERTEILUNGEN

### 2.1 Diskrete Gleichverteilungen

#### Erinnerung (Definition 1.10)

Ist  $\Omega$  endlich, so heißt Wahrscheinlichkeitsmaß mit Zähldichte  $\rho(\omega) = \frac{1}{\#\Omega}$  für  $\omega \in \Omega$  (**diskrete**) **Gleichverteilung** auf  $\Omega \sim U(\Omega)$

Es gilt das für jedes  $A \in \mathcal{P}(\Omega)$

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega}$$

Anwendungsbeispiele sind faires Würfeln, fairer Münzwurf, Zahlenlotto, ...

### 2.2 Urnenmodelle

Ein “Urnenmodell” ist eine abstrakte Darstellung von Zufallsexperimenten, bei denen zufällig Stichproben aus einer gegebenen Menge “gezogen” werden. Eine Urne ist ein Behältnis, in welchem sich farbige/nummerierte Kugeln befinden, die ansonsten ununterscheidbar sind. Aus der Urne ziehe man blind/zufällig eine oder mehrere Kugeln und notiere ihre Farbe/Zahl.

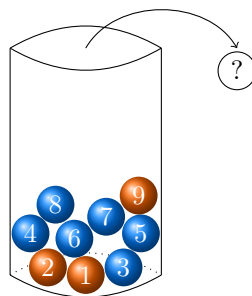


Abbildung 2.1: Urnenmodell mit nummerierten, farbigen Kugeln

#### 2.2.1 Urnenmodell mit Zurücklegen: Multinomial-Verteilung

Gegeben sei eine Urne mit  $N$  Kugeln, verschiedenfarbig mit Farben aus  $E$ , wobei  $|E| \geq 2$

Man ziehe  $n$  Stichproben/Kugeln, wobei nach jedem Zug die Kugel wieder zurückgelegt wird.

Uns interessiert die Farbe in jedem Zug, setze also

$$\Omega = E^n \text{ und } \mathcal{F} = \mathcal{P}(\Omega)$$

Zur Bestimmung eines geeigneten Wahrscheinlichkeitsmaßes nummerieren wir die Kugeln mit  $1, \dots, N$ , so dass alle Kugeln der Farbe  $a \in E$  eine Nummer aus  $F_a \subset \{1, \dots, N\}$  tragen. Würden wir die Nummern notieren, so wäre

$$\bar{\Omega} = \{1, \dots, N\}^n \text{ und } \bar{\mathcal{F}} = \mathcal{P}(\bar{\Omega})$$

und wir könnten die Gleichverteilung  $\bar{\mathbb{P}} = U(\bar{\Omega})$  als WMaß für einen einzelnen Zug verwenden. Für den Übergang zu  $\Omega$  konstruieren wir Zufallsvariablen. Die Farbe im  $i$ -ten Zug wird beschrieben durch

$$X_i: \begin{cases} \bar{\Omega} & \rightarrow E \\ \bar{\omega} = (\bar{\omega}_1, \dots, \bar{\omega}_n) & \mapsto a \text{ falls } \bar{\omega}_i \in F_a \end{cases}$$

Der Zufallsvektor

$$X = (X_1, \dots, X_n): \bar{\Omega} \rightarrow \Omega$$

beschreibt dann die Abfolge der Farben. Für jedes  $\omega \in \Omega$  gilt dann

$$\{X = \omega\} = F_{\omega_1} \times \dots \times F_{\omega_n} = \bigtimes_{i=1}^n F_{\omega_i}$$

und damit

$$\mathbb{P}(\{\omega\}) = \bar{\mathbb{P}}(X^{-1}(\{\omega\})) = \mathbb{P}(X = \omega) = \frac{|F_{\omega_1}| \cdots |F_{\omega_n}|}{|\bar{\Omega}|} = \prod_{i=1}^n \frac{|F_{\omega_i}|}{N} =: \prod_{i=1}^n \rho(\omega_i)$$

Zähldichten, die sich als Produkt von Zähldichten schreiben lassen, werden auch als **Produkt-dichten** bezeichnet (↗ Abschnitt 3.2).

Sehr oft interessiert uns bei einem Urnenexperiment nicht die Reihenfolge der gezogenen Farben, sondern nur die Anzahl der Kugeln in Farbe  $a \in E$  nach  $n$  Zügen. Dies entspricht

$$\hat{\Omega} = \left\{ k = (k_a)_{a \in E} \in \mathbb{N}_0^{|E|} : \sum_{a \in E} k_a = n \right\} \text{ und } \hat{\mathcal{F}} = \mathcal{P}(\hat{\Omega})$$

Den Übergang  $\Omega \rightarrow \hat{\Omega}$  beschreiben wir durch die Zufallsvariablen

$$Y_a(\omega): \begin{cases} \Omega & \rightarrow \mathbb{N}_0 \\ \omega = (\omega_1, \dots, \omega_n) & \mapsto \sum_{a \in E} \mathbb{1}_{\{a\}}(\omega_i) \end{cases}$$

und

$$Y = (Y_a)_{a \in E}: \Omega \rightarrow \hat{\Omega}$$

Wir erhalten

$$\begin{aligned}\mathbb{P}(Y = k) &= \mathbb{P}(Y_a = k_a : a \in E) \\ &= \sum_{\omega \in \Omega : Y(\omega) = k} \prod_{i=1}^n \rho(\omega_i) = \sum_{\omega \in \Omega : Y(\omega) = k} \prod_{a \in E} \rho(a)^{k_a} = \binom{n}{(k_a)_{a \in E}} \prod_{a \in E} \rho(a)^{k_a}\end{aligned}$$

wobei

$$\binom{n}{(k_1, \dots, k_\ell)} := \begin{cases} \frac{n!}{k_1! \cdots k_\ell!} & \text{falls } \sum_{i=1}^{\ell} k_i = n \\ 0 & \text{sonst} \end{cases}$$

der Multinomialkoeffizient ist, welcher die Anzahl der Möglichkeiten beschreibt  $n$  Objekte in  $\ell$  Gruppen aufzuteilen, sodass die Gruppe  $i$  gerade  $k_i$  Objekte enthält.

### Definition 2.1

Sei  $\ell \geq 2$ ,  $p = (p_1, \dots, p_\ell)$  eine Zähldichte und  $n \in \mathbb{N}$ , dann heißt die Verteilung auf

$$\left\{ k = (k_i)_{i=1, \dots, \ell} \in \mathbb{N}_0^\ell : \sum_{i=1}^{\ell} k_i = n \right\}$$

mit Zähldichte

$$m((k_1, \dots, k_\ell)) = \binom{n}{k_1, \dots, k_\ell} \prod_{i=1}^{\ell} p_i^{k_i}$$

**Multinomialverteilung** mit Parametern  $n$  und  $p$ . Wir schreiben auch  $\text{Multi}(n, p)$ .

### Beispiel 2.2

Eine Urne enthalte nur schwarze (“1”) und weiße (“0”) Kugeln ( $E = \{0, 1\}$ ) und es sei  $\rho(1) = p$  gerade die Proportion der schwarzen Kugeln (Wahrscheinlichkeit bei einem Zug schwarz zu ziehen). Dann ist die Wahrscheinlichkeit in  $n$  Zügen  $k$ -mal schwarz zu ziehen

$$\binom{n}{k} \prod_{i=0,1} \rho(i)^{k_i} = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

Ein solches (wiederholtes) Experiment mit nur zwei möglichen Ergebnissen und feste Wahrscheinlichkeit  $p \in [0, 1]$  nennen wir auch (wiederholtes) **BERNOULLI-Experiment**.

**Definition 2.3**

Sei  $p \in [0, 1]$  und  $n \in \mathbb{N}$ , dann heißt die Verteilung mit Zähldichte

$$\rho(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k \in \{0, 1, \dots, n\}$$

**Binomialverteilung** auf  $\{0, \dots, n\}$  mit Parameter  $p$  (Erfolgswahrscheinlichkeit). Wir schreiben auch  $\text{Bin}(n, p)$ . Im Fall  $n = 1$  nennen wir die Verteilung mit Zähldichte

$$\rho(0) = 1 - p \quad \rho(1) = p$$

auch **Bernoulliverteilung** mit Parameter  $p$  und schreiben  $\text{Bernoulli}(p)$ .

## 2.2.2 Urnenmodell ohne Zurücklegen: Hypergeometrische Verteilung

Gegeben sei eine Urne mit  $N$  Kugeln verschiedener Farben aus  $E$  mit  $|E| \geq 2$ . Es werden  $n \leq N$  Stichproben entnommen, wobei die gezogenen Kugeln nicht in die Urne zurückgelegt werden.

**Beispiel 2.4**

Eine Urne enthalte  $S$  schwarze ("1") und  $W$  weiße ("0") Kugeln, d.h.  $E = \{0, 1\}$  und  $S + W = N$ . Dann ist die Wahrscheinlichkeit in  $n$  Zügen ohne Zurücklegen gerade  $s$  schwarze und  $w$  weiße Kugeln zu ziehen

$$\rho(s) = \frac{\binom{W}{w} \cdot \binom{S}{s}}{\binom{N}{n}} \quad 0 \leq s \leq S, 0 \leq w \leq W, s + w = n, S + W = N$$

**Definition 2.5**

Seien  $N \in \mathbb{N}, W \leq N, n \leq N$ . Dann heißt die Verteilung auf  $\{0, \dots, n\}$  mit Zähldichte

$$\rho(w) = \frac{\binom{W}{w} \cdot \binom{N-W}{n-w}}{\binom{N}{n}} \quad w = \max\{0, n - N + W\}, \dots, \min\{W, n\}$$

**Hypergeometrische Verteilung** mit Parametern  $N, W$  und  $n$ . Wir schreiben auch  $\text{Hyper}(N, W, n)$ .

## 2.3 Poisson-Approximation und -Verteilung

$\text{Bin}(n, p)$  ist zwar explizit und elementar definiert, jedoch für große  $n$  mühsam auszuwerten. Für seltene Ereignisse ( $n$  groß,  $p$  klein) kann man daher eher den folgenden Satz anwenden.



**Satz 2.6 (Poisson-Approximation)**

Sei  $\lambda > 0$  und  $(p_n)_{n \in \mathbb{N}}$  eine Folge in  $[0, 1]$  mit

$$n \cdot p_n \rightarrow \lambda, \quad n \rightarrow \infty$$

Dann gilt für alle  $k \in \mathbb{N}_0$  für die Zähldichte der  $\text{Bin}(n, p_n)$ -Verteilung

$$\lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}$$

*Beweis.* Sei  $k \in \mathbb{N}_0$  fix, dann

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n^k}{k!} \frac{n(n-1) \cdots (n-k+1)}{n^k} = \frac{n^k}{k!} \cdot 1 \cdot \left(1 - \frac{1}{n} \cdots \frac{k-1}{n}\right) \xrightarrow{n \rightarrow \infty} \frac{n^k}{k!}$$

wobei  $a(l) \xrightarrow{n \rightarrow \infty} b(l) \Leftrightarrow \frac{a(l)}{b(l)} \xrightarrow{n \rightarrow \infty} 1$ . Damit gilt

$$\begin{aligned} \binom{n}{k} p_n^k (1 - p_n)^{n-k} &\xrightarrow{n \rightarrow \infty} \frac{n^k}{k!} p_n^k (1 - p_n)^{n-k} \\ &\xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} (1 - p_n)^n = \frac{\lambda^k}{k!} \left(1 - \frac{n \cdot p_n}{n}\right)^n \\ &\xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

□

Der erhaltene Grenzwert liefert die Zähldichte auf  $\mathbb{N}_0$ , denn

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

**Definition 2.7**

Sei  $\lambda > 0$ . Dann heißt das auf  $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$  definierte Wahrscheinlichkeitsmaß mit

$$\mathbb{P}(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k \in \mathbb{N}_0$$

**Poissonverteilung** mit Parameter  $\lambda$ . Wir schreiben auch  $\text{Poisson}(\lambda)$ .

Die Poisson-Verteilung ist ein natürliches Modell für die Anzahl von zufälligen, seltenen Ereignissen (z.B. Tore im Fußballspiel, Schadensfälle einer Versicherung)

# Kapitel 3

## BEDINGTE WAHRSCHEINLICHKEITEN UND UNABHÄNGIGKEIT

### 3.1 Bedingte Wahrscheinlichkeiten

#### Beispiel 3.1

Das Würfeln mit zwei fairen, sechsseitigen Würfeln können wir mit

$$\Omega = \{(i, j) : i, j \in \{1, \dots, 6\}\} \quad \text{und} \quad \mathbb{P} = \mathbb{U}(\Omega)$$

modellieren. Da  $|\Omega| = 36$  gilt also  $\mathbb{P}(\{\omega\}) = 1/36$  für alle  $\omega \in \Omega$ . Betrachten wir das Ereignis

$$A = \{(i, j) \in \Omega : i + j = 8\}$$

dann folgt  $\mathbb{P}(A) = \frac{5}{36}$ . Werden die beiden Würfe nacheinander ausgeführt, so kann nach dem ersten Wurf eine Neubewertung der Wahrscheinlichkeit von  $A$  erfolgen. Ist beispielsweise

$$B = \{(i, j) \in \Omega, i = 4\}$$

eingetreten, so kann die Summe 8 nur durch eine weitere 4 realisiert werden, also mit Wahrscheinlichkeit

$$\frac{1}{6} = \frac{|A \cap B|}{|B|}.$$

Das Eintreten von  $B$  führt also dazu, dass das Wahrscheinlichkeitsmaß  $\mathbb{P}$  durch ein neues Wahrscheinlichkeitsmaß  $\mathbb{P}_B$  ersetzt werden muss. Hierbei sollte gelten:

$$\text{Renormierung:} \quad \mathbb{P}_B = 1 \quad (\text{R})$$

$$\text{Proportionalität:} \quad \begin{aligned} &\text{Für alle } A \subseteq \mathcal{F} \text{ mit } A \subseteq B \text{ gilt } \mathbb{P}_B(A) = c_B \cdot \mathbb{P}(A) \\ &\text{mit einer Konstante } c_B \end{aligned} \quad (\text{P})$$

#### Lemma 3.2

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  Wahrscheinlichkeitsraum und  $B \in \mathcal{F}$  mit  $\mathbb{P}(B) > 0$ . Dann gibt es genau ein Wahrscheinlichkeitsmaß  $\mathbb{P}_B$  auf  $(\Omega, \mathcal{F})$  mit den Eigenschaften (R) und (P). Dieses ist gegeben durch

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad \text{für alle } A \in \mathcal{F}$$

*Beweis.* Offenbar erfüllt  $\mathbb{P}_B$  wie definiert (R) und (P). Umgekehrt erfülle  $\mathbb{P}_B$  die Eigenschaften

(R) und (P). Dann folgt für  $A \in \mathcal{F}$

$$\mathbb{P}_B(A) = \mathbb{P}_B(A \cap B) + \underbrace{\mathbb{P}_B(A \setminus B)}_{=0, \text{ wegen (R)}} \stackrel{(P)}{=} c_B \cdot \mathbb{P}(A \cap B).$$

Für  $A = B$  folgt zudem aus (R)

$$1 = \mathbb{P}_B(B) = c_B \cdot \mathbb{P}(B)$$

also  $c_B = \mathbb{P}(B)^{-1}$ . □

### Definition 3.3

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  Wahrscheinlichkeitsraum und  $B \in \mathcal{F}$  mit  $\mathbb{P}(B) > 0$ . Dann heißt

$$\mathbb{P}(A | B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \text{ mit } A \in \mathcal{F}$$

die **bedingte Wahrscheinlichkeit** von  $A$  gegeben  $B$ . Falls  $\mathbb{P}(B) = 0$ , setze  $\mathbb{P}(A | B) = 0$  für alle  $A \in \mathcal{F}$ .

### Beispiel 3.4

In der Situation Beispiel 3.1 gilt  $A \cap B = \{(4, 4)\}$  und damit

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$$

Aus Definition 3.3 ergibt sich das folgende Lemma.

### Lemma 3.5 (Multiplikationsformel)

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $A_1, \dots, A_n \in \mathcal{F}$ . Dann gilt

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \dots \mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1})$$

*Beweis.* Ist  $\mathbb{P}(A_1 \cap \dots \cap A_n) = 0$ , so gilt auch  $\mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1}) = 0$ . Andernfalls sind alle Faktoren der rechten Seite ungleich Null und

$$\begin{aligned} & \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \dots \mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1}) \\ &= \mathbb{P}(A_1) \cdot \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} \dots \frac{\mathbb{P}(A_1 \cap \dots \cap A_n)}{\mathbb{P}(A_1 \cap \dots \cap A_{n-1})} \\ &= \mathbb{P}(A_1 \cap \dots \cap A_n) \end{aligned} \quad \square$$

Stehen die  $A_i$  in Lemma 3.5 in einer (zeitlichen) Abfolge, so liefert Formel einen Hinweis wie Wahrscheinlichkeitsmaße für **Stufenexperimente** konstruiert werden können. Ein Stufenexperiment aus  $n$  nacheinander ausgeführten Telexperimenten lässt sich als **Baumdiagramm** darstellen.

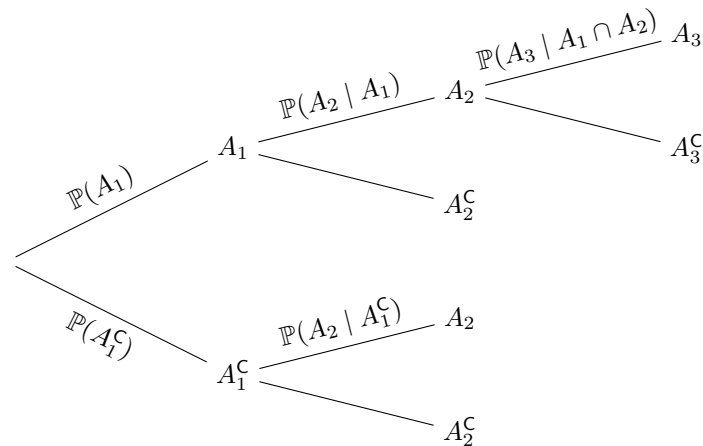


Abbildung 3.1: Darstellung eines Stufenexperiments

**Satz 3.6 (Wahrscheinlichkeitsmaß eines Stufenexperiments)**

Gegeben seien  $n$  Ergebnisräume  $\Omega_i = \{\omega_i(1), \dots, \omega_i(k)\}, k \in \mathbb{N} \cup \{\infty\}$  und es sei  $\Omega = \Omega_1 \times \dots \times \Omega_n$  der zugehörige Produktraum. Weiter seien  $\mathcal{F}_i$   $\sigma$ -Algebren auf  $\Omega_i$  und  $\mathcal{F} = \bigotimes_{i=1}^n \mathcal{F}_i$  die Produkt- $\sigma$ -Algebra auf  $\Omega$ . Setze  $\omega = (\omega_1, \dots, \omega_n)$  und

$$[\omega_1, \dots, \omega_m] := \{\omega_1\} \times \dots \times \{\omega_m\} \times \Omega_{m+1} \times \dots \times \Omega_n \quad (m \leq n)$$

und

$$\mathbb{P}(\{\omega_m\} \mid [\omega_1, \dots, \omega_{m-1}])$$

für die Wahrscheinlichkeit in der  $m$ -ten Stufe des Experiments  $\omega_m$  zu beobachten, falls in den vorausgehenden Stufen  $\omega_1, \dots, \omega_{m-1}$  beobachtet wurden. Dann definiert

$$\mathbb{P}(\{\omega\}) := \mathbb{P}(\{\omega_1\}) \prod_{m=2}^n \mathbb{P}(\{\omega_m\} \mid [\omega_1, \dots, \omega_{m-1}])$$

ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{F}, \mathbb{P})$ .

*Beweis.* Nachrechnen! □

**Beispiel 3.7 (POLYA-Urne)**

Gegeben sei eine Urne mit  $s$  schwarzen und  $w$  weißen Kugeln. Bei jedem Zug wird die gezogene Kugel zusammen mit  $c \in \mathbb{N}_0 \cup \{-1\}$  weiteren Kugeln derselben Farbe zurückgelegt.

- $c = 0$ : Urnenmodell mit Zurücklegen
- $c = -1$ : Urnenmodell ohne Zurücklegen

Beide haben wir schon in Kapitel 2.2 gesehen. Sei deshalb  $c \in \mathbb{N}$  (Bsp. Modell für zwei konkurrierende Populationen). Ziehen wir  $n$ -mal, so haben wir ein  $n$ -Stufenexperiment mit

$$\Omega = \{0, 1\}^n \quad \text{mit} \quad 0 = \text{“weiß”}, 1 = \text{“schwarz”} \quad (\Omega_i = \{0, 1\})$$

Zudem gelten im ersten Schritt

$$\mathbb{P}(\{0\}) = \frac{w}{s+w} \text{ und } \mathbb{P}(\{1\}) = \frac{s}{s+w}$$

sowie

$$\mathbb{P}(\{\omega_m\} \mid [\omega_1, \dots, \omega_{m-1}]) = \begin{cases} \frac{w+c(m-1-\sum_{i=1}^{m-1} \omega_i)}{s+w+c(m-1)} & \omega_m = 0 \\ \frac{s+c \sum_{i=1}^{m-1} \omega_i}{s+w+c(m-1)} & \omega_m = 1 \end{cases}$$

Mit Satz 3.6 folgt als Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{P}(\Omega))$

$$\begin{aligned} \mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) &= \mathbb{P}(\{\omega_1\}) \prod_{m=2}^n \mathbb{P}(\{\omega_m\} \mid [\omega_1, \dots, \omega_{m-1}]) \\ &= \frac{\prod_{i=0}^{l-1} (s+c \cdot i) \prod_{j=0}^{n-l-1} (w+c \cdot j)}{\prod_{i=0}^n (s+w+c \cdot i)} \end{aligned}$$

mit  $l = \sum_{i=1}^n \omega_i$ . Definiere wir nun die Zufallsvariable

$$S_n: \Omega \rightarrow \mathbb{N}_0 \text{ mit } (\omega_1, \dots, \omega_n) \mapsto \sum_{i=1}^n \omega_i$$

welche die Anzahl der gezogenen schwarzen Kugeln modelliert, so folgt

$$\mathbb{P}(S_n = l) = \binom{n}{l} \frac{\prod_{i=0}^{l-1} (s+c \cdot i) \prod_{j=0}^{n-l-1} (w+c \cdot j)}{\prod_{i=0}^n (s+w+c \cdot i)}$$

Mittels  $a := s/c$  und  $b := w/c$  folgt

$$\mathbb{P}(S_n = l) = \binom{n}{l} \frac{\prod_{i=0}^{l-1} (-a-i) \prod_{j=0}^{n-l-1} (-b-j)}{\prod_{i=0}^n (-a-b-i)} = \frac{\binom{-a}{l} \binom{-b}{n-l}}{\binom{-a-b}{n}} \text{ mit } l \in \{0, \dots, n\}$$

Dies ist die **Polya-Verteilung** auf  $\{0, \dots, n\}$  ( $n \in \mathbb{N}$ ) mit Parametern  $a, b > 0$ .

### Beispiel 3.8

Ein Student beantwortet eine Multiple-Choice-Frage mit vier Antwortmöglichkeiten, eine davon ist richtig. Er kennt die richtige Antwort mit Wahrscheinlichkeit  $2/3$ . Wenn er die richtige Antwort kennt, so wählt er diese aus. Andernfalls wählt er zufällig (gleichverteilt) eine Antwort. Betrachten wir

$$W = \{\text{richtige Antwort gewusst}\} \text{ und } R = \{\text{Richtige Antwort gewählt}\}$$

Dann gilt

$$\mathbb{P}(W) = \frac{2}{3} \quad \mathbb{P}(R \mid W) = 1 \quad \mathbb{P}(R \mid W^c) = \frac{1}{4}$$

Angenommen der Student gibt die richtige Antwort. Mit welcher Wahrscheinlichkeit hat er diese gewusst?  $\rightarrow \mathbb{P}(W \mid R) = ?$

**Satz 3.9**

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $\Omega = \bigcup_{i \in I} B_i$  eine höchstens abzählbare Zerlegung in paarweise disjunkte Ereignisse  $B_i \in \mathcal{F}$ .

(1) *Satz von der totalen Wahrscheinlichkeit:* Für alle  $A \in \mathcal{F}$  gilt

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i) \quad (3.1)$$

(2) *Satz von Bayes:* Für alle  $A \in \mathcal{F}$  mit  $\mathbb{P}(A) > 0$  und alle  $k \in I$  gilt

$$\mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(A \mid B_k) \cdot \mathbb{P}(B_k)}{\sum_{i \in I} \mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i)} \quad (3.2)$$

*Beweis.* Es gilt:

$$\sum_{i \in I} \mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i) = \sum_{i \in I} \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(B_i)} \cdot \mathbb{P}(B_i) = \sum_{i \in I} \mathbb{P}(A \cap B_i) \stackrel{\sigma\text{-Add.}}{=} \mathbb{P}(A)$$

und

$$\mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(A \cap B_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_k) \cdot \mathbb{P}(B_k)}{\mathbb{P}(A)}$$

also folgt (b) aus (a). □

**Beispiel 3.10**

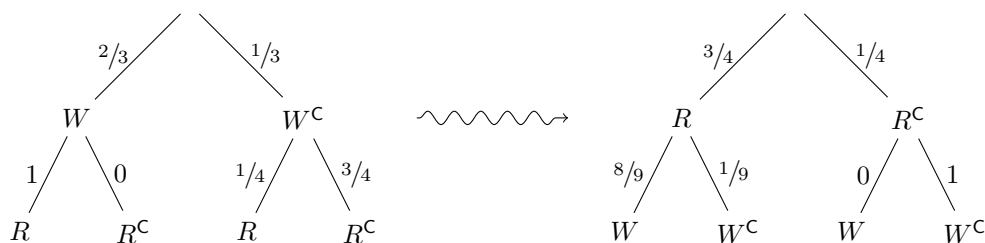
In der Situation von Beispiel 3.8 folgt mit (3.1)

$$\begin{aligned} \mathbb{P}(R) &= \mathbb{P}(R \mid W) \cdot \mathbb{P}(W) + \mathbb{P}(R \mid W^c) \cdot \mathbb{P}(W^c) \\ &= 1 \cdot \frac{2}{3} + \frac{1}{4} \cdot \frac{1}{3} = \frac{3}{4} \end{aligned}$$

und mit (3.2)

$$\mathbb{P}(W \mid R) = \frac{\mathbb{P}(R \mid W) \cdot \mathbb{P}(W)}{\mathbb{P}(R)} = \frac{1 \cdot \frac{2}{3}}{\frac{3}{4}} = \frac{8}{9}$$

für die gesuchte Wahrscheinlichkeit.



**Abbildung 3.2:** Bedingte Wahrscheinlichkeit im Baumdiagramm

## 3.2 Stochastische Unabhängigkeit

In vielen Fällen besagt die Intuition über verschiedene Zufallsexperimente / Ereignisse, dass diese sich *nicht* gegenseitig beeinflussen. Für solche  $A, B \in \mathcal{F}$  mit  $\mathbb{P}(A) > 0$  und  $\mathbb{P}(B) > 0$  sollte gelten

$$\mathbb{P}(A \mid B) = \mathbb{P}(A) \quad \text{und} \quad \mathbb{P}(B \mid A) = \mathbb{P}(B).$$

### Definition 3.11

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  Wahrscheinlichkeitsraum. Zwei Ereignisse  $A, B \in \mathcal{F}$  heißt **(stochastisch) unabhängig** bezüglich  $\mathbb{P}$ , falls

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Wir schreiben auch  $A \perp\!\!\!\perp B$ .

### Beispiel 3.12

Wir betrachten wieder das Würfeln mit 2 fairen, sechsseitigen Würfeln.

$$\Omega = \{(i, j) \mid i, j \in \{1, \dots, n\}\} \quad \mathcal{F} = \mathcal{P}(\Omega) \quad \mathbb{P} = \mathbb{U}(\Omega)$$

Betrachte

$$A := \{(i, j) \in \Omega : i \text{ gerade}\}$$

$$B := \{(i, j) \in \Omega : j \leq 2\}$$

In diesem Fall, erwarten wir intuitiv die Unabhängigkeit von  $A$  und  $B$ . In der Tat ist  $\mathbb{P}(A) = 1/2$ ,  $\mathbb{P}(B) = 1/3$  und  $\mathbb{P}(A \cap B) = 1/6$ , womit  $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$  erfüllt ist. Betrachte nun

$$C := \{(i, j) \in \Omega : i + j = 7\}$$

$$D := \{(i, j) \in \Omega : i = 6\}$$

Dann gilt  $\mathbb{P}(C) = 1/6$  und  $\mathbb{P}(D) = 1/6$ . Wegen  $C \cap D = \{(6, 1)\}$  folgt

$$\mathbb{P}(C \cap D) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(C) \cdot \mathbb{P}(D)$$

$C$  und  $D$  sind also *stochastisch* unabhängig, obwohl eine kausale Abhängigkeit vorliegt.

### Definition 3.13

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $I \neq \emptyset$  eine endliche Indexmenge. Dann heißt die Familie  $(A_i)_{i \in I}$  von Ereignissen in  $\mathcal{F}$  **unabhängig** bezüglich  $\mathbb{P}$ , falls für alle  $\emptyset \neq J \subseteq I$  gilt

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

Offensichtlich impliziert die Unabhängigkeit einer Familie die paarweise Unabhängigkeit je zweier Familienmitglieder nach Definition 3.11. Umgekehrt gilt dies nicht.

**Beispiel 3.14 (Abhängigkeit trotz paarweiser Unabhängigkeit)**

Wir betrachten ein zweifaches Bernoulliexperiment mit Erfolgswahrscheinlichkeit  $1/2$ , d.h.

$$\Omega = \{0, 1\}^2 \quad \mathcal{F} = \mathcal{P}(\Omega) \quad \mathbb{P} = \mathbb{U}(\Omega)$$

sowie

$$\begin{aligned} A &= \{1\} \times \{0, 1\} && \text{(Münzwurf: erster Wurf ist Zahl)} \\ B &= \{0, 1\} \times \{1\} && \text{(Münzwurf: zweiter Wurf ist Zahl)} \\ C &= \{(0, 0), (1, 1)\} && \text{(beide Würfe haben selbes Ergebnis)} \end{aligned}$$

Dann gelten  $\mathbb{P}(A) = 1/2 = \mathbb{P}(B) = \mathbb{P}(C)$  und

$$\begin{aligned} \mathbb{P}(A \cap B) &= \mathbb{P}(\{(1, 1)\}) = 1/4 = \mathbb{P}(A) \cdot \mathbb{P}(B) \\ \mathbb{P}(A \cap C) &= \mathbb{P}(\{(1, 1)\}) = 1/4 = \mathbb{P}(A) \cdot \mathbb{P}(C) \\ \mathbb{P}(B \cap C) &= \mathbb{P}(\{(1, 1)\}) = 1/4 = \mathbb{P}(B) \cdot \mathbb{P}(C) \end{aligned}$$

Daraus folgt also paarweise Unabhängigkeit. Jedoch ist

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

und  $A, B, C$  sind *nicht* stochastisch unabhängig.

**Definition 3.15 (Unabhängige  $\sigma$ -Algebren)**

Seien  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum,  $I \neq \emptyset$  eine Indexmenge und  $(E_i, \mathcal{E}_i)$  Messräume.

- (1) Die Familie  $\mathcal{F}_i \subset \mathcal{F} (i \in I)$ , heißen **unabhängig**, wenn für die  $\emptyset \neq J \subseteq I$  mit  $|J| < \infty$  gilt

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i) \quad \text{für beliebige } A_i \in \mathcal{F}_i (i \in J)$$

- (2) Die Zufallsvariable  $X_i: (\Omega, \mathcal{F}) \rightarrow (E_i, \mathcal{E}_i) (i \in I)$ , heißen **unabhängig**, wenn die  $\sigma$ -Algebren

$$\sigma(X_i) = X^{-1}(\mathcal{E}_i) = \{\{X_i \in F\} : F \in \mathcal{E}_i\} \quad (i \in I)$$

unabhängig sind.

**Lemma 3.16 (Zusammenhang der Definitionen)**

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum,  $I \neq \emptyset$ ,  $A_i \in \mathcal{F}_i$  für alle  $i \in I$ . Die folgenden Aussagen sind äquivalent:

- (1) Die Ereignisse  $A_i (i \in I)$  sind unabhängig.
- (2) Die  $\sigma$ -Algebren  $\sigma(A_i) (i \in I)$  sind unabhängig.
- (3) Die Zufallsvariablen  $\mathbb{1}_{A_i} (i \in I)$  sind unabhängig.

*Beweis.* Da die Unabhängigkeit über endliche Teilmengen definiert ist, können wir oBdA  $I = \{1, \dots, n\}$  annehmen.



- Da  $\sigma(\mathbb{1}_{A_i}) = \sigma(A_i)$  folgt die Äquivalenz von (2) und (3) direkt aus Definition 3.15.
- Zudem ist (2)  $\rightarrow$  (1) klar.
- Für (1)  $\rightarrow$  (2) genügt es zu zeigen, dass

$$A_1, \dots, A_n \text{ unabhängig} \Rightarrow B_1, \dots, B_n \text{ unabhängig mit } B_i \in \{\emptyset, A_i, A_i^c, \Omega\}$$

Rekursiv folgt dies bereits aus

$$A_1, \dots, A_n \text{ unabhängig} \Rightarrow B_1, A_2, \dots, A_n \text{ unabhängig mit } B_1 \in \{\emptyset, A_1, A_1^c, \Omega\}.$$

Für  $B_1 \in \{\emptyset, A_1, \Omega\}$  ist dies klar.

Sei also  $B_1 = A_1^c$  und  $J \subseteq I$ ,  $J \neq \emptyset$ . Falls  $1 \notin J$ , ist nichts zu zeigen. Sei  $1 \in J$ , dann gilt mit  $A = \bigcap_{i \in J, i \neq 1} A_i$  sicherlich

$$\begin{aligned} \mathbb{P}(A_1^c \cap A) &= \mathbb{P}(A \setminus (A_1 \cap A)) \\ &= \mathbb{P}(A) - \mathbb{P}(A_1 \cap A) \\ &= \prod_{i \in J \setminus \{1\}} \mathbb{P}(A_i) - \prod_{i \in J} \mathbb{P}(A_i) \\ &= (1 - \mathbb{P}(A_1)) \prod_{i \in J \setminus \{1\}} \mathbb{P}(A_i) \\ &= \mathbb{P}(A_1^c) \prod_{i \in J \setminus \{1\}} \mathbb{P}(A_i) \end{aligned}$$

□

Insbesondere zeigt Lemma 3.16, dass wir in einer Familie unabhängiger Ereignisse beliebig viele Ereignisse durch ihr Komplement,  $\emptyset$  oder  $\Omega$  ersetzen können ohne die Unabhängigkeit zu verlieren.

### Satz 3.17

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  Wahrscheinlichkeitsraum und  $\mathcal{F}_i \subseteq \mathcal{F}$ ,  $i \in I$ , seien  $\cap$ -stabile Familien von Ereignissen. Dann gilt

$$\mathcal{F}_i (i \in I) \text{ unabhängig} \Leftrightarrow \sigma(\mathcal{F}_i) (i \in I) \text{ unabhängig.}$$

*Beweis.* oBdA sei  $I = \{1, \dots, n\}$  und  $\Omega \in \mathcal{F}_i$ ,  $i \in I$ .

( $\Leftarrow$ ) trivial, da  $\mathcal{F}_i \subseteq \sigma(\mathcal{F}_i)$  und das Weglassen von Mengen erlaubt ist.

( $\Rightarrow$ ) zeigen wir rekursiv

- (i) Wähle  $F_i \in \mathcal{F}_i$ ,  $i = 2, \dots, n$  und definiere für  $F \in \sigma(\mathcal{F}_1)$  die endlichen Maße

$$\mu(F) = \mathbb{P}(F \cap F_2 \cap \dots \cap F_n) \text{ und } \nu(F) = \mathbb{P}(F) \mathbb{P}(F_2) \dots \mathbb{P}(F_n)$$

- (ii) Da die Familien  $\mathcal{F}_i$  unabhängig sind, gilt  $\mu|_{\mathcal{F}_1} = \nu|_{\mathcal{F}_1}$ . Nach dem Eindeutigkeitssatz für Maße (Satz 1.9) folgt  $\mu|_{\sigma(\mathcal{F}_1)} = \nu|_{\sigma(\mathcal{F}_1)}$  also

$$\mathbb{P}(F \cap F_2 \cap \dots \cap F_n) = \mathbb{P}(F) \mathbb{P}(F_2) \dots \mathbb{P}(F_n)$$

für alle  $F \in \sigma(\mathcal{F}_i)$  und  $F_i \in \mathcal{F}_i, i = 1, \dots, n$ . Da  $\Omega \in \mathcal{F}_i$  für alle  $i$  gilt die erhaltene Produktformel auf für alle Teilmengen  $J \subseteq I$ .

Also sind  $\sigma(\mathcal{F}_1), \mathcal{F}_2, \dots, \mathcal{F}_n$  unabhängig.

(iii) Wiederholtes Anwenden von (i) und (ii) liefert den Satz.  $\square$

Mit Satz 3.17 folgen:

### Korollar 3.18

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $\mathcal{F}_{i,j} \subseteq \mathcal{F}$  mit  $1 \leq i \leq n, 1 \leq j \leq m(i)$  unabhängige,  $\cap$ -stabile Familien. Dann sind auch

$$\mathcal{G}_i = \sigma(\mathcal{F}_{i,1}, \dots, \mathcal{F}_{i,m(i)}), \quad 1 \leq i \leq n$$

unabhängig.

*Beweis.* oBdA sei  $\Omega \in \mathcal{F}_{i,j} \forall i, j$ , da  $\Omega$  die Unabhängigkeit nach Lemma 3.16 nicht zerstört. Dann sind die Familien

$$\mathcal{F}_i^\cap := \left\{ F_{i,1} \cap \dots \cap F_{i,m(i)} : F_{i,j} \in \mathcal{F}_{i,j}, 1 \leq j \leq m(i) \right\} \quad (1 \leq i \leq n)$$

$\cap$ -stabil, unabhängig und es gilt  $\mathcal{F}_{i,1}, \dots, \mathcal{F}_{i,m(i)} \subseteq \mathcal{F}_i^\cap$  (vgl. Hausaufgabe). Nach Satz 3.17 sind auch  $\sigma(\mathcal{F}_i^\cap)$  unabhängig. Damit folgt die Behauptung, da aus

$$\mathcal{F}_{i,1}, \dots, \mathcal{F}_{i,m(i)} \subseteq \mathcal{F}_i^\cap \subseteq \sigma(\mathcal{F}_{i,1}, \dots, \mathcal{F}_{i,m(i)}) = \mathcal{G}_i$$

unter Anwendung des Erzeugenden-Operators  $\mathcal{G}_i = \sigma(\mathcal{F}_{i,1}, \dots, \mathcal{F}_{i,m(i)}) \subseteq \sigma(\mathcal{F}_i^\cap) \subseteq \mathcal{G}_i$  folgt, d.h.  $\sigma(\mathcal{F}_i^\cap) = \mathcal{G}_i$ .  $\square$

### Korollar 3.19

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $X_{i,j}: \Omega \rightarrow E$  ( $1 \leq i \leq n, 1 \leq j \leq m(i)$ ) unabhängige Zufallsvariablen. Zudem seien  $f_i: E^{m(i)} \rightarrow \mathbb{R}$  messbar. Dann sind auch die Zufallsvariablen  $f_i(X_{i,1}, \dots, X_{i,m(i)})$  mit  $1 \leq i \leq n$  unabhängig.

*Beweis.* Setze  $\mathcal{F}_{i,j} = \sigma(X_{i,j})$  und  $\mathcal{G}_i = \sigma(\mathcal{F}_{i,1}, \dots, \mathcal{F}_{i,m(i)})$ . Dann sind nach Korollar 3.18 die  $\mathcal{G}_i$  ( $i = 1, \dots, n$ ) unabhängig. Zudem ist  $Y_i := f_i(X_{i,1}, \dots, X_{i,m(i)})$   $\mathcal{G}_i$ -messbar, also  $\sigma(Y_i) \subseteq \mathcal{G}_i$ . Damit erben die  $Y_i$  die Unabhängigkeit der  $\mathcal{G}_i$ .  $\square$

### Beispiel 3.20

Seien  $X_1, \dots, X_n$  unabhängige, reelle Zufallsvariablen. Dann sind auch

$$Y_1 = X_1, Y_2 = X_2 + \dots + X_n$$

unabhängig.

**Satz 3.21 (Unabhängigkeit von Zufallsvariablen)**

Seien  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $X_1, \dots, X_n: (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$  Zufallsvariablen. Dann sind äquivalent:

- (1)  $X_1, \dots, X_n$  sind unabhängig.
- (2)  $\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i) \quad \forall A_1, \dots, A_n \in \mathcal{E}$
- (3) Die gemeinsame Verteilung der  $X_i$  entspricht dem Produktmaß der einzelnen Verteilungen

$$\mathbb{P}_{X_1, \dots, X_n} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$$

Beweis. per Ringschluss

(1  $\Rightarrow$  2) Seien  $A_1, \dots, A_n \in \mathcal{E}$  beliebig. Dann gilt per Definition

$$\begin{aligned} \mathbb{P}_{X_1, \dots, X_n}(A_1 \times \dots \times A_n) &= \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) \\ &= \mathbb{P}\left(\bigcap_{i=1}^n \{X_i \in A_i\}\right) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \in A_i) \\ &= \prod_{i=1}^n \mathbb{P}_{X_i}(A_i) \\ &= \left(\bigotimes_{i=1}^n \mathbb{P}_{X_i}\right)(A_1 \times \dots \times A_n) \end{aligned}$$

(2  $\Rightarrow$  3) Aus der obigen Gleichung sehen wir, dass (2) bereits (3) impliziert für alle Rechtecke  $A_1 \times \dots \times A_n$ . Da die Familie der Rechtecke  $\cap$ -stabil ist und  $\mathcal{E}^{\otimes n}$  erzeugt, folgt die Aussage aus dem Eindeutigkeitssatz für Maße.

(3  $\Rightarrow$  1) Es sei  $J \subseteq \{1, \dots, n\}$  und

$$A_i := \begin{cases} \text{beliebig in } \mathcal{E} & \text{falls } i \in J \\ E & \text{falls } i \notin J \end{cases}$$

Dann ist

$$\begin{aligned} \mathbb{P}(X_i \in A_i : i \in J) &= \mathbb{P}(X_i \in A_i : i = 1, \dots, n) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \in A_i) \\ &= \prod_{i \in J} \mathbb{P}(X_i \in A_i) \end{aligned}$$

□

**Beispiel 3.22**

Im Urnenmodell mit Zurücklegen hat der Vektor  $X = (X_1, \dots, X_n)$  mit  $X_i = \text{Farbe im } i\text{-ten Zug}$  als Zähldichte die Produktdichte der  $X_i$ .  $X_1, \dots, X_n$  sind also unabhängig.

### 3.2.1 Konstruktion unabhängiger Zufallsvariablen

Kapitel 1: Zu beliebiger Wahrscheinlichkeitsverteilung  $\mathbb{P}_X$  existiert ein Wahrscheinlichkeitsraum mit Zufallsvariable  $X$  auf diesem Wahrscheinlichkeitsraum, so dass  $X \sim \mathbb{P}_X$ .

- (1) Seien  $\mathbb{P}_{X_1}, \mathbb{P}_{X_2}$  Wahrscheinlichkeitsverteilungen auf  $(E, \mathcal{E})$ . Gibt es einen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, \mathbb{P})$  und Zufallsvariablen  $X_1, X_2$  unabhängig, so dass  $X_i \sim \mathbb{P}_{X_i}$ ?
- (2) Wie kann ich beliebig (unendlich) viele unabhängige Zufallsvariablen konstruieren?

Wir beginnen mit Schritt (1):

Konstruiere zwei Wahrscheinlichkeitsräume  $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$  ( $i = 1, 2$ ) und Zufallsvariablen  $X_1, X_2$  mit  $X_i \sim \mathbb{P}_{X_i}$ . Auf dem Produktraum

$$\Omega := \Omega_1 \times \Omega_2 \quad \mathcal{F} := \mathcal{F}_1 \otimes \mathcal{F}_2 \quad \mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$$

definieren wir

$$X'_1: \Omega_1 \times \Omega_2 \rightarrow E: (\omega_1, \omega_2) \mapsto X_1(\omega_1)$$

$$X'_2: \Omega_1 \times \Omega_2 \rightarrow E: (\omega_1, \omega_2) \mapsto X_2(\omega_2)$$

Dann gilt für beliebige Ereignisse:  $F_1, F_2 \in \mathcal{E}$

$$\underbrace{\{X'_1 \in F_1\} \cap \{X'_2 \in F_2\}}_{\supseteq \Omega = \Omega_1 \times \Omega_2} = \underbrace{\{X_1 \in F_1\}}_{\supseteq \Omega_1} \times \underbrace{\{X_2 \in F_2\}}_{\supseteq \Omega_2} \in \mathcal{F}_1 \times \mathcal{F}_2$$

und damit folgt die Messbarkeit der Abbildungen  $X'_1, X'_2$ , d.h.  $X'_1, X'_2$  sind Zufallsvariablen auf  $(\Omega, \mathcal{F})$ . Zudem gilt

$$\begin{aligned} \mathbb{P}(X'_1 \in F_1, X'_2 \in F_2) &= \mathbb{P}_1 \otimes \mathbb{P}_2(\{X_1 \in F_1\} \times \{X_2 \in F_2\}) \\ &= \mathbb{P}_1(X_1 \in F_1) \mathbb{P}_2(X_2 \in F_2) \end{aligned}$$

also  $\mathbb{P}(X'_i \in F_i) = \mathbb{P}_i(X'_i \in F_i)$  sowie nach Satz 3.21  $X'_1 \perp\!\!\!\perp X'_2$ .

Wenn  $(\Omega_2, \mathcal{F}_1, \mathbb{P}_1) = (\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ , so liefert die obige Konstruktion zwei unabhängige Zufallsvariablen auf einem Wahrscheinlichkeitsraum. Andernfalls können wir auf den Produktraum ausweichen und  $X'_i$  anstelle von  $X_i$  betrachten. Die obige Konstruktion lässt sich direkt auf endlich viele Zufallsvariablen übertragen.

Betrachten wir nun den Fall (2). Wir benötigen dafür den folgenden Satz.

**Satz 3.23 (Satz von Kolmogorov)**

Sei  $I$  eine beliebige Indexmenge und  $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$  ( $i \in I$ ) Wahrscheinlichkeitsräume. Setze

$$\Omega_I := \prod_{i \in I} \Omega_i = \left\{ \omega : I \rightarrow \bigcup_{i \in I} \Omega_i, \omega_i \in \Omega_i, i \in I \right\}$$

$$\mathcal{F}_I := \sigma(\pi^{-1}(\mathcal{F}_i) : i \in I)$$

wobei  $\pi_i : \Omega_I \rightarrow \Omega_i$  mit  $\omega \mapsto \omega_i$  die Projektionsabbildung ist. Dann existiert auf  $(\Omega_I, \mathcal{F}_I)$  genau ein Wahrscheinlichkeitsmaß  $\mathbb{P}_I$ , sodass für alle  $H \subseteq I$  mit  $0 \leq |H| < \infty$  gilt

$$\pi_H(\mathbb{P}_I) = \bigotimes_{i \in H} \mathbb{P}_i,$$

wobei  $\pi_i : \Omega_I \rightarrow \Omega_H$  wiederum die Projektionsabbildung.

*Beweis.* ↗ Schilling: Maß und Integral, Satz 17.4. □

Sind auf den Wahrscheinlichkeitsräumen  $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$  ( $i \in I$ ), nun Zufallsvariablen  $X_i : \Omega_i \rightarrow E$  gegeben, so definieren wir wie im Satz von Kolmogorov (Satz 3.23)

$$(\Omega, \mathcal{F}, \mathbb{P}) := \left( \Omega_I, \mathcal{F}_I, \mathbb{P}_I = \bigotimes_{i \in I} \mathbb{P}_i \right) \text{ mit } \omega = (\omega_i)_{i \in I}$$

und wie im endlichen Fall

$$X'_i : \Omega \rightarrow E \text{ mit } X'_i(\omega) = X_i(\omega_i)$$

definieren. Da die Unabhängigkeit der Zufallsvariablen über endliche Teilfamilien definiert ist, folgt diese wie im endlichen Fall.

### 3.3 Faltungen

Seien  $X, Y$  zwei reelle unabhängige Zufallsvariablen mit  $X \sim \mathbb{P}_X$  und  $Y \sim \mathbb{P}_Y$ . Dann hat  $(X, Y)$  die Verteilung  $\mathbb{P}_X \otimes \mathbb{P}_Y$  auf  $\mathbb{R}^2$ . Andererseits ist auch  $X + Y$  eine reelle Zufallsvariable, denn  $X + Y = A(X, Y)$  für  $A : \mathbb{R}^2 \rightarrow \mathbb{R}$  mit  $(x, y) \mapsto x + y$ .  $A$  ist stetig, also messbar. Die Verteilung von  $X + Y$  ist dann  $(\mathbb{P}_X \otimes \mathbb{P}_Y) \circ A^{-1}$ .

**Definition 3.24**

Seien  $\mathbb{P}_1$  und  $\mathbb{P}_2$  Wahrscheinlichkeitsmaße auf  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ . Das durch

$$\mathbb{P}_1 * \mathbb{P}_2(F) := \iint \mathbb{1}_F(x + y) \mathbb{P}_1(dx) \mathbb{P}_2(dy)$$

definierte Wahrscheinlichkeitsmaß  $\mathbb{P}_1 * \mathbb{P}_2 = (\mathbb{P}_1 \otimes \mathbb{P}_2) \circ A^{-1}$  auf  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$  heißt **Faltung** von  $\mathbb{P}_1$  und  $\mathbb{P}_2$ .

**Satz 3.25**

Seien  $X, Y: \Omega \rightarrow \mathbb{R}^n$  unabhängige Zufallsvariablen mit Verteilungen  $\mathbb{P}_X$  bzw.  $\mathbb{P}_Y$ . Dann ist

$$\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y$$

die Verteilung von  $X + Y$ .

*Beweis.* siehe Herleitung Faltung oben □

Faltungen von Wahrscheinlichkeitsmaßen mit Dichten besitzen wiederum eine Dichte.

**Proposition 3.26**

Seien  $\mathbb{P}_1$  und  $\mathbb{P}_2$  Wahrscheinlichkeitsmaße auf  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

- (1) **diskreter Fall:** Sind  $\mathbb{P}_1$  und  $\mathbb{P}_2$  de facto Wahrscheinlichkeitsmaße auf  $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$  mit Zähldichten  $\rho_1$  und  $\rho_2$ , dann ist  $\mathbb{P}_1 * \mathbb{P}_2$  ein Wahrscheinlichkeitsmaß auf  $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$  mit Zähldichte

$$\rho_1 * \rho_2(k) = \sum_{\ell \in \mathbb{Z}} \rho_1(\ell) \rho_2(k - \ell) \quad (k \in \mathbb{Z})$$

- (2) **stetiger Fall:** Wenn  $\mathbb{P}_1$  und  $\mathbb{P}_2$  Dichtefunktionen  $\rho_1$  und  $\rho_2$  besitzen, dann besitzt  $\mathbb{P}_1 * \mathbb{P}_2$  die Dichtefunktion

$$\rho_1 * \rho_2(x) := \int_{\mathbb{R}} \rho_1(y) \rho_2(x - y) \, dy \quad (x \in \mathbb{R})$$

*Beweis. Diskreter Fall:* Sei  $k \in \mathbb{Z}$ .

$$(\mathbb{P}_1 \otimes \mathbb{P}_2)(A = k) = \sum_{\ell_1, \ell_2 \in \mathbb{Z}; \ell_1 + \ell_2 = k} \rho_1(\ell_1) \rho_2(\ell_2) = \sum_{\ell, k \in \mathbb{Z}} \rho_1(\ell) \rho_2(k - \ell) = \rho_1 * \rho_2(k)$$

**Stetiger Fall:** Sei  $c \in \mathbb{R}$ .

$$\begin{aligned} \mathbb{P}_1 * \mathbb{P}_2((-\infty, c]) &= (\mathbb{P}_1 \otimes \mathbb{P}_2)(A \leq c) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{(-\infty, c]}(x + y) \rho_1(x) \rho_2(y) \, dx \, dy \\ &\stackrel{y=z-x}{=} \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{(-\infty, c]}(z) \rho_1(x) \rho_2(z - x) \, dx \, dz \\ &= \int_{-\infty}^c \underbrace{\int_{\mathbb{R}} \rho_1(x) \rho_2(z - x) \, dx}_{\rho_1 * \rho_2(z)} \, dz \end{aligned}$$

□

**Beispiel 3.27**

Seien  $X \sim \text{Poisson}(\lambda)$  und  $Y \sim \text{Poisson}(\mu)$  zwei unabhängige reelle Zufallsvariablen (mit Werten

in  $\mathbb{N}_0$ ). Dann ist  $X + Y$  eine Zufallsvariable mit Werten in  $\mathbb{N}_0$  und Zähldichte

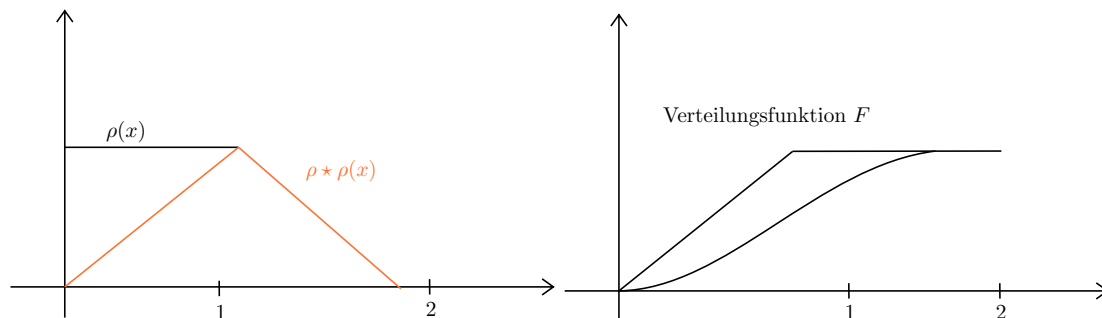
$$\begin{aligned}\mathbb{P}(X + Y = k) &= \sum_{\ell \in \mathbb{Z}} \mathbb{P}(X = \ell) \cdot \mathbb{P}(Y = k - \ell) \\ &= \sum_{\ell=0}^k \frac{\lambda^\ell}{\ell!} e^{-\lambda} \cdot \frac{\mu^{k-\ell}}{(k-\ell)!} e^{-\mu} = e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{\ell=0}^k \binom{k}{\ell} \lambda^\ell \mu^{k-\ell} \\ &= e^{-(\lambda+\mu)} \frac{1}{k!} (\lambda + \mu)^k\end{aligned}$$

sodass  $X + Y \sim \text{Poisson}(\lambda + \mu)$ , d.h. der Typ der Verteilung ist bei der Faltung erhalten geblieben. Das ist aber nicht immer der Fall.

### Beispiel 3.28

Seien  $X, Y \sim U([0, 1])$  zwei unabhängige Zufallsvariablen mit Dichten  $\rho(x) = \mathbb{1}_{[0,1]}(x)$ . Dann ist  $X + Y$  eine Zufallsvariable mit Werten in  $[0, 2]$  und Dichte

$$\begin{aligned}\rho * \rho(x) &= \int_{\mathbb{R}} \rho(y) \rho(y - x) \, dy = \int_{1 \wedge x} \mathbb{1}_{[0,1]}(y) \mathbb{1}_{[0,1]}(y - x) \, dy = \int_{1 \vee (x-1)} dy \\ &= \begin{cases} x & 0 \leq x \leq 1 \\ 2 - x & 1 \leq x \leq 2 \\ 0 & \text{sonst} \end{cases}\end{aligned}$$



# Kapitel 4

## WEITERE STANDARDMODELLE: STETIGE VERTEILUNGEN

### 4.1 Stetige Gleichverteilung

#### Erinnerung

$\Omega \subset \mathbb{R}^n$  Borel-messbar mit Lebesgue-Volumen  $0 < \lambda(\Omega) < \infty$ . Wahrscheinlichkeitsmaß ist  $(\Omega, \mathcal{B}(\Omega))$  mit Dichte

$$\rho(x) = \frac{1}{\lambda(\Omega)}$$

heißt stetige Gleichverteilung auf  $\Omega$ , notiert mit  $U(\Omega)$ .

Für alle  $A \in \mathcal{B}(\Omega)$  gilt:

$$\mathbb{P}(A) = \int_A \rho(x) \, dx = \frac{\lambda(A)}{\lambda(\Omega)}.$$

Meist verwenden wir  $U([a, b])$  für  $a < b$  (Gleichverteilung auf Intervall) mit

$$\rho(x) = \frac{1}{(b-a)} \text{ mit } a \leq x \leq b$$

und Verteilungsfunktion

$$F(x) = \begin{cases} 0 & x < a \\ \int_a^x \frac{1}{b-a} \, dy = \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b. \end{cases}$$

### 4.2 Wartezeitverteilungen

#### 4.2.1 Negative Binomialverteilung

Wir wiederholen ein Bernoulliexperiment mit Erfolgswahrscheinlichkeit  $p \in [0, 1]$  unendlich oft. Gesucht ist die Anzahl der Misserfolge bis zum  $r$ -ten Erfolg ( $r \in \mathbb{N}$ ). Ein passender Ergebnisraum ist  $\Omega = \mathbb{N}_0$ . Für Modellierung ist es jedoch leichter in jedem Versuch erfolgt ("1") oder Misserfolg ("0") festzuhalten und mit dem unendlichen Produktmaß des Bernoullimaßes auf  $\{0, 1\}^{\mathbb{N}}$  zu arbeiten.

Als Zufallsvariable wählen wir

$$X_r: \{0, 1\}^{\mathbb{N}} \rightarrow \Omega$$



welche die Anzahl der Misserfolge bis zum  $r$ -ten Erfolg darstellt. Setze

$$X_r(\omega) = \min \left\{ k : \sum_{i=1}^k \omega_i = r \right\} - r$$

Dann ist

$$\mathbb{P}(X_r = k) = \sum_{\omega \in \{0,1\}^{\mathbb{N}}, X_r(\omega)=k} \prod_{i=1}^{\infty} \rho(\omega_i) \quad (\star)$$

mit  $\rho(0) = 1 - p$  und  $\rho(1) = p$  (Zähldichte der Binomialverteilung), also

$$\mathbb{P}(X_r = k) = \binom{r+k-1}{k} (1-p)^k p^r \quad (k \in \mathbb{N})$$

**Anmerkung.** Alternativ kann Gleichung  $(\star)$  auch durch

$$\mathbb{P}(X_r = k) = \sum_{\omega \in \{0,1\}^{r+k}, X_r(\omega)=k} \prod_{i=1}^{r+k} \rho(\omega_i)$$

ersetzt werden.

#### Definition 4.1 (negative Binomialverteilung)

Sei  $p \in [0, 1]$  und  $r \in \mathbb{N}$ , dann heißt die Verteilung auf  $\mathbb{N}_0$  mit Zähldichte

$$\rho(k) = \binom{r+k-1}{k} (1-p)^k p^r$$

die **negative Binomialverteilung** mit Parametern  $(r, p)$ . Schreibe  $\text{negBin}(r, p)$ . Im Fall  $r = 1$  nennen wir die Verteilung mit Zähldichte

$$\rho(k) = p(1-p)^k \quad (k \in \mathbb{N}_0)$$

**geometrische Verteilung** mit Parameter  $p$ . Schreibe  $\text{Geom}(p)$ .

### 4.2.2 Exponential- und Gammaverteilung

**Ziel:** Modelliere die Wartezeit auf  $r$  Ereignisse in kontinuierlicher Zeit

**Wähle:**  $(\Omega, \mathcal{F}) = (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$

**Annahmen:**

- Jedes Ereignis geschieht zu einer zufälligen Zeit.
- Die Anzahl der Ereignisse bis zur Zeit  $t$  sei  $\text{Poisson}(\lambda t)$  verteilt.

Die zweite Annahme macht Sinn, denn

- Poissonverteilung ist Modell für Anzahl seltener Ereignisse

- Nach Beispiel 3.26 gilt  $\text{Poisson}(\lambda t) * \text{Poisson}(\lambda s) = \text{Poisson}(\lambda(t + s))$ . Die Linearität des Parameters entspricht also einer Stationaritätsvoraussetzung – Modelliert  $X \sim \text{Poisson}(\lambda t)$  die Ereignisse in  $(0, t]$  und  $Y \sim \text{Poisson}(\lambda s)$  die Ereignisse in  $(t, t + s]$ , so modelliert  $X + Y \sim \text{Poisson}(\lambda(t + s))$  die Ereignisse in  $(0, t + s]$ .

Unter diesen Annahmen folgt für die Wahrscheinlichkeit in  $(0, t]$  mindestens  $r$  Ereignisse zu beobachten

$$\mathbb{P}((0, t)) = 1 - \sum_{k=0}^{r-1} \underbrace{e^{-\lambda t} \frac{(\lambda t)^k}{k!}}_{\text{Zähldichte Poisson}(\lambda t) \text{ in } k}$$

Wegen

$$\begin{aligned} \frac{d}{dt} \left( 1 - \sum_{k=0}^{r-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \right) &= -(-\lambda) e^{-\lambda t} \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} - e^{-\lambda t} \sum_{k=1}^{r-1} \frac{\lambda^k t^{k-1}}{(k-1)!} \\ &= e^{-\lambda t} \left( \sum_{k=0}^{r-1} \frac{\lambda^{k+1} t^k}{k!} - \sum_{k=0}^{r-2} \frac{\lambda^{k+1} t^k}{k!} \right) \\ &= e^{-\lambda t} \frac{\lambda^r t^{r-1}}{(r-1)!} \end{aligned}$$

gilt

$$\mathbb{P}((0, t)) = \int_0^t e^{-\lambda x} \frac{\lambda^r x^{r-1}}{(r-1)!} dx$$

Wir definieren allgemeiner:

#### Definition 4.2 (Gammaverteilung, Gammafunktion)

Seien  $\lambda > 0$  und  $r > 0$ . Dann heißt die Verteilung auf  $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$  mit Dichte

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \quad \text{mit} \quad \Gamma(r) = \int_0^\infty y^{r-1} e^{-y} dy \quad (r > 0)$$

**Gammaverteilung** mit Parametern  $\lambda, r$ . Schreibe  $\text{Gamma}(\lambda, r)$ .

Insbesondere ist  $\text{Gamma}(\lambda, 1)$  gerade die **Exponentialverteilung**  $\text{Exp}(\lambda)$  (vgl. Beispiel 1.7).

**Anmerkung.** Die Funktion  $\Gamma$  in Definition 4.2 wird auch als *Gamma-Funktion* bezeichnet.

Die Gammaverteilung ist **reproduktiv**: Die Wartezeit auf  $r + s$  Ereignisse entspricht der Wartezeit auf  $r$  Ereignisse +  $s$  (weitere) Ereignisse:

#### Lemma 4.3

Seien  $X \sim \text{Gamma}(\lambda, r), Y \sim \text{Gamma}(\lambda, s)$  unabhängig. Dann gilt

$$X + Y \sim \text{Gamma}(\lambda, r + s)$$

*Beweis.* Hier nur für  $r, s \in \mathbb{N}$ , allgemein später mit momenterzeugenden Funktionen. Seien  $\rho_X,$

$\rho_Y$  Dichten von  $X, Y$ . Nach Satz 3.25 folgt

$$\begin{aligned}
 \rho_{X+Y}(x) &= (\rho_X * \rho_Y)(x) = \int_{\mathbb{R}} \rho_X(y) \rho_Y(x-y) \, dy \\
 &= \int_0^x \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y} \frac{\lambda^s}{\Gamma(s)} (x-y)^{s-1} e^{-\lambda(x-y)} \, dy \\
 &= \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} e^{-\lambda x} \int_0^x y^{r-1} (x-y)^{s-1} \, dy \\
 &\stackrel{\text{P.I.}}{=} \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} e^{-\lambda x} \left( \underbrace{\left[ \frac{1}{r} y^r (x-y)^{s-1} \right]_{y=0}^x}_{=0} + \frac{s-1}{r} \int_0^x y^r (x-y)^{s-2} \, dy \right) \\
 &\stackrel{\text{Induktion}}{=} \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(r)} e^{-\lambda x} \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} = \frac{\lambda^{r+s}}{\Gamma(r+s)} e^{-\lambda x} \quad \square
 \end{aligned}$$

Exponentialverteilungen sind zudem **gedächtnislos**:

#### Lemma 4.4 (Gedächtnislosigkeit der Exponentialverteilung)

Sei  $X \sim \text{Exp}(\lambda)$ . Dann gilt

$$\mathbb{P}(X > t) = \mathbb{P}(X > t+s \mid X > s) \quad (t, s \geq 0) \quad (\star)$$

*Beweis.*

$$\begin{aligned}
 \mathbb{P}(X > t+s \mid X \geq s) &= \frac{\mathbb{P}(X > t+s, X > s)}{\mathbb{P}(X > s)} \\
 &= \frac{\mathbb{P}(X > t+s)}{\mathbb{P}(X > s)} \\
 &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda t} = \mathbb{P}(X > t) \quad \square
 \end{aligned}$$

#### Beispiel 4.5

Eine Studentin wartet morgens eine  $\text{Exp}(1/5)$  verteilte Zeit  $X$  auf den Bus zur Uni. Die Wahrscheinlichkeit einer Wartezeit  $\geq 5$  Minuten beträgt

$$\mathbb{P}(X \geq 5) = e^{-\frac{1}{5} \cdot 5} = e^{-1} \approx 0,37$$

An einem kalten, stürmischen Frühlingstag hat die Studentin bereits 10 Minuten gewartet. Die Wahrscheinlichkeit mindestens fünf weitere Minuten zu warten ist

$$\mathbb{P}(X \geq 15 \mid X \geq 10) = \mathbb{P}(X \geq 5) = e^{-1} \approx 0,37$$

#### Bemerkung

Man kann sogar zeigen, dass die Exponentialverteilung die einzige absolutstetige Verteilung mit der Eigenschaft  $(\star)$  ist.

## ERWARTUNGSWERT UND VARIANZ

### 5.1 Der Erwartungswert

#### Definition 5.1 (Erwartungswert)

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  Wahrscheinlichkeitsraum und  $X: (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  Zufallsvariable. Dann ist

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

der **Erwartungswert** von  $X$ .

#### Bemerkung

Der Erwartungswert von  $X$  existiert genau dann, wenn

$$\int_{\Omega} |X| d\mathbb{P} < \infty \quad \text{bzw.} \quad \mathbb{E}[|X|] < \infty$$

d.h. genau dann, wenn  $X \in \mathcal{L}^1(\mathbb{P})$ . Für nichtnegative Zufallsvariablen ist der Erwartungswert immer definiert, wenn wir  $+\infty$  als zulässigen Wert annehmen, was wir in der Folge auch tun.

#### Beispiel 5.2

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum,  $A \in \mathcal{F}$  und sei  $X: (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  die Indikatorvariable  $X(\omega) = \mathbb{1}_A(\omega)$ . Dann gilt:  $X \in \mathcal{L}^1(\mathbb{P})$  und

$$\mathbb{E}[X] = \int_{\Omega} \mathbb{1}_A(\omega) \mathbb{P}(d\omega) = \int_A \mathbb{P}(d\omega) = \mathbb{P}(A)$$

#### Proposition 5.3

Sei  $X: (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$  eine Zufallsvariable und  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  Borel-messbar. Dann gilt

$$\mathbb{E}[f \circ X] = \mathbb{E}[f(X)] = \int_{\Omega} f(X) d\mathbb{P} = \int_{\Omega} f(X(\omega)) d\mathbb{P}(\omega) = \int_{\mathbb{R}^n} f(x) \mathbb{P}(X \in dx)$$

*Beweis.*  $f \circ X = f(X)$  ist eine reelle Zufallsvariable. Die Formel folgt direkt auf dem Transformationssatz für Bildmaße (↗ Schilling MINT Satz 18.1).  $\square$

**Proposition 5.4 (Erwartungswerte bei Existenz einer (Zähl-)dichte)**

Sei  $X: (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$  eine Zufallsvariable und  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  Borel-messbar.

- (1) **diskreter Fall:** Ist  $\mathbb{P}_X$  ein Wahrscheinlichkeitsmaß auf  $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$  mit Zähldichte  $\rho$ , so ist

$$\mathbb{E}[f(X)] = \sum_{x \in \mathbb{Z}} f(x) \rho(x)$$

- (2) **stetiger Fall:** Besitzt  $\mathbb{P}_X$  eine Dichte  $\rho$  (bzgl. Lebesguemaß), so ist

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) \rho(x) \, dx$$

*Beweis.* Klar aus Definition 5.1 und Proposition 5.3. □

**Beispiel 5.5**

Sei  $X \sim \text{Bin}(n, p)$ . Dann gilt

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=0}^n k \cdot \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n \frac{n!}{(n-k)!(k-1)!} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \underbrace{\binom{n-1}{k-1} p^{k-1} (1-p)^{n-1-(k-1)}}_{\text{Zähldichte Bin}(n-1, p)} \\ &= np(p + 1 - p)^{n-1} \\ &= np \end{aligned}$$

Da der Erwartungswert ein Integral ist, übertragen sich viele Eigenschaften.

**Proposition 5.6 (Eigenschaften des Erwartungswertes)**

Seien  $X, X_n, Y: (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  ( $n \in \mathbb{N}$ ) Zufallsvariablen in  $\mathcal{L}^1(\mathbb{P})$  und  $a, b \in \mathbb{R}$  konstant.

- (1) **Linearität:**  $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
- (2) **Monotonie:**  $X \leq Y$ , d.h.  $X(\omega) \leq Y(\omega) \quad \forall \omega \in \Omega$ . Dann gilt  $\mathbb{E}[X] \leq \mathbb{E}[Y]$  und insbesondere gilt  $X \geq 0 \Rightarrow \mathbb{E}[X] \geq 0$ .
- (3) **Lemma von Fatou:**  $\mathbb{E}[\liminf_{n \rightarrow \infty} X_n] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n]$
- (4) **Satz von Beppo-Levi:** Wenn  $X_n \geq 0$  und  $X_n \uparrow X$  so gilt  $\mathbb{E}[X] = \sup_{n \in \mathbb{N}} \mathbb{E}[X_n] = \lim_{n \rightarrow \infty} \mathbb{E}[X_n]$
- (5) **Dominierte Konvergenz / Satz von Lebesgue:** Sei  $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$  und  $\mathbb{P}(\{\omega : |X_n(\omega)| \leq Y(\omega)\}) = 1$  (d.h.  $|X| \leq Y$   $\mathbb{P}$ -fast sicher) für  $Y \in \mathcal{L}^1(\mathbb{P})$ . Dann gilt  $X \in \mathcal{L}^1(\mathbb{P})$  und  $\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}[X]$ .
- (6) **Markov-Ungleichung:** Sei  $\varepsilon > 0$ . Dann gilt  $\mathbb{P}(|X| \geq \varepsilon) \leq \frac{1}{\varepsilon} \mathbb{E}[|X|]$
- (7) **Hölder-Ungleichung:** Sei  $1 \leq p, q \leq \infty, \frac{1}{p} + \frac{1}{q} = 1$ . Dann gilt

$$\mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^p])^{1/p} (\mathbb{E}[|Y|^q])^{1/q}$$

Für  $p = q = 2$  ergibt sich die CAUCHY-SCHWARZ-Ungleichung.

- (8) **Jensen-Ungleichung:** Sei  $X \geq 0$  und  $\Phi: [0, \infty) \rightarrow [0, \infty)$  konvex, messbar. Dann gilt  $\Phi(\mathbb{E}[X]) \leq \mathbb{E}[\Phi(X)]$ .

Beweis. ↗ Schilling MINT. □

**Beispiel 5.7**

Da für  $X_1, \dots, X_n \sim \text{Bernoulli}(p)$  unabhängig gilt, dass  $\underbrace{X_1 + \dots + X_n}_{=: X} \sim \text{Bin}(n, p)$ . Damit folgt

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[X_1 + \dots + X_n] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_n] = n\mathbb{E}[X_1] \\ &= n \cdot (1 \cdot \underbrace{\mathbb{P}(X_1 = 1)}_{=p} + 0 \cdot \underbrace{\mathbb{P}(X_1 = 0)}_{=1-p}) \\ &= n \cdot p \end{aligned}$$

**Proposition 5.8 (Produktformel für Erwartungswerte)**

Seien  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $X_1, \dots, X_n: \Omega \rightarrow \mathbb{R}^d$  unabhängige Zufallsvariablen und  $f_1, \dots, f_n: \mathbb{R}^d \rightarrow \mathbb{R}$  messbare Funktionen. Wenn entweder  $f_i(X_i) \geq 0$  ( $i = 1, \dots, n$ ) oder  $f_i(X_i) \in \mathcal{L}^1(\mathbb{P})$  ( $i = 1, \dots, n$ ), dann gilt

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

Für den Beweis von Proposition 5.8 benötigen wir:

**Lemma 5.9**

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum,  $X, Y: \Omega \rightarrow \mathbb{R}^d$  unabhängige Zufallsvariablen und  $h: \mathbb{R}^{2d} \rightarrow \mathbb{R}$  messbar. Falls  $h \geq 0$  oder  $h(X, Y) \in \mathcal{L}^1(\mathbb{P})$ , dann gilt

$$\begin{aligned} \mathbb{E}[h(X, Y)] &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(x, y) \mathbb{P}(X \in dx) \mathbb{P}(Y \in dy) \\ &= \mathbb{E} \left[ \int_{\mathbb{R}^d} h(X, y) \mathbb{P}(Y \in dy) \right] \\ &= \mathbb{E} \left[ \int_{\mathbb{R}^d} h(x, Y) \mathbb{P}(X \in dx) \right] \end{aligned}$$

*Beweis.*  $h(X, Y)$  ist eine reelle Zufallsvariable und

$$\begin{aligned} \mathbb{E}[h(X, Y)] &= \int_{\Omega} h(X(\omega), Y(\omega)) \, d\mathbb{P}(\omega) \\ &\stackrel{5.3}{=} \int_{\mathbb{R}^{2d}} h(x, y) \mathbb{P}(X \in dx, Y \in dy) \\ &\stackrel{X \perp\!\!\!\perp Y, 3.21}{=} \int_{\mathbb{R}^{2d}} h(x, y) \mathbb{P}_X \otimes \mathbb{P}_Y(dx, dy) \\ &\stackrel{\text{FUBINI}}{=} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(x, y) \mathbb{P}_X(dx) \mathbb{P}_Y(dy) \\ &\stackrel{5.3}{=} \int_{\Omega} \int_{\mathbb{R}^d} h(x, Y(\omega)) \mathbb{P}_X(dx) \, d\mathbb{P}(\omega) \\ &= \mathbb{E} \left[ \int_{\mathbb{R}^d} h(x, Y) \mathbb{P}_X(dx) \right] \end{aligned}$$

□

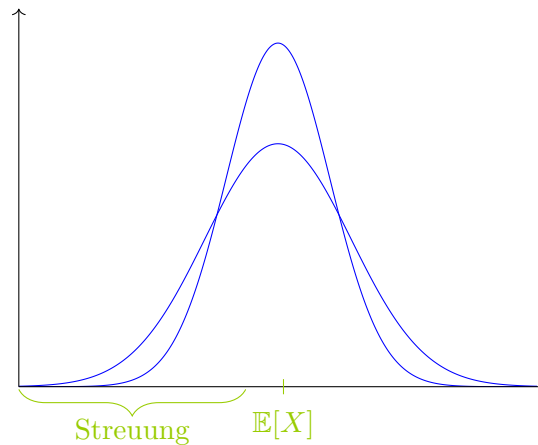
*Beweis* (Proposition 5.8). Betrachte  $n = 2$ , Zufallsvariablen  $X, Y$  und Abbildungen  $f, g$ . Setze  $h(x, y) = f(x)g(y)$ , dann folgt für  $f, g \geq 0$  mit 5.9

$$\begin{aligned} \mathbb{E}[f(X)g(Y)] &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x)g(y) \mathbb{P}(X \in dx) \mathbb{P}(Y \in dy) \\ &= \int_{\mathbb{R}^d} f(x) \mathbb{P}(X \in dx) \int_{\mathbb{R}^d} g(y) \mathbb{P}(Y \in dy) \\ &= \mathbb{E}[f(X)] \cdot \mathbb{E}[g(Y)] \end{aligned}$$

Für  $f(X), g(Y) \in \mathcal{L}^1(\mathbb{P})$  zeigt obige Rechnung

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[|f(X)|] \mathbb{E}[|g(Y)|] < \infty$$

also  $f(X)g(Y) \in \mathcal{L}^1(\mathbb{P})$ . Die Aussage folgt über die obige Rechnung. Für allgemeines  $n$  folgt Proposition 5.8 durch Iteration mit Korollar 3.19. □



## 5.2 Varianz und höhere Momente

### Definition 5.10 ( $k$ -te Momente)

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $X: (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  eine reelle Zufallsvariable. Dann ist für  $k \in \mathbb{N}$

$$\mathbb{E}[X^k] = \int_{\Omega} X^k(\omega) \mathbb{P}(d\omega) = \int_{\mathbb{R}} x^k \mathbb{P}(X \in dx)$$

das  $k$ -te **Moment** von  $X$  (sofern definiert).

### Bemerkung

- Erwartungswert  $\cong$  erstes Moment
- Das  $k$ -te Moment existiert genau dann, wenn  $\int_{\Omega} |X(\omega)^k| \mathbb{P}(d\omega) < \infty$  bzw.  $X \in \mathcal{L}^k(\mathbb{P})$ .
- MINT:  $\mathcal{L}^r(\mathbb{P}) \subseteq \mathcal{L}^s(\mathbb{P})$  für  $s \leq r$

Von Interesse ist insbesondere das zweite Moment.

### Definition 5.11 (Varianz, Standardabweichung)

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $X, Y \in \mathcal{L}^2(\mathbb{P})$  reelle Zufallsvariablen.

- (1) Die **Varianz** von  $X$  ist  $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- (2) Die **Standardabweichung**/Streuung von  $X$  ist  $\sqrt{\text{Var}(X)}$ .
- (3) Die **Kovarianz** von  $X$  und  $Y$  ist

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- (4) Sind  $\text{Var}(X), \text{Var}(Y) \geq 0$ , dann ist die **Korrelation** von  $X$  und  $Y$  definiert als

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

- (5) Gilt  $\text{Corr}(X, Y) = 0$ , so heißen  $X, Y$  **unkorreliert**.



**Bemerkung**

- Die Endlichkeit der Ausdrücke in Definition 5.11 folgt aus der CAUCHY-SCHWARZ-Ungleichung  

$$\mathbb{E}[|XY|] \leq (\mathbb{E}[X|X|^2])^{\frac{1}{2}} \cdot (\mathbb{E}[X|Y|^2])^{\frac{1}{2}}$$
- Für die (Ko)varianz gilt

$$\begin{aligned} \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] &= \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \end{aligned}$$

**Beispiel 5.12**

Sei  $X \sim \text{Bin}(n, p)$ , dann gilt

$$\text{Var}(X) = \mathbb{E}[X^2] - \underbrace{\mathbb{E}[X]^2}_{=np}$$

mit

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n k \frac{(n-1)!}{(n-1-(k-1))!(k-1)!} p^{k-1} (1-p)^{n-1-(k-1)} \\ &= np \sum_{\ell=0}^{n-1} (\ell+1) \binom{n-1}{\ell} p^\ell (1-p)^{n-1-\ell} \quad \text{mit } \ell = k-1 \\ &= np \left( 1 + \sum_{\ell=0}^{n-1} \ell \binom{n-1}{\ell} p^\ell (1-p)^{n-1-\ell} \right) \\ &= np(1 + (n-1)p \cdot 1) \quad \text{mit binomischem Lehrsatz} \\ &= np + n(n-1)p^2 \end{aligned}$$

Damit ist  $\text{Var}(X) = np + n^2p^2 - np^2 - n^2p^2 = np(1-p)$ .

**Proposition 5.13 (Eigenschaften der (Ko-)Varianz)**

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum,  $X, Y, X_1, \dots, X_n \in \mathcal{L}^2(\mathbb{P})$ ,  $a, b \in \mathbb{R}$ .

- (1)  $\text{Var}(aX + b) = a^2 \text{Var}(X)$
- (2) Sei  $\text{Cov}(X, Y)^2 \leq \text{Var}(X) \text{Var}(Y)$  und insbesondere  $|\text{Corr}(X, Y)| \leq 1$
- (3)  $\text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i,j=1, i \neq j}^n \text{Cov}(X_i, X_j)$ . Sind die  $X_1, \dots, X_n$  paarweise unkorreliert, so gilt die **Formel von Bienaymé**:

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

- (4)  $X \perp\!\!\!\perp Y \Rightarrow \text{Corr}(X, Y) = 0$
- (5) **Tschebyscheff-Ungleichung:** Für  $\varepsilon > 0$  gilt

$$\mathbb{P}(|X - \mathbb{E}[X]| > \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}$$

*Beweis.* (zu 1) Da  $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ , folgt

$$\mathbb{V}\text{ar}(aX + b) = \mathbb{E}[(a\mathbb{E}[X] + b - (a\mathbb{E}[X] + b))^2] = \mathbb{E}[a^2(X - \mathbb{E}[X])^2] = a^2\mathbb{V}\text{ar}(X)$$

(zu 2) Wegen (1) können wir oBdA annehmen, dass  $\mathbb{E}[X] = 0 = \mathbb{E}[Y]$ . Dann wird (2) zu  $\mathbb{E}[XY]^2 \leq \mathbb{E}[X] \cdot \mathbb{E}[Y]$  und dies ist die CAUCHY-SCHWARZ Ungleichung.

(zu 3) Wähle wieder oBdA  $\mathbb{E}[X_i] = 0$ . Dann

$$\begin{aligned} \mathbb{V}\text{ar}\left(\sum_{i=1}^n X_i\right) &= \mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^2\right] \\ &= \mathbb{E}\left[\sum_{i,j=1}^n X_i X_j\right] \\ &= \sum_{i,j=1}^n \mathbb{E}[X_i X_j] \\ &= \sum_{i,j=1}^n \mathbb{C}\text{ov}(X_i, X_j) \\ &= \sum_{i=1}^n \mathbb{V}\text{ar}(X_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^n \mathbb{C}\text{ov}(X_i, X_j) \end{aligned}$$

(zu 4) Aus der Unabhängigkeit von  $X$  und  $Y$  folgt  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ . Somit gilt

$$\mathbb{C}\text{ov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0 \quad \Rightarrow \quad \mathbb{C}\text{orr}(X, Y) = 0$$

(zu 5) Wende die MARKOV-Ungleichung auf  $X' = (X - \mathbb{E}[X])^2$  an, dann folgt

$$\begin{aligned} \mathbb{P}(|X - \mathbb{E}[X]| > \varepsilon) &= \mathbb{P}((X - \mathbb{E}[X])^2 > \varepsilon^2) \\ &\stackrel{\text{MARKOV}}{\leq} \frac{1}{\varepsilon^2} \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \frac{\mathbb{V}\text{ar}(X')}{\varepsilon^2} \end{aligned}$$

□

## 5.3 Wahrscheinlichkeitserzeugende Funktionen

**Definition 5.14 (Wahrscheinlichkeitserzeugende Funktion)**

(1) Ist  $\mathbb{P}$  ein Wahrscheinlichkeitsmaß auf  $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$  mit Zähldichte  $\rho$ , so heißt

$$\psi_{\mathbb{P}}(s) := \sum_{k \in \mathbb{N}_0} s^k \rho(k) \quad (0 \leq s \leq 1)$$

**wahrscheinlichkeitserzeugende Funktion** von  $\mathbb{P}$  (probability generating function - pgf).

(2) Ist  $X$  eine  $\mathbb{N}_0$ -wertige Zufallsvariable auf  $(\Omega, \mathcal{F}, \mathbb{P})$ , so heißt

$$\psi_X(s) := \sum_{k \in \mathbb{N}_0} s^k \mathbb{P}(X = k) \quad (0 \leq s \leq 1)$$

**wahrscheinlichkeitserzeugende Funktion** von  $X$ .

**Bemerkung**

- Da  $\sum_{k \in \mathbb{N}_0} \rho(k) = 1$  ist, ist die pgf auf  $0 \leq s \leq 1$  wohldefiniert. Zudem ist  $\psi$  auf  $[0, 1)$  unendlich oft differenzierbar.
- Da  $\rho(k) \geq 0 \forall k$  ist die pgf stets konvex.
- Durch Taylorentwicklung von  $\psi$  in 0 erhält man:

$$\psi_{\mathbb{P}}(s) = \sum_{k \in \mathbb{N}_0} s^k \frac{\psi_{\mathbb{P}}^{(k)}(0)}{k!}$$

so dass für alle  $k \in \mathbb{N}_0$  folgt

$$\rho(k) = \frac{\psi_{\mathbb{P}}^{(k)}(0)}{k!}$$

Die Verteilung  $\mathbb{P}$  (bzw.  $\mathbb{P} \circ X^{-1}$ ) ist durch  $\psi_{\mathbb{P}}$  (bzw.  $\psi_X$ ) eindeutig bestimmt. Also “erzeugt  $\psi$  die Zähldichte”.

**Beispiel 5.15**

Ist  $X \sim \text{Bin}(n, p)$ , so folgt

$$\begin{aligned} \psi_X(s) &= \sum_{k=0}^n s^k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} (ps)^k (1-p)^{n-k} \\ &= (ps + (1-p))^n \\ &= (1 + p(s-1))^n \end{aligned}$$

**Proposition 5.16 (Momentenberechnung mit pgf)**

Sei  $X$  eine  $\mathbb{N}_0$ -wertige Zufallsvariable, dann gilt

$$\mathbb{E}[X^n] < \infty \quad (n \geq 1) \iff \psi_X^{(n)}(1) = \lim_{s \nearrow 1} \psi_X^{(n)}(s) < \infty.$$

Insbesondere gilt dann

$$\psi_X^{(n)}(1) = \mathbb{E}[X(X-1)\cdots(X-n+1)]$$

*Beweis.* Sei  $\rho(k) = \mathbb{P}(X = k)$ . Durch  $n$ -faches gliedweises Differenzieren der Potenzreihe  $\psi_X$  folgt

$$\psi_X^{(n)}(s) = \sum_{k \in \mathbb{N}_0} k(k-1)\cdots(k-n+1)\rho(k)s^{k-n} \quad (0 \leq s < 1)$$

Dann existieren in  $[0, \infty)$

$$\begin{aligned} \lim_{s \nearrow 1} \psi_X^{(n)}(s) &= \lim_{s \nearrow 1} \sum_{k=0}^{\infty} k(k-1)\cdots(k-n+1)\rho(k)s^{k-n} \\ &= \sum_{k=n}^{\infty} \rho(k)k(k-1)\cdots(k-n+1) \\ &= \mathbb{E}[X(X-1)\cdots(X-n+1)] \end{aligned}$$

sowie induktiv

$$\begin{aligned} \psi_X^{(n)}(1) &= \lim_{s \nearrow 1} \frac{\psi_X^{(n-1)}(1) - \psi_X^{(n-1)}(s)}{1-s} \\ &= \lim_{s \nearrow 1} \sum_{k \in \mathbb{N}_0} \rho(k)k(k-1)\cdots(k-(n-1)+1) \frac{1-s^{k-(n-1)}}{1-s} \\ &= \lim_{s \nearrow 1} \sum_{k \in \mathbb{N}_0} \rho(k)k(k-1)\cdots(k-n+2) \sum_{\ell=0}^{k-(n-1)} s^\ell && \text{(Geom. Reihe)} \\ &= \sum_{k \in \mathbb{N}_0} \rho(k)k(k-1)\cdots(k-n+2)(k-n+1) && \text{(Monotonie)} \\ &= \mathbb{E}[X(X-1)\cdots(X-n+1)] \end{aligned}$$

Insbesondere gilt  $\mathbb{E}[X^n] < \infty$  genau dann, wenn  $\psi_X^{(n)}(1) < \infty$  bzw.  $\lim_{s \nearrow 1} \psi_X^{(n)}(s) < \infty$  □

**Beispiel 5.17**

Sei  $X \sim \text{Bin}(n, p)$ , dann gilt nach Beispiel 5.15  $\psi_X(s) = (1 + p(s-1))^n$ . Damit

$$\begin{aligned} \psi_X'(s) &= n(1 + p(s-1))^{n-1}p \\ \psi_X''(s) &= n(n-1)(1 + p(s-1))^{n-2} \cdot p^2 \end{aligned}$$

so dass

$$\begin{aligned}\mathbb{E}[X] &= \psi'_X(1) = np \\ \text{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 \\ &= \psi''_X(1) + \psi'_X(1) - (\psi'_X(1))^2 \\ &= n(n-1)p^2 + np - (np)^2 = np - np^2 = np(1-p)\end{aligned}$$

### Proposition 5.18

Seien  $X, Y$  unabhängige,  $\mathbb{N}_0$ -wertige Zufallsvariablen auf  $(\Omega, \mathcal{F}, \mathbb{P})$ . Dann gilt

$$\begin{aligned}\psi_{X+Y}(s) &= \mathbb{E}[s^{X+Y}] = \mathbb{E}[s^X s^Y] \\ &= \mathbb{E}[s^X] \mathbb{E}[s^Y] \\ &= \psi_X(s) \psi_Y(s)\end{aligned}$$

### Proposition 5.19

Sind  $\mathbb{P}_1, \mathbb{P}_2$  Wahrscheinlichkeitsmaße auf  $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$ , so gilt

$$\psi_{\mathbb{P}_1 * \mathbb{P}_2}(s) = \psi_{\mathbb{P}_1}(s) \psi_{\mathbb{P}_2}(s) \quad (0 \leq s \leq 1)$$

# Kapitel 6

## NORMALVERTEILUNG

### Definition 6.1

Sei  $\mu \in \mathbb{R}$ ,  $\sigma^2 > 0$ . Das Wahrscheinlichkeitsmaß  $\mathcal{N}(\mu, \sigma^2)$  auf  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  mit Dichtefunktion

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

heißt **Normalverteilung** mit Parametern  $\mu$  und  $\sigma^2$ . Im Fall  $\mu = 0$  und  $\sigma^2 = 1$ , heißt  $\mathcal{N}(0, 1)$  **Standardnormalverteilung**.

### Bemerkung

- $g_{\mu, \sigma^2}$  ist eine Wahrscheinlichkeitsdichte, da  $\int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$ .
- $\mu$  und  $\sigma^2$  sind gerade *Erwartungswert* und *Varianz* der Normalverteilung  $\mathcal{N}(\mu, \sigma^2)$ . Sei dazu  $X \sim \mathcal{N}(\mu, \sigma^2)$ .

$$\begin{aligned} \mathbb{E}[X] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx && \text{substituiere } y = \frac{x-\mu}{\sqrt{\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (\sqrt{\sigma^2}y + \mu) e^{-\frac{y^2}{2}} dy \\ &= \underbrace{\sqrt{\frac{\sigma^2}{2\pi}} \int_{\mathbb{R}} y e^{-\frac{y^2}{2}} dy}_{=0, \text{ wg. Symmetrie}} + \underbrace{\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mu e^{-\frac{y^2}{2}} dy}_{\mu \sqrt{2\pi}} \\ &= \mu \\ \text{Var}(X) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx && \text{substituiere } y = \frac{x-\mu}{\sqrt{\sigma^2}} \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} y^2 e^{-\frac{y^2}{2}} dy \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} (-y) \left( -y e^{-\frac{y^2}{2}} \right) dy \\ &\stackrel{\text{part. Int.}}{=} \frac{\sigma^2}{\sqrt{2\pi}} \left( \underbrace{\left[ -y e^{-\frac{y^2}{2}} \right]_{-\infty}^{\infty}}_{=0} + \underbrace{\int_{\mathbb{R}} e^{-\frac{y^2}{2}} dy}_{=\sqrt{2\pi}} \right) \\ &= \sigma^2 \end{aligned}$$

Die Popularität der Normalverteilung erklärt sich insbesondere aus dem zentralen Grenzwertsatz. Hier eine einfache Form davon:

**Proposition 6.2 (de MOIVRE-LAPLACE, lokaler Grenzwertsatz)**

Seien  $p \in (0, 1)$  und  $B_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}$  die Zähldichte von  $\text{Bin}(n, p)$ ,  $g(x)$  die Dichte der Standardnormalverteilung. Dann gilt für beliebiges  $c > 0$  mit

$$x_n(k) = \frac{k - np}{\sqrt{np(1-p)}}$$

der Grenzwert

$$\lim_{n \rightarrow \infty} \max_{k: |x_n(k)| < c} \left| \frac{\sqrt{np(1-p)} \cdot B_{n,p}(k)}{g(x_n(k))} - 1 \right| = 0$$

*Beweis.* siehe Handout bzw. später als Korollar des zentralen Grenzwertsatzes □

# Kapitel 7

## MOMENTERZEUGENDE UND CHARAKTERISTISCHE FUNKTIONEN

### 7.1 Momenterzeugende Funktionen

**Ziel:** “Übersetze” Verteilungen in Funktionen. Insbesondere einfache Faltungsoperation ( $\nearrow$  5.3).

#### Definition 7.1

(1) Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $X: \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable. Dann heißt

$$m_X(u) := \mathbb{E} [e^{uX}] \quad (u \in \mathbb{R}), \text{ so dass } m_X(u) < \infty$$

**momenterzeugende Funktion** von  $X$  (mgf = moment generating function)

(2) Ist  $\mathbb{P}$  eine Verteilung auf  $\mathbb{R}$ , so heißt

$$m_{\mathbb{P}}(u) := \int_{\mathbb{R}} e^{ux} \mathbb{P}(dx) \quad (u \in \mathbb{R}), \text{ so dass } m_{\mathbb{P}}(u) < \infty$$

**momenterzeugende Funktion** von  $\mathbb{P}$ .

#### Beispiel 7.2

Sei  $X \sim \text{Gamma}(\lambda, r)$ .

$$\begin{aligned} m_X(u) &= \mathbb{E} [e^{uX}] \\ &= \int_0^\infty e^{ux} \lambda e^{-\lambda x} \frac{(\lambda x)^{r-1}}{\Gamma(r)} dx \\ &= \frac{\lambda^r}{\Gamma(r)} \int_0^\infty e^{-(\lambda-u)x} x^{r-1} dx \quad \text{substituiere } y = (\lambda - u)x \\ &= \frac{\lambda^r}{\Gamma(r)} \int_0^\infty e^{-y} \frac{y^{r-1}}{(\lambda - u)^{r-1}} \frac{dy}{(\lambda - u)} \\ &= \left( \frac{\lambda}{\lambda - u} \right)^r \quad u < \lambda \end{aligned}$$

#### Lemma 7.3

Ist  $X$   $\mathbb{N}_0$ -wertig, so gilt für alle  $u \in \mathbb{R}$  mit  $m_X(u) < \infty$

$$m_X(u) = \mathbb{E} [e^{uX}] = \psi_X(e^u)$$

*Beweis.* Klar, da folgt aus Definition 7.1. □



**Proposition 7.4**

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $X, Y: \Omega \rightarrow \mathbb{R}$  Zufallsvariablen mit mgf's  $m_X$  bzw.  $m_Y$ . Es gelten:

- (1)  $m_X(0) = 1$
- (2) Für  $a, b \in \mathbb{R}$  gilt  $m_{aX+b}(u) = e^{bu} \cdot m_X(au)$  für  $u$  so dass  $m_X(au) < \infty$ .
- (3)  $X \perp\!\!\!\perp Y \Rightarrow m_{X+Y}(u) = m_X(u) \cdot m_Y(u)$  für alle  $u$ , so dass  $m_X(u), m_Y(u) < \infty$

*Beweis.* (zu 1) Klar.

$$(zu\ 2) \ m_{aX+b}(u) = \mathbb{E}[e^{aXu+bu}] = e^{bu} \mathbb{E}[e^{auX}] = e^{bu} m_X(au).$$

$$(zu\ 3) \ m_{X+Y}(u) = \mathbb{E}[e^{Xu+Yu}] \stackrel{X \perp\!\!\!\perp Y}{=} \mathbb{E}[e^{uX}] \mathbb{E}[e^{uY}] = m_X(u) \cdot m_Y(u) \quad \square$$

Der Bezeichnung “momenterzeugend” erklärt sich mit der folgenden Proposition:

**Proposition 7.5**

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $X: \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable mit momenterzeugender Funktion  $m_X$ , so dass ein  $\varepsilon > 0$  existiert mit  $m_X(u) < \infty$  auf  $[0, \varepsilon)$ . Dann gilt

$$\mathbb{E}[X^n] = \frac{d^n}{du^n} m_X(0) \quad \forall n \in \mathbb{N}$$

*Beweis.* Für alle  $u \in \mathbb{R}$  mit  $m_X(u) < \infty$  folgt

$$m_X(u) = \mathbb{E}[e^{uX}] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(uX)^k}{k!}\right] \stackrel{\text{LEBESGUE}}{=} \sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]}{k!} u^k$$

$n$ -fachen Differenzieren liefert

$$\frac{d^n}{du^n} m_X(0) = \sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]}{k!} k(k-1) \cdots (k-n+1) u^{k-n}$$

so dass

$$\frac{d^n}{du^n} m_X(0) = \frac{\mathbb{E}[X^n]}{n!} n(n-1) \cdots (n-n+1) = \mathbb{E}[X^n] \quad \square$$

Die mgf charakterisiert eine Verteilung eindeutig:

**Satz 7.6**

Seien  $(\Omega, \mathcal{F}, \mathbb{R}), (\Omega', \mathcal{F}', \mathbb{P}')$  Wahrscheinlichkeitsräume und  $X: \Omega \rightarrow \mathbb{R}$  sowie  $Y: \Omega' \rightarrow \mathbb{R}$  Zufallsvariablen mit mgfs  $m_X, m_Y$ . Wenn  $m_X(u)$  und  $m_Y(u)$  in einer Umgebung um Null definiert sind und im Definitionsbereich gilt  $m_X(u) = m_Y(u)$ , so haben  $X$  und  $Y$  die selben Verteilungen ( $X \stackrel{d}{=} Y$ ).

*Beweis.* Sind  $X, Y$   $\mathbb{N}_0$ -wertig, so folgt dies aus Lemma 7.3 und dem entsprechenden Resultat für pgfs. Der allgemeine Fall folgt aus dem Resultat zu charakteristischen Funktionen (??).  $\square$

**Beispiel 7.7 ((vgl. Lemma 4.3))**

Seien  $X \sim \text{Gamma}(\lambda, r)$  und  $Y \sim \text{Gamma}(\lambda, s)$  unabhängig, dann gilt nach Beispiel 7.2

$$m_X(u) = \left(\frac{\lambda}{\lambda - u}\right)^r \quad (u < \lambda) \quad \text{und} \quad m_Y(u) = \left(\frac{\lambda}{\lambda - u}\right)^s \quad (u < \lambda)$$

und nach ??

$$m_{X+Y}(u) = m_X(u) \cdot m_Y(u) = \left(\frac{\lambda}{\lambda - u}\right)^r \cdot \left(\frac{\lambda}{\lambda - u}\right)^s = \left(\frac{\lambda}{\lambda - u}\right)^{r+s}$$

Dies ist die mgf einer  $\text{Gamma}(\lambda, r+s)$  Verteilung. Nach Satz 7.6 folgt  $X+Y \sim \text{Gamma}(\lambda, r+s)$ .

## 7.2 Charakteristische Funktionen

**Definition 7.8**

(1) Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $X: \Omega \rightarrow \mathbb{R}^d$  eine Zufallsvariable. Dann ist

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

**charakteristische Funktion** von  $X$ .

(2) Ist  $\mathbb{P}$  Verteilung in  $\mathbb{R}^d$ , dann ist

$$\varphi_{\mathbb{P}}(u) := \int_{\mathbb{R}^d} e^{i\langle u, x \rangle} \mathbb{P}(dx) \quad (u \in \mathbb{R}^d)$$

**charakteristische Funktion** von  $\mathbb{P}$ .

**Bemerkung**

- Da  $|e^{i\langle u, X \rangle}| = 1$ , ist die charakteristische Funktion für *alle*  $u \in \mathbb{R}^d$  definiert.
- Die charakteristische Funktion ist die inverse FOURIERtransformation des Maßes  $\mathbb{P}$  bzw.  $\mathbb{P}_X$ . Die mgf ist hingegen mit der LAPLACETransformation verwandt.
- Existiert die mgf in einer Umgebung der Null, dann ist sie dort holomorph (komplex differenzierbar) und kann daher in die komplexe Ebene fortgesetzt werden (siehe Funktionentheorie). Dies führt auf die charakteristische Funktion.
- Ist  $X$  eine  $\mathbb{N}_0$ -wertige Funktion, so gilt  $\varphi_X(u) = \psi_X(e^{iu})$ .

**Beispiel 7.9**

Sei  $X \sim \text{Gamma}(\lambda, r)$ . Dann folgt mittels Rechnung wie in Beispiel 7.2 und dem Integralsatz von CAUCHY:

$$\varphi_X(u) = \left(\frac{\lambda}{\lambda - iu}\right)^r \quad (u \in \mathbb{R})$$

**Proposition 7.10 (Rechenregeln)**

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum,  $X, Y: \Omega \rightarrow \mathbb{R}^d$  mit charakteristischer Funktion  $\varphi_X, \varphi_Y$ . Dann

- (1)  $\varphi_X(0) = 1$
- (2) Seien  $A \in \mathbb{R}^{n \times n}$ ,  $b \in \mathbb{R}^n$ . Dann gilt

$$\varphi_{AX+b}(u) = e^{i \cdot \langle u, b \rangle} \varphi_X(A^\top u)$$

- (3) Wenn  $X \perp\!\!\!\perp Y$ , dann folgt  $\varphi_{X+Y}(u) = \varphi_X(u) \cdot \varphi_Y(u)$  ( $u \in \mathbb{R}^d$ )

*Beweis.* Analog zu Proposition 7.4. □

**Proposition 7.11**

Sei  $X \sim \mathcal{N}(\mu, \sigma^2)$ , dann gilt

$$\varphi_X(u) = e^{i\mu u - \frac{\sigma^2 u^2}{2}} \quad (u \in \mathbb{R})$$

und insbesondere

$$\varphi_{\mathcal{N}(0,1)}(u) = e^{-\frac{u^2}{2}} \quad (u \in \mathbb{R})$$

*Beweis.* Betrachte die Standardnormalverteilung:

$$\varphi(u) := \varphi_{\mathcal{N}(0,1)}(u) = \int_{\mathbb{R}} e^{iux} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Mit dem Differenzierbarkeitslemma für Parameterintegrale ( $\nearrow$  Schilling MINT, Satz 12.2)

$$\begin{aligned} \varphi'(u) &= \int_{\mathbb{R}} \left( \frac{d}{du} e^{iux} \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} ixe^{iux} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (-i)e^{iux} \left( \frac{d}{dx} e^{-\frac{x^2}{2}} \right) dx \\ &\stackrel{\text{P.I.}}{=} \underbrace{\frac{-i}{\sqrt{2\pi}} \left[ e^{iux} e^{-\frac{x^2}{2}} \right]_{x=-\infty}^{\infty}}_{=0} + \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} i \left( \frac{d}{dx} e^{iux} \right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} i \cdot iue^{iux} e^{-\frac{x^2}{2}} dx \\ &= -u \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iux} e^{-\frac{x^2}{2}} dx \\ &= -u\varphi(u) \end{aligned}$$

Die DGL  $\varphi'(u) = -u \cdot \varphi(u)$  besitzt die Lösung

$$\varphi(u) = \varphi(0) \cdot e^{-\frac{u^2}{2}}$$

mit  $\varphi(0) = 1$  nach Proposition 7.10, also folgt  $\varphi(u) = e^{-\frac{u^2}{2}}$ . Für  $X \sim \mathcal{N}(\mu, \sigma^2)$  gilt  $Z := \frac{X-\mu}{\sqrt{\sigma^2}} \sim \mathcal{N}(0, 1)$  ( $\nearrow$  HA 9.1) und nach Proposition 7.10 folgt

$$\varphi_X(u) = \varphi_{\sigma Z + \mu}(u) = e^{i\mu u} \cdot \varphi_Z(\sigma u) = e^{i\mu u} \cdot e^{-\frac{\sigma^2 u^2}{2}} \quad \square$$

Die Charakteristische Funktion charakterisiert eine Verteilung eindeutig:

### Proposition 7.12

Seien  $(\Omega, \mathcal{F}, \mathbb{P}), (\Omega', \mathcal{F}', \mathbb{P}')$  Wahrscheinlichkeitsräume und  $X: \Omega \rightarrow \mathbb{R}^d$  sowie  $Y: \Omega' \rightarrow \mathbb{R}^d$  Zufallsvariablen mit charakteristischen Funktionen  $\varphi_X, \varphi_Y$ . Dann:

$$X \stackrel{d}{=} Y \Leftrightarrow \varphi_X(u) = \varphi_Y(u) \quad \forall u \in \mathbb{R}^d$$

Für den Beweis benötigen wir:

### Lemma 7.13

Seien  $(\Omega, \mathcal{F}, \mathbb{P}), (\Omega', \mathcal{F}', \mathbb{P}')$  Wahrscheinlichkeitsräume und  $X: \Omega \rightarrow \mathbb{R}^d$  sowie  $Y: \Omega' \rightarrow \mathbb{R}^d$  Zufallsvariablen. Dann gilt  $X \stackrel{d}{=} Y$  genau dann, wenn

$$\mathbb{E}[f(X)] = \mathbb{E}'[f(Y)] \quad \forall f \in C_c(\mathbb{R}^d)$$

Dabei ist  $\mathbb{E}'[f(Y)]$  der Erwartungswert bezüglich  $\mathbb{P}'$  und  $C_c$  ist die Menge der stetigen Funktionen mit kompakten Träger.

*Beweis.* ( $\Rightarrow$ ) klar

( $\Leftarrow$ ) Es genügt zu zeigen

$$\mathbb{E}[\mathbb{1}_K(X)] = \mathbb{P}(X \in K) = \mathbb{P}'(Y \in K) = \mathbb{E}'[\mathbb{1}_K(Y)] \quad (\star)$$

für alle  $K \subseteq \mathbb{R}^d$  kompakt, denn die kompakten Mengen sind ein  $\cap$ -stabiler Erzeuger von  $\mathcal{B}(\mathbb{R}^d)$  und es gibt eine aufsteigende Folge  $K_n \nearrow \mathbb{R}^d$ . Die Indikatoren  $\mathbb{1}_K$  können wir durch  $C_c$ -Funktionen approximieren: Sei

$$\begin{aligned} \text{dist}(x, A) &:= \inf_{y \in A} |x - y| \quad \text{mit} \quad A \subset \mathbb{R}^d \\ f_n(x) &:= \frac{\text{dist}(x, U_n^c)}{\text{dist}(x, U_n^c) + \text{dist}(x, K)} \quad \text{mit} \quad U_n = \{y \in \mathbb{R}^d : \text{dist}(y, K) < 1/n\} \end{aligned}$$

Dann ist  $f_n \in C_c$  mit  $f_n \searrow \mathbb{1}_K$ . Mit monotoner Konvergenz folgt  $(\star)$  aus der Voraussetzung.  $\square$

Folgende Proposition ist analog zu Proposition 7.5.

**Proposition 7.14**

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $X: \Omega \rightarrow \mathbb{R}^d$  eine Zufallsvariable mit charakteristischer Funktion  $\varphi_X$ . Wenn  $\mathbb{E}[\|X\|^n] < \infty$  für ein  $n \in \mathbb{N}$ , dann existiert

$$\partial_u^\alpha \varphi_X(u) = \frac{\partial^{|\alpha|}}{\partial_{u_1}^{\alpha_1} \dots \partial_{u_d}^{\alpha_d}} \varphi_X(u) \quad \forall \alpha \in \mathbb{N}_0^d \text{ mit } |\alpha| = \sum_{i=1}^d \alpha_i \leq n$$

Zudem ist  $\partial_u^\alpha \varphi_X(u)$  stetig und es gilt

$$\mathbb{E}[X^\alpha] = i^{-|\alpha|} \partial_u^\alpha \varphi_X(0)$$

*Beweis.* Da  $\|x^\alpha\| \leq \|x\|^{|\alpha|}$  für  $x \in \mathbb{R}^d, \alpha \in \mathbb{N}_0^d$  folgt aus der Annahme  $\mathbb{E}[\|X^\alpha\|] < \infty$ . Damit

$$\begin{aligned} \partial_u^\alpha \varphi_X(u) &= \partial_u^\alpha \int_{\mathbb{R}^d} e^{i\langle u, x \rangle} \mathbb{P}(X \in dx) \\ &\stackrel{(\star)}{=} \int_{\mathbb{R}^d} \partial_u^\alpha \int_{\mathbb{R}^d} e^{i\langle u, x \rangle} \mathbb{P}(X \in dx) \\ &= \int_{\mathbb{R}^d} (ix)^\alpha e^{i\langle u, x \rangle} \mathbb{P}(X \in dx) \\ &= i^{|\alpha|} \int_{\mathbb{R}^d} x e^{i\langle u, x \rangle} \mathbb{P}(X \in dx) \end{aligned}$$

so dass die Ableitung existiert und für  $u = 0$  die Momentenformel folgt. In  $(\star)$  haben wir  $|\alpha|$ -mal das Differenzierbarkeitslemma ( $\nearrow$  MINT Satz 12.1) verwendet, wobei  $\partial_u^\alpha e^{i\langle u, x \rangle}$  integrierbar ist, da

$$\left\| \partial_u^\alpha e^{i\langle u, x \rangle} \right\| = \left\| (ix)^\alpha e^{i\langle u, x \rangle} \right\| = \|(ix)^\alpha\| = \|x^\alpha\| \quad \square$$