

# Einführung in die Stochastik

16. August 2017

## Inhaltsverzeichnis

<b>1 Grundlagen</b>	<b>2</b>
2. Urnenmodelle mit Reihenfolge, mit Zurücklegen . . . . .	9
2.A Urnenmodelle, ohne Reihenfolge, mit Zurücklegen . . . . .	10
<b>3 Bedingte Wahrscheinlichkeit, Unabhängigkeit</b>	<b>14</b>
3.A Galton-Watson-Prozesse und Wahrscheinlichkeitsbäume . . . . .	15
<b>4 Erwartungswert und Varianz</b>	<b>25</b>
4.A Erzeugende Funktionen . . . . .	33
<b>5 Grenzwertsätze</b>	<b>35</b>
<b>6 Grundzüge der Statistik</b>	<b>43</b>
6.A Fragestellung und Modellbildung . . . . .	43
6.B Maximum-Likelihood-Schätzer . . . . .	45
6.C Tests . . . . .	49
<b>7 Markovketten</b>	<b>51</b>

# 1 Grundlagen

**Erinnerung 1.1.** Naive Grundidee der Modellierung des Zufalls:

Konzept	mathematisches Objekt	Symbol
„Alle denkbaren Ergebnisse eines zufälligen Geschehens“	Menge	$\Omega$
Wahrscheinlichkeit, dass $\omega \in \Omega$ beobachtet wird	Abbildung $\Omega \rightarrow [0, 1]$	
Alle denkbaren Ja-Nein-Fragen, die zum zufälligen Geschehen gestellt werden können	Potenzmenge von $\Omega$	$\mathcal{P}(\Omega)$
Wahrscheinlichkeit, dass die zu $a \in \mathcal{P}(\Omega)$ gehörige Frage mit „ja“ beantwortet wird.	Abbildung $\mathcal{P}(\Omega) \rightarrow [0, 1]$	$\mathbb{P}$

**Beispiel 1.2.1** (6-Seitiger Würfel).  $\Omega = \{1, 2, 3, 4, 5, 6\}$  und  $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$ .

Ein Beispiel für eine Ja-Nein-Frage: „Ist die gewürfelte Zahl durch 3 teilbar?“ dann ist  $A \in \mathcal{P}(\Omega) : A = \{3, 6\}$  und  $\mathbb{P}(A) = \frac{1}{3}$ .

**Beispiel 1.2.2.** Speziell zufällige natürliche Zahl:  $\Omega = \mathbb{N}$ ,  $p(1) = \frac{1}{2}$ ,  $p(2) = \frac{1}{4}$ , ...,  $p(n) = 2^{-n}$ .

Dann gilt  $\sum_{\omega=1}^{\infty} p(\omega) = 1$ .

a) Ja-Nein-Frage: „Ist die Zahl gerade?“

Zugelassene Menge:  $A = \{2, 3, 6, \dots\}$

$$\mathbb{P}(A) = \sum_{j=1}^{\infty} p(2j) = \sum_{j=1}^{\infty} 2^{-2j} = \frac{1}{1 - \frac{1}{4}} - 1 = \frac{1}{3}$$

b) Ja-Nein-Frage: „Ist die Zahl Primzahl?“

Zugelassene Menge:  $B = \{n \in \mathbb{N} \mid n \text{ ist prim}\}$

$$\mathbb{P}(B) = \sum_{j \in B} p(j) = ???$$

$$\text{Abschätzung } \mathbb{P}(B) \leq 1 - \mathbb{P}(A) + \mathbb{P}(\{2\}) \leq \frac{2}{3} + \frac{1}{4}$$

**Definition 1.3.** Sei  $\Omega$  eine abzählbare Menge.

Eine Abbildung  $p : \Omega \rightarrow [0, 1]$  mit  $\sum_{\omega \in \Omega} p(\omega) = 1$  heißt **Zähldichte** oder **Wahrscheinlichkeitsdichte** auf  $\Omega$ .

**Definition 1.4.** Man nennt dann  $\Omega$  den „Ergebnisraum“, die „Grundmenge“ oder „Grundgesamtheit“.

Ein spezielles  $A \in \mathcal{P}(\Omega)$  nennt man „Ergebnis“ und falls  $A = \{\omega\}$  **Elementarereignis**.

**Definition 1.5.** Sei  $\Omega$  eine abzählbare Menge und  $\mathcal{P}(\Omega)$  die Potenzmenge,  $p$  sei eine Zähldichte.

Dann heißt die Abbildung

$$\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1], A \mapsto \sum_{\omega \in A} p(\omega)$$

das von  $p$  erzeugte **Wahrscheinlichkeitsmaß** (kurz W-Maß).

*Bemerkung 1.6.* Beachte:  $p$  wird in der Notation unterdrückt. Alternativ schreibe  $\mathbb{P}_p$ .

Außerdem: Statt  $\mathbb{P}(\{\omega\})$  wird oft  $\mathbb{P}(\omega)$  geschrieben.

**Lemma 1.7.** Sei  $p$  eine Zähldichte auf  $\Omega$ . Das von  $p$  erzeugte Wahrscheinlichkeitsmaß hat folgende Eigenschaften:

- a)  $\mathbb{P}(\Omega) = 1$
- b) Falls  $(A_n)_{n \in \mathbb{N}}$  eine Folge von paarweise disjunkten Ereignissen ist, dann ist  $\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ .

*Beweis.* Sei  $p$  Zähldichte auf  $\Omega$ .

- a) Nach Definition ?? :  $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} p(\omega) = 1$

- b) Es gilt:

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}(\{\omega \in \Omega : \exists n \in \mathbb{N} \text{ mit } \omega \in A_n\}) = \sum_{\{\omega \in \Omega : \exists n \in \mathbb{N} \text{ mit } \omega \in A_n\}} p(\omega) \quad (1)$$

Sei nun  $N(\omega) = \sum_{n=1}^{\infty} \chi_{A_n}(\omega)$  „Anzahl der  $A_n$  die  $\omega$  enthalten“. Dann ist

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \sum_{\omega \in A_n} p(\omega) = \sum_{\{\omega \in \Omega : \exists n \in \mathbb{N} \text{ mit } \omega \in A_n\}} p(\omega) N(\omega) \quad (2)$$

Da aber die  $A_n$  paarweise disjunkt sind ist  $N(\omega) = 1$  für alle  $\omega \in \{\omega \in \Omega : \exists n \in \mathbb{N} \text{ mit } \omega \in A_n\}$ , ist ist (1)=(2).  $\square$

**Definition 1.8.** Für  $A \subseteq \Omega$  heißt

$$\chi_A : \Omega \rightarrow \{0, 1\}, \quad \omega \mapsto \begin{cases} 1 & \omega \in A \\ 0 & \text{sonst} \end{cases}$$

die **Charakteristische Funktion** oder **Indikator** von  $A$ . Man schreibt auch  $\mathbb{1}_A$ .

*Beispiel 1.9.* Sei  $\Omega = \mathbb{N}$ ,  $(a_n)_{n \in \mathbb{N}}$  eine Folge mit  $a_n \geq 0$ ,  $a := \sum_{n=1}^{\infty} a_n \leq \infty$  und  $a > 0$ .

Dann ist  $p : \Omega \rightarrow [0, 1]$ ,  $n \mapsto p(n) := \frac{a_n}{a}$  eine Zähldichte.

Es gilt sogar die Isomorphie:

$$\{\text{Zähldichte auf } \mathbb{N}\} \cong \{\text{Nicht negative Folgen mit } \sum_{n=1}^{\infty} a_n = 1\} \cong \{\text{Nichtnegative summierbare Folgen } \neq 0\}$$

*Bemerkung 1.10* (Notation). Sei  $\Omega$  eine Menge.  $|\Omega| \leq \infty$  bezeichne die Anzahl der Elemente von  $\Omega$  „Mächtigkeit der Menge“.

*Beispiel 1.11.1* (Einfache Irrfahrt). Dimension  $d$ ,  $N$  Schritte.  $\Omega = \{(x_0, x_1, \dots, x_N) : X_j \in \mathbb{Z}^d \forall j, x_0 = 0, |x_{j+1} - x_j| = 1 \forall j\}$ .

Also ist  $|\Omega_N| = (2d)^N$ , Setze  $p(\omega = \frac{1}{(2d)^N}) \forall \omega \in \Omega$ .

Man kann nun folgende Fragestellungen formulieren:

- a)  $A_N := \{(x_1, \dots, x_N)\} \in \Omega_N : \exists j > 0 \text{ mit } x_j = 0\}$ . (“Rückkehr zum Startpunkt”).

Es ist klar, dass  $\mathbb{P}(A_N) \geq \frac{1}{2d} > 0$ , falls  $N \geq 2$ .

Es ist leicht zu zeigen, dass  $N \mapsto \mathbb{P}(A_N)$  wächst monoton.

**Knifflig:** Was ist  $\lim_{N \rightarrow \infty} ? < 1? = 1?$

**Antwort:**  $= 1$  für  $d \leq 2$ ,  $< 1$  für  $d \geq 3$ .

- b)  $B_n, \alpha := \{\omega = (x_0, x_1, \dots, x_N) \in \Omega_N : |x_N| \geq N^\alpha\}$  für  $0 < \alpha \leq 1$

**Frage:**  $\lim_{N \rightarrow \infty} \mathbb{P}(B_n, \alpha)?$

**Antwort:** 0, falls  $\alpha > \frac{1}{2}$

1, falls  $\alpha < \frac{1}{2}$

Für  $\alpha = \frac{1}{2}$  gilt

$$\lim_{N \rightarrow \infty} \mathbb{P}(B_n, \alpha) = \frac{V_k(d)}{(2\pi)^{\frac{d}{2}}} \int_1^\infty r^{d-1} \exp\left(-\frac{1}{2}r^2\right) dr$$

(dabei ist  $V_k(d)$  das Volumen der  $d$ -Dimensionalen Einheitskugel).

*Beispiel 1.11.2* (Selbstvermeidende Irrfahrt). Dimension  $d$ ,  $N$  Schritte.

- a)  $\Omega_N^0 = \{(x_0, x_1, \dots, x_N) \in \Omega_N : x_i \neq x_j \text{ falls } i \neq j\}$  Dann gilt für die Anzahl der Pfade:

$$|\Omega_N^0| = \begin{cases} 2, & \text{falls } d = 1 \\ ??, & \text{falls } d > 1 \end{cases}$$

und es ist  $p(\omega) = \frac{1}{|\Omega_N^0| \forall \omega \in \Omega_N^0}$ .

- b) Wie in a)2.

**Frage** Was ist  $\lim_{N \rightarrow \infty} \mathbb{P}(B_{N, \alpha}^0)$ .

**Bekannt**  $\exists \alpha_c > 0$  mit

$$\lim_{N \rightarrow \infty} \mathbb{P}(\mathbb{B}_{\times, \alpha}^\times) = \begin{cases} 0, & \text{falls } \alpha > \alpha_c \\ 1, & \text{falls } \alpha < \alpha_c \end{cases}$$

**Bekannte Werte:**  $d = 1 \quad \alpha_c = 1$

$d = 2 \quad \alpha_c = \frac{3}{4}$ , falls SLE-Conjecture stimmt

$d = 3 \quad \alpha_c \approx 0,5876$  (Numerik)

$d \geq 4 \quad \alpha_c = \frac{1}{2}$

*Beispiel 1.12.* Auswählen einer Zufälligen reellen Zahl in  $[0, 1]$ , alle Zahlen sollen die gleiche Wahrscheinlichkeit haben:

$[0, 1]$  ist nicht endlich, also ist Gleiche Wahrscheinlichkeit für alle Zahlen unmöglich.

$[0, 1]$  ist nicht abzählbar, also scheitert der bisherige Ansatz mit der Zähldichte.

**Ein möglicher Ausweg** Definiere  $\mathbb{P}([a, b]) = \mathbb{P}((a, b)) = \mathbb{P}([a, b)) = \mathbb{P}((a, b])$ .  
Die Erweiterung, sodass  $\forall A \in \mathcal{P}([0, 1])$   $\mathbb{P}(A)$  definiert ist, ist nicht möglich.

**Lösung** Definiere  $\mathbb{P}$  nicht auf allen Mengen  $\mathcal{P}([0, 1])$ .

**Definition 1.13.** Sei  $\Omega$  eine nichtleere Menge.  
Ein Mengensystem  $\mathcal{F} \subset \mathcal{P}(\Omega)$  heißt  **$\sigma$ -Algebra**, falls

- a)  $\Omega \in \mathcal{F}$
- b) Falls  $A \in \mathcal{F}$ , dann auch  $A^C \in \mathcal{F}$ .
- c) Falls  $A_1, A_2, \dots \in \mathcal{F}$ , dann auch  $\bigcap A_i \in \mathcal{F}$ .

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

*Bemerkung 1.14.*  $\mathcal{F}$  ist “die Menge aller Teilmengen von  $\Omega$ , für die die zugehörige Ja-Nein-Frage beantwortbar ist”.

Daher meint

- a) “Ist  $\omega \in \Omega$ ” muss beantwortbar sein.
- b) Falls “Ist  $\omega \in A$ ?” beantwortbar, so ist auch “Ist  $\omega \notin A$ ?” beantwortbar.
- c) Falls “Ist  $\omega \in A_i$ ?” beantwortbar für alle  $i$ , dann ist auch “Ist  $\omega$  in irgendeinem  $A_i$ ?” beantwortbar.

*Beispiel 1.15.* Sei  $\Omega = [0, 1)$ , dann ist

- a)  $\mathcal{F}_0 := \{\emptyset, \Omega\}$
- b)  $\mathcal{F}_1 := \{\emptyset, [0, \frac{1}{3}), [\frac{1}{3}, 1), \Omega\}$ .  
Die Frage “Ist  $\omega \geq \frac{1}{2}$ ” ist hier nicht beantwortbar!
- c)  $A_{j,n} := [\frac{j}{n}, \frac{j+1}{n})$ ,  $n$  ist fest,  $j \geq n$ .  
 $\mathcal{F}_2 = \{\bigcup_{k=1}^n B_{k,n} : B_{k,n} \in \{\emptyset, A_{k,n}\}\}$
- d)  $\mathcal{F}_3 = \mathcal{P}(\Omega)$  ist ebenfalls eine  $\sigma$ -Algebra.

**Satz 1.16.** Sei  $\mathcal{G} \subset \mathcal{P}(\Omega)$  ein Mengensystem. Sei  $\Sigma := \{\mathcal{A} \subset \mathcal{P}(\Omega) : \mathcal{A} \text{ ist } \sigma\text{-Algebra und } \mathcal{G} \subset \mathcal{A}\}$ .

Dann ist auch  $\bigcap_{\mathcal{A} \in \Sigma} \mathcal{A}$  eine  $\sigma$ -Algebra.

**Definition 1.16.**  $\sigma(\mathcal{G}) := \bigcap_{\mathcal{A} \in \Sigma} \mathcal{A}$  heißt **die von  $\mathcal{G}$  erzeugte  $\sigma$ -Algebra**.

**Definition 1.17.** Sei  $\Omega = \mathbb{R}$ ,  $\mathcal{G} := \{[a, b] : a, b \in \mathbb{R}, a < b\}$ .

$\mathcal{B} := \sigma(\mathcal{G})$  heißt **Borel- $\sigma$ -Algebra**.

*Bemerkung 1.18.* a)  $\mathcal{B}$  enthält alle offenen Mengen, alle abgeschlossenen Mengen und alle halboffenen Intervalle.

- b)  $\mathcal{B} \subsetneq \mathcal{P}(\Omega)$ .
- c)  $\mathcal{B}$  kann nicht abzählbar konstruiert werden.
- d)  $\mathcal{B} = \sigma(\{(-\infty, c] : c \in \mathbb{R}\})$ .

e) Falls  $\Omega_0 \subset \mathbb{R}$ ,  $\Omega_0 \neq \emptyset$ , dann ist

$$\mathcal{B}_{\Omega_0} := \{A \cap \Omega_0 : A \in \mathcal{B}(\mathbb{R})\}$$

eine  $\sigma$ -Algebra, die **Einschränkung** von  $\mathcal{B}$  auf  $\Omega_0$ .

**Definition 1.19.** Seien  $E_1, E_2, \dots, E_N$  Mengen,  $N \leq \infty$ .  
 $\mathcal{E}_i$  seien  $\sigma$ -Algebren auf  $E_i$  und es sei

$$\Omega = \prod_{i=1}^N E_i = \{(e_1, \dots, e_N) : e_i \in E_i \forall i \leq N\}$$

Eine Menge der Form

$$A_{j, B_j} = \{(e_1, \dots, e_N) : e_j \in B_j, \text{ andere } e_k \text{ beliebig}\}$$

mit  $B_j \in \mathcal{E}_j, j \leq N$  heißt **Zylindermenge**.

**Definition 1.19.** Die  $\sigma$ -Algebra in  $\Omega$  die von allen Zylindermengen Erzeugt wird heißt **Produkt- $\sigma$ -Algebra**. Man nennt  $\mathcal{Z}$  das System der Zylindermenge und  $\mathcal{E}_1 \otimes \mathcal{E}_2 \otimes \dots \otimes \mathcal{E}_N := \sigma(\mathcal{Z})$ .

**Definition 1.20.** Sei  $\Omega, \mathcal{F}$  ein messabere Raum. Eine Abbildung  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$  heißt **Wahrscheinlichkeitsmaß** (W-Maß) auf  $\mathcal{F}$  (teilweise auch „auf  $\Omega$ “), falls

- a)  $\mathbb{P}(\Omega) = 1$
- b) Für paarweise disjunkte  $A_n \in \mathcal{F}$  gilt  $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$  („ $\sigma$ -Additivität“).

Dann heißt  $\Omega, \mathcal{F}, \mathbb{P}$  **Wahrscheinlichkeitsraum**.

*Beispiel 1.21.* Sei  $(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \delta_x)$ . Dabei ist  $\mathbb{P}(A) = \begin{cases} 1; & \text{falls } x \in A \\ 0, & \text{sonst.} \end{cases}$

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

Es Modelliert ein „Zufalls“-Experiment, welches sicher  $x$  ergibt.

**Satz 1.22** (Eigenschaften von Wahrscheinlichkeitsmaßen). *Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $A, B, A_1, A_2, \dots \in \mathcal{F}$ . Dann ist*

- a)  $\mathbb{P}(\emptyset) = 0$
- b)  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$   
*insbesondere  $\mathbb{P}(A) + \mathbb{P}(A^C) = 1$*
- c) Falls  $A \subseteq B$ , dann ist  $\mathbb{P}(A) \leq \mathbb{P}(B)$  („Monotonie“)
- d)  $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$  ( $\sigma$ -Subadditivität)
- e) Falls  $A_n \nearrow A_n$  (d.h.  $A_1 \subseteq A_2 \subseteq \dots$  und  $\bigcap_{n=1}^{\infty} A_n = A$ ),  
dann ist  $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}(A)$  („ $\sigma$ -Stetigkeit“)  
Falls  $A_n \searrow A_n$  (d.h.  $A_1 \supseteq A_2 \supseteq \dots$  und  $\bigcup_{n=1}^{\infty} A_n = A$ ),  
dann ist  $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}(A)$

*Beweis.* a)  $\mathbb{P}(\emptyset) = \mathbb{P}(\bigcup_{n=1}^{\infty} \emptyset) = \sum_{n=1}^{\infty} \mathbb{P}(\emptyset)$ . Also  $\mathbb{P}(\emptyset) = 0$

- b) Falls  $A \cap B = \emptyset$ , dann ist  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(\emptyset) = \mathbb{P}(A) + \mathbb{P}(B)$ .  
Sei nun  $A \cap B \neq \emptyset$ .

□

**Definition 1.27.** Die Abbildung  $\lambda : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, \infty], A \mapsto \int_{\mathbb{R}^n} \chi_A(x) dx = \lambda(A)$  heißt **Lebesgue-Maß**

*Beispiel 1.28.* Sei  $\varrho : \mathbb{R}^n \rightarrow [0, \infty)$  Borel-messbar und  $\int \varrho(x) dx = 1$ .  
Dann ist die Abbildung  $\mathbb{P}_\varrho : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, 1], A \mapsto \int_A \varrho(x) dx = \int \chi_A(y) \varrho(x) dx$  ein Wahrscheinlichkeitsmaß.

**Definition 1.29.** Sei  $\varrho : \mathbb{R}^n \rightarrow [0, \infty)$  Borel-messbar und  $\int \varrho(x) dx = 1$ , (=1.28) dann heißt  $\varrho$  **Dichte** von  $\mathbb{P}_\varrho$ .

*Beispiel 1.30.* Sei  $\varrho$  eine Dichte,  $x \in \mathbb{R}$ . Dann hat  $\mathbb{P} := \frac{1}{3}\mathbb{P}_\varrho + \frac{2}{3}\delta_x$  keine Dichte (siehe ??)

*Bemerkung.* Wenn  $\varrho$  Dichte ist schreibt man auch  $\mathbb{P}_\varrho \equiv \varrho(x) dx$

**Definition 1.31.** Sei  $\Omega \in \mathcal{B}(\mathbb{R}^n)$  mit  $\lambda(\Omega) < \infty$ . Das Wahrscheinlichkeitsmaß auf  $\Omega$  mit Dichte  $\varrho(y) = \frac{1}{\lambda(\Omega)}$  heißt **Gleichverteilung** auf  $\Omega$ .  
Man fasst dann  $\tilde{\varrho}(x) = \frac{1}{\lambda(\Omega)} \chi_A \Omega(x)$  als Einbettung in den  $\mathbb{R}^n$  auf.

**Erinnerung** (Zufallsvariable). Der Begriff „Zufallsvariable“ ist historisch gewachsen. (Keine Variable einer Funktion).

**Problemstellung** Von einem komplizierten Zufälligen Geschehen will man nur gewisse Aspekte betrachten.

*Beispiel 1.32.1* (2 mal Würfeln, Würfelsumme). Sei  $\Omega : \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 4, 5, 6\}$ .  
 $\omega = (\omega_1, \omega_2)$ . Der zu betrachtende Aspekt:  $S(\omega_1, \omega_2) = \omega_1 + \omega_2 \in \{2, 3, \dots, 12\} \neq \Omega$ .

*Beispiel 1.32.2.* Sei  $\Omega = \Omega_N$  (siehe ??, einfache Irrfahrt).  
 $\omega = (\omega_1, \omega_2, \dots, \omega_N) \in (\mathbb{Z}^d)^N$ .

**Aspekt 1** Position nach  $N$  Schritten.

Modell:  $X_N((\omega_1, \dots, \omega_N)) = \omega_N \in \mathbb{Z}^d \neq \Omega$

**Aspekt 2** Maximaler Abstand vom Ursprung bis zum Schritt  $N$ .

Modell  $M_N((\omega_1, \dots, \omega_N)) = \max\{|\omega_j|, j \leq N\}$ .

*Beispiel 1.32.3.* Sei  $\Omega = [0, 1], \mathcal{F} \in \mathcal{B}([0, 1]), \omega = x \in [0, 1]$ .

**Aspekt 1** Erste Ziffer nach dem Komma?

Modell:  $y_1(x) = \lfloor 10x \rfloor$

**Aspekt 2** Fläche des Quadrates mit Kantenlänge  $x$

Modell:  $Q(x) = x^2 \in [0, 1]$ .

Fazit: Modellierung durch Abbildungen.

**Definition 1.33.** Seine  $(\Omega, \mathcal{F})$  und  $(\Omega', \mathcal{F}')$  Ereignisräume.

eine Abbildung  $X : \Omega \rightarrow \Omega'$  heißt **Zufallsvariable** (ZV) [oder messbare Abbildung, zufälliges Element von  $\Omega'$ ], falls gilt:

$\forall A' \in \mathcal{F}'$  ist  $X^{-1}(A') \in \mathcal{F}$ .

Hierbei ist  $X^{-1}$  das Urbild von  $A'$  unter  $X$ .

*Bemerkung 1.34.* Die Urbild-Abbildung bilde Mengen in  $\mathcal{F}'$  (d.h. erlaube Ja-Nein-Fragen) auf Mengen in  $\mathcal{P}(\Omega)$  (d.h. Ja-Nein-Fragen) ab.

*Beispiel.* In 1.32.1 ?? ist  $S^{-1}(\{4\}) = \{(1, 3), (2, 2), (3, 1)\}$ .  
In 1.32.1 ?? ist  $Y_1^{-1}(\{3, 7\}) = [0, 3, 0, 4] \cup [0, 7, 0, 8]$

*Bemerkung.* Die Bedingung (\*) bedeutet, dass für alle durch  $A \in \mathcal{F}'$  erzeugte erlaubten Ja-Nein-Fragen auch die Frage „Liegt  $X(\omega)$  in  $A'$ ?“ erlaubt ist.

*Bemerkung.* Oft nimmt man nur  $\Omega$  und  $(\Omega', \mathcal{F}')$  als gegeben.  
Dann ist  $X^{-1}(\mathcal{F}') := \{X^{-1}(A') \mid A' \in \mathcal{F}'\}$  die von  $X$  erzeugte  $\sigma$ -Algebra.

*Bemerkung.* Falls  $\mathcal{F} = \mathcal{P}(\Omega)$ , dann ist jede Abbildung eine Zufallsvariable.

**Lemma 1.35.** *Sei  $(\Omega, \mathcal{F})$  und  $(\Omega', \mathcal{F}')$  Ereignisräume,  $X : \Omega \rightarrow \Omega'$  und sei  $\mathcal{G}'$  ein Mengensystem mit  $\mathcal{F}' = \sigma(\mathcal{G}')$ . Dann ist  $X$  genau dann Zufallsvariable, wenn  $X^{-1}(A') \in \mathcal{F} \forall A' \in \mathcal{G}'$ .*

*Beweis.* „ $\Rightarrow$ “ ist klar, da  $\mathcal{G}' \subset \mathcal{F}'$

„ $\Leftarrow$ “ Sei  $\mathcal{A}' := \{A' \in \Omega' \mid X^{-1}(A') \in \mathcal{F}\}$  ist eine  $\sigma$ -Algebra und  $\mathcal{A}' \supset \mathcal{G}'$  nach Annahme.

Daher ist  $\mathcal{F}' = \sigma(\mathcal{G}') \subset \mathcal{A}'$ , sodass  $X^{-1}(A') \in \mathcal{F} \forall A' \in \mathcal{F}'$ . □

*Beispiel 1.36.1.* Sei  $(\Omega', \mathcal{F}') = (\mathbb{R}, \mathbb{R})$ . Nach 1.35 gilt:  
 $X : \Omega \rightarrow \Omega'$  ist genau dann Zufallsvariable, wenn  $X^{-1}((-\infty, c)) \in \mathcal{F} \forall c \in \mathbb{R}$   
Für  $\Omega' = \mathbb{R}$  heißt  $X$  **reelle Zufallsvariable**

*Beispiel 1.36.2.* Es ist  $\overline{\mathbb{R}} := [-\infty, \infty]$  mit  $\sigma$ -Algebra  $\mathcal{B}(\overline{\mathbb{R}}) = \sigma(\{[-\infty - c] : c \in \mathbb{R}\})$ .

Die Abbildung  $X : \Omega \rightarrow \overline{\mathbb{R}}$  ist genau dann Zufallsvariable, wenn  $X^{-1}([-\infty, c]) \in \mathcal{F} \forall c$ .

Dann heißt  $X$  **numerische Zufallsvariable**

**Theorem 1.37.** *Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum,  $(\Omega', \mathcal{F}')$  ein Ereignisraum,  $X : \Omega \rightarrow \Omega'$  eine Zufallsvariable.*

*Dann ist die Abbildung*

$$\mathbb{P}' : \mathcal{F}' \rightarrow [0, 1], \quad A' \mapsto \mathbb{P}'(A') := \mathbb{P}(X^{-1}(A'))$$

*ein Wahrscheinlichkeitsmaß auf  $(\Omega', \mathcal{F}')$ .*

**Definition 1.38.**  $\mathbb{P}'$  heißt **Bildmaß von  $\mathbb{P}$  unter  $X$**  oder **Verteilung von  $X$  unter  $\mathbb{P}$** .

Man schreibt  $\mathbb{P}' = \mathbb{P} \circ X^{-1}$  oder  $\mathbb{P}' = \mathbb{P}_X$

*Beweis.* Da  $X$  eine Zufallsvariable ist, ist  $X^{-1}(A') \in \mathcal{F} \forall A' \in \mathcal{F}'$ , daher im Definitionsbereich von  $\mathbb{P}$ .

Also ist  $\mathbb{P}'$  wohldefiniert. Prüfe Definition 1.20 ??.

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



- b)  $A'_1, A'_2, \dots \in \mathcal{F}'$  seien paarweise disjunkt. Dann sind  $X^{-1}(A'_1), X^{-1}(A'_2), \dots$  auch paarweise disjunkt.

$$\begin{aligned}\mathbb{P}'\left(\bigcup_{i=1}^{\infty} A'_i\right) &= \mathbb{P}\left(X^{-1}\left(\bigcup_{i=1}^{\infty} A'_i\right)\right) \\ &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} X^{-1}(A'_i)\right) \\ &= \sum_{i=1}^{\infty} \mathbb{P}(X^{-1}(A'_i)) \\ &= \sum_{i=1}^{\infty} \mathbb{P}'(A'_i)\end{aligned}$$

□

**Definition 1.39.** Seien  $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$  und  $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$  Wahrscheinlichkeitsräume,  $X_1 : \Omega_1 \rightarrow \Omega'_1$ ,  $X_2 : \Omega_2 \rightarrow \Omega'_2$  Zufallsvariablen. Falls  $\mathbb{P}_1(X_1^{-1}(A')) = \mathbb{P}_2(X_2^{-1}(A')) \forall A' \in \mathcal{F}'$ , dann heißen  $X_1$  und  $X_2$  identisch verteilt.

*Bemerkung 1.40 (Notation).* Man schreibt oft:

- $\{X \in A'\}$  statt  $X^{-1}(A')$
- $\mathbb{P}(\{X \in A'\})$  oder  $\mathbb{P}(X \in A')$  statt  $\mathbb{P}(X^{-1}(A'))$
- $\mathbb{P}_X$  statt  $\mathbb{P} \circ X^{-1}$ .

## 2. Urnenmodelle mit Reihenfolge, mit Zurücklegen

*Beispiel 2.5.1* (Einmal ziehen:). Bsp: Tulpenzwiebeln,  $N$  Stück:  $k_r$  rote,  $k_g$  gelbe,  $k_o$  orange.

Dann ist  $\mathbb{P}(\text{rot}) = \frac{k_r}{N}, \dots$

Allgemein: Menge der Merkmale  $A = (a_1, \dots, a_m)$  mit  $p_i = \text{Bruchteil der „Kugeln“ mit Merkmal } a_i$ , sodass  $\mathbb{P}(\{a_i\}) = p_i$ .

Dann ist  $\P$  eine Zähldichte mit  $\sum_{i=1}^m p_i = 1$ .

*Beispiel 2.5.2* ( $N$ -Mal ziehen). Anzahl der Ziehungen  $N \in \mathbb{N}$ ,  $\Omega = A^N$ ,  $\P((a_{j_1}, \dots, a_{j_N})) := p_{j_1} \cdot \dots \cdot p_{j_N}$ , für  $j_1, \dots, j_N \in \{1, \dots, m\}$  mit  $p_1, \dots, p_m$  wie in 1.

Modelliert  $N$ -mal „unabhängig“ (siehe Kapitel 3) ziehen mit Zurücklegen.

**Definition 2.6.** Spezialfall von 2.:  $A = \{0, 1\}$ ,  $p_1 = p$ ,  $p_0 = 1 - p$ .

$1 \hat{=}$  „Erfolg“,  $p_1 \hat{=}$  „Erfolgswahrscheinlichkeit“.

Dann kann für  $\omega \in A^N$ ,  $\omega = (\omega_1, \dots, \omega_N)$  schreiben:

$$\mathbb{P}(\{\omega\}) = p^{\sum_{i=1}^N \omega_i} (1 - p)^{\sum_{i=1}^N (1 - \omega_i)}$$

Man nennt dies die **Bernoulli-Verteilung** für  $N$  Versuche mit Erfolgswahrscheinlichkeit  $p$ .

**Definition 2.7.** Die Zähldichte  $\bar{p} : A \times \dots \times A \rightarrow [0, 1]$ ,  $\bar{p}(a_1, \dots, a_n) \mapsto \mathbb{P}(\{a_1\}) \cdot \dots \cdot \mathbb{P}(\{a_n\}) = p_1 \cdot \dots \cdot p_n$  aus 2.5.2 heißt  $N$ -fache **Produktdichte** der Zähldichte  $p : A \rightarrow [0, 1]$ ,  $a_j \mapsto p(a_j) = p_j$ .

Das zugehörige Wahrscheinlichkeitsmaß heißt  $N$ -faches **Produktmaß** von  $\P$ .

## 2.A Urnenmodelle, ohne Reihenfolge, mit Zurücklegen

*Beispiel.* In Situation Das ist ein „Aspekt“ des Experiments, daher eine Zufallsvariable:

$$C : A^N \mapsto \mathbb{N}_0^m, \quad (a_{j_1}, \dots, a_{j_N}) \mapsto \left( \sum_{k=1}^N \chi_{\{a_1\}}(a_{j_k}), \sum_{k=1}^N \chi_{\{a_2\}}(a_{j_k}), \dots, \sum_{k=1}^N \chi_{\{a_m\}}(a_{j_k}) \right)$$

und

**Definition 2.8.** Sei  $\lambda > 0$ . Das Wahrscheinlichkeitsmaß auf  $\mathbb{N}_0$  mit Zähldichte  $p_\lambda(k) = e^{-\lambda} \frac{\lambda^k}{k!}$  heißt **Poisson-Verteilung** zum Parameter  $\lambda$ . Schreibe  $\text{Poi}(\{k\}) := \mathbb{P}_{p_k}(\{k\})$ .

*Bemerkung* (Bedeutung der Poisson-Verteilung). Modelliert die Anzahl der Erfolge, wenn  $N = \text{Anzahl der Versuche} \rightarrow \infty$ ,  $\mathbb{P}(\text{Erfolg pro Versuch}) \xrightarrow{N \rightarrow \infty} 0$ , sodass  $N\mathbb{P}(\text{Erfolg}) \xrightarrow{N \rightarrow \infty} \lambda$ .

**Definition 2.16.** Seien  $(a_n), (b_n) \subseteq \mathbb{R}^+$  Folgen, dann heißen  $(a_n), (b_n)$  zueinander **asymptotisch äquivalent**, falls  $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$ .

**Lemma 2.17.** Sei  $\sim$  die Asymptotische Äquivalenz. Dann gilt

a)  $\sim$  ist Äquivalenzrelation (reflexiv, symmetrisch, transitiv)

b) Falls  $\lim_{n \rightarrow \infty} a_n = a \leq \infty$  und  $(a_n) \sim (b_n)$ , dann ist  $\lim_{n \rightarrow \infty} b_n = a$ .

**Satz 2.15** (Poisson-Approximation). Sei  $\lambda > 0$  und  $(p_n) \subset [0, 1]$  eine Folge, mit  $\lim_{n \rightarrow \infty} np_n = \lambda$ . Sei  $B_{n,p_n}$  Binomialverteilung, dann gilt  $\forall k \in \mathbb{N}$ :

$$\lim_{n \rightarrow \infty} B_{n,p_n}(\{k\}) = \text{Poi}_\lambda(\{k\})$$

*Beweis.* Es gilt für den Binomialkoeffizienten:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{1}{k!} n(n-1)\dots(n-k+1) \sim \frac{1}{k!} n^k$$

denn für festes  $k \in \mathbb{N}$  ist  $(n-k) \sim n$  und  $\sim$  ist transitiv. Es folgt, dass

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

denn falls  $\lim_{n \rightarrow \infty} a_n = a > 0$ , dann  $\lim_{n \rightarrow \infty} \left(1 - \frac{1}{n} a_n\right)^n = e^{-a}$ . □

*Bemerkung 2.18.* Sei  $\alpha$  die „Rate des Eintreffens“ eines Ereignisses  $E$ , d.h.

$$\lim_{\delta \rightarrow 0} \mathbb{P}(E \text{ tritt in } [0, \delta] \text{ ein}) \frac{1}{\delta} = \alpha$$

Sei  $[0, t]$  der Beobachtungszeitraum. Teile  $[0, t]$  in  $n$  Intervalle der Länge  $t/n$ . Zusätzlich nehmen wir an

- (i) Ob  $E$  in einem gewissen Teilintervall auftritt beeinflusst die anderen Teilintervalle nicht.
- (ii) Zwei Ereignisse in einem Teilintervall sind verschwindend unwahrscheinlich.

Dann ist

$$\mathbb{P}(E \text{ passiert } k\text{-mal in } [0, t]) \stackrel{(ii)}{\approx} \mathbb{P}(E \text{ passiert in } k \text{ Intervallen}) \stackrel{(i)}{\approx} B_{n, p_n}(k) \xrightarrow{n \rightarrow \infty} \text{Poi}_{\alpha t}(k)$$

Also beschreibt der Parameter  $\lambda$  der Verteilung  $\text{Poi}_\lambda$  beschreibt „Rate mal Zeit“.

**Definition 2.19.** Für  $0 < p \leq 1$  heißt das Wahrscheinlichkeitsmaß  $\mathcal{G}_p$  auf  $\mathbb{N}_0$  mit Zähldichte  $g(k) = p(1-p)^k$  die **geometrische Verteilung** zur Erfolgswahrscheinlichkeit  $p$ .

$\mathcal{G}(\{k\})$  = „Wahrscheinlichkeit bei Erfolgswahrscheinlichkeit  $p$  pro Versuch genau  $k$  Fehlversuche vor dem Ersten Erfolg zu sehen“.

Allgemein: Da Wahrscheinlichkeitsmaß  $\overline{\mathcal{B}}_{r,p}$  auf  $\mathbb{N}_0$  mit  $r \in \mathbb{N}$  und  $0 < p \leq 1$  mit

$$\overline{\mathcal{B}}(\{k\}) = \frac{r(r+1)\dots(r+k-1)}{k!} p^r (1-p)^k$$

heißt **negative Binomialverteilung** und modelliert die Wahrscheinlichkeit vor dem  $r$ -ten Erfolg genau  $k$  Misserfolge zu haben.

**Definition 2.20.** Das Wahrscheinlichkeitsmaß  $\text{Exp}_\alpha$  auf  $(\mathbb{R}_0^+, \mathcal{B}(\mathbb{R}_0^+))$  mit Dichte  $\rho_\alpha(x) := \alpha e^{-\alpha x}$  heißt **Exponentialverteilung** zum Parameter  $\alpha > 0$

**Definition 2.20.** Das Wahrscheinlichkeitsmaß  $\text{Exp}_\alpha$  auf  $(\mathbb{R}_0^+, \mathcal{B}(\mathbb{R}_0^+))$  mit  $\alpha > 0, r \geq 1$  und Dichte

$$\gamma_{\alpha,r} := \frac{\alpha^r}{(r-1)!} x^{r-1} e^{-\alpha x} = \frac{\alpha^r}{\Gamma(r)} x^{r-1} e^{-\alpha x}$$

heißt **Gammaverteilung** zu den Parametern  $\alpha, r$ .

*Bemerkung 2.21.* a)  $\int \gamma_{\alpha,r}(x) dx = 1$

b)

$$\begin{aligned} \text{Exp}_\alpha([0, t]) &= \alpha \int_0^t e^{-\alpha x} dx = \alpha \left[ -\frac{1}{\alpha} e^{-\alpha x} \right]_0^t = 1 - e^{-\alpha t} = e^{-\alpha t} \sum_{k=1}^{\infty} \frac{\alpha^k}{k!} t^k \\ &= \text{Poi}_{\alpha t}(\{\mathbb{N}\}) = \text{Poi}_{\alpha t}(\text{„Midestens ein Erfolg vor Zeit } t \text{ bei Rate } \alpha\text{“}) \\ &= \mathbb{P}(\text{„Wartezeit auf ersten Erfolg } \leq t\text{“}) \end{aligned}$$

c) Es gilt

$$\mathbb{P}_{\gamma_{\alpha,r}}([0, t]) = \frac{\alpha^r}{(r-1)!} \int_0^t x^{r-1} e^{-\alpha x} dx = e^{-\alpha t} \sum_{k=r}^{\infty} \frac{(\alpha t)^k}{k!} \quad (\star)$$

denn

$$\begin{aligned}\frac{d}{dt}(\star) &= -\alpha e^{-\alpha t} \sum_{k=r}^{\infty} \frac{(\alpha t)^k}{k!} + e^{-\alpha t} \sum_{k=r}^{\infty} \frac{\alpha^k}{k!} k t^{k-1} \\ &= e^{-\alpha t} \left( -\alpha \sum_{k=r}^{\infty} \frac{(\alpha t)^k}{k!} + \alpha \sum_{k=r}^{\infty} \frac{(\alpha t)^k}{(k-1)!} \right) \\ &= e^{-\alpha t} \alpha \frac{(\alpha t)^{r-1}}{(r-1)!} = \frac{\alpha^r}{(r-1)!} \frac{d}{dt} \int_0^t x^{r-1} e^{-\alpha x} dx\end{aligned}$$

Somit ist also

$$(\star) = \text{Poi}_{\alpha,t}(\{r, r+1, \dots\}) = \mathbb{P}(\text{Wartezeit bis zum } r\text{-ten Erfolg ist } \leq t)$$

Daher heit  $\text{Exp}_{\alpha}$  und  $\mathbb{P}_{\gamma_{\alpha,r}}$  **Wartezeitverteilungen**.

**Definition 2.22.** Das Wahrscheinlichkeitsma  $\mathcal{N}$  auf  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  mit Dichte  $\phi_{0,1}(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$  heit **Standard-Normalverteilung**. Das Wahrscheinlichkeitsma  $\mathcal{N}_{m,v}$  auf  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  mit Dichte  $\phi_{m,v}(x) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(x-m)^2}{2v}\right)$  heit **Normalverteilung** (oder Gauverteilung/Gauma) mit Mittelwert  $m$  und Varianz  $v$ .

**Proposition 2.23.**  $\phi_{m,v}(x) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(x-m)^2}{2v}\right)$  ist eine Dichte auf  $\mathbb{R}$ .

*Beweis.* Es gilt

$$\begin{aligned}\frac{1}{\sqrt{2\pi v}} \int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2v}} dx &\stackrel{y=x-m}{=} \frac{1}{\sqrt{2\pi v}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2v}} dy \\ &\stackrel{z=\frac{y}{\sqrt{v}}}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz\end{aligned}$$

Berechne dann stattdessen

$$\begin{aligned}\left( \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right) &= \left( \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right) \left( \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \right) = \int_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx dy \\ &= \int_0^{\infty} \int_0^{2\pi} e^{-\frac{r^2}{2}} r d\theta dr \\ &= 2\pi \int_0^{\infty} e^{-\frac{r^2}{2}} r dr\end{aligned}$$

mit  $s = -\frac{r^2}{2}$

$$\begin{aligned}&= 2\pi \int_{-\infty}^0 e^s ds \\ &= 2\pi(e^0 - e^{-\infty}) = 2\pi\end{aligned}$$

Es folgt, dass  $\int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2v}} dx = \sqrt{2\pi}$ . □

*Bemerkung 2.24.* a) Bedeutung der Normalverteilung:

„universeller“ Grenzwert unabhngiger Summen in der „einzig sinnvollen“

Skalierung:  $\frac{1}{n} \sum_{i=1}^n X_i \rightarrow m$ .

Es folgt, dass  $\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - m) \sim \mathcal{N}(0, v)$ .

- b) Geometrische Bedeutung:  $\mathcal{N}_{0,v}$  ist die erste Koordinate der sogenannte „Gleichverteilung auf  $\mathbb{R}^\infty$ “.

**Satz 2.25.** Sei  $B_N(r) := \{x \in \mathbb{R}^N : |x| \leq r\}$  Kugel mit Radius  $r$  im  $\mathbb{R}^N$ , sei  $\mathbb{P}_N$  die Gleichverteilung auf  $B_N(r)$  und sei  $X_1 : B_N(r) \rightarrow \mathbb{R}, x = (x_1, \dots, x_N) \mapsto x_1$  die Zufallsvariable „Projektion auf die erste Koordinate“. Dann gilt

$$\lim_{N \rightarrow \infty} \mathbb{P}_{N,r_N}(a \leq X_1 \leq b) = \begin{cases} 0, & \text{falls } \lim_{n \rightarrow \infty} \frac{r_N}{\sqrt{N-1}} = \infty \\ \delta_0([a, b]), & \text{falls } \lim_{N \rightarrow \infty} \frac{r_N}{\sqrt{N-1}} = 0 \text{ und } a, b \neq 0 \\ \frac{1}{\sqrt{2\pi v}} \int_a^b e^{-\frac{x^2}{2v}} dx, & \text{falls } \lim_{N \rightarrow \infty} \frac{r_N}{\sqrt{N-1}} = \sqrt{v} > 0 \end{cases}$$

*Beweis.* (nur Fall  $a, b \neq 0$  im Fall 2) Integration einer Kugel durch Zerlegen in „Kreisscheiben“.

$$\begin{aligned} h_r(a, b) &:= \int_{\mathbb{R}^N} \chi_{x_1^2 + \dots + x_N^2 \leq r^2}(x) \chi_{\{a \leq x_1 \leq b\}} dx \\ &= \int_a^b dz \int_{\mathbb{R}^{N-1}} \chi_{x_1^2 + \dots + x_N^2 \leq r^2 - z^2}(x) dx \\ &= \int_a^b \int_{B_{N-1}(\sqrt{r^2 - z^2})} dx = \int_a^b (r^2 - z^2)^{N-1/2} dz \underbrace{\int_{B_{N-1}} (1) dx}_{=V_1(N-1)} \\ &= r^{N-1} V_1(N-1) \int_a^b \left(1 - \frac{z^2}{r^2}\right)^{\frac{N-1}{2}} dz \end{aligned}$$

Durch Substitution  $y = \frac{z}{r} \sqrt{N-1}$

$$= r^{N-1} V_1(N-1) \frac{r}{\sqrt{N-1}} \int_{\frac{\sqrt{N-1}}{r} a}^{\frac{\sqrt{N-1}}{r} b} \overbrace{\left(1 - \frac{1}{N-1} y^2\right)}^{=: f_N(y)} dy$$

Da  $\mathbb{P}_{N,r_N}(a \leq x_1 \leq b) = \frac{h_{r_N}(a,b)}{h_{r_N}(-r_N, r_N)}$  ist kürzen sich die Faktoren vor den Integralen.

Außerdem ist  $\lim_{N \rightarrow \infty} f_N(y) = \exp(-y^2)^{1/2} = e^{-\frac{1}{2}y^2}$  gleichmäßig auf Kompakta.

$$(\text{d.h. } \lim_{N \rightarrow \infty} \sup \left\{ \left| f_N(y) - e^{-\frac{y^2}{2}} \right| : |y| \leq C \right\} = 0)$$

Mit dieser Information kann man relativ leicht zeigen, dass für  $a_\infty = \lim_{N \rightarrow \infty} \frac{\sqrt{N-1}}{r_N} a$ ,

$$b_\infty = \lim_{N \rightarrow \infty} \frac{\sqrt{N-1}}{r_N} b \text{ gilt}$$

$$\lim_{N \rightarrow \infty} \int_{\frac{\sqrt{N-1}}{r_N} a}^{\frac{\sqrt{N-1}}{r_N} b} f_N(x) dy = \begin{cases} \int_{a_\infty}^{b_\infty} e^{-\frac{y^2}{2}} dy, & \text{falls } -\infty \leq a_\infty b_\infty \leq \infty \\ 0, & \text{falls } a_\infty = b_\infty \in \{-\infty, 0, \infty\} \end{cases}$$

Aus ?? für  $h_{r_N}(-r_N, r_N)$  folgt die Behauptung in allen Fällen.  $\square$

### 3 Bedingte Wahrscheinlichkeit, Unabhängigkeit

**Definition 3.1.** Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum,  $A \in \mathcal{F}$  mit  $\mathbb{P}(A) > 0$ . Dann heißt die Abbildung

$$\mathbb{P}_A : \mathcal{F} \rightarrow [0, 1], \quad B \mapsto \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

**bedingtes Wahrscheinlichkeitsmaß** unter der Bedingung  $A$ .  
Für festes  $B \in \mathcal{F}$  heißt die Zahl  $\mathbb{P}(B|A) := \mathbb{P}_A(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$  die **bedingte Wahrscheinlichkeit** von  $B$  unter der Bedingung  $A$ .

**Proposition 3.2.**  $\mathbb{P}_A$  ist das eindeutige Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{F})$  mit den Eigenschaften

- (i)  $\mathbb{P}_A(A) = 1$
- (ii)  $\exists c > 0$  mit  $\mathbb{P}_A(B) = c\mathbb{P}(B) \forall B \in \mathcal{F}$  mit  $B \subseteq A$ .

*Beweis.* a)  $\mathbb{P}_A$  ist ein Wahrscheinlichkeitsmaß und erfüllt (i),(ii).

b)  $\mathbb{P}$  ist eindeutig:

Sei  $\tilde{\mathbb{P}}_A$  ein Wahrscheinlichkeitsmaß mit den Eigenschaften (i),(ii).

Dann gilt  $\tilde{\mathbb{P}}_A(B) = \tilde{\mathbb{P}}_A(A \cap B) + \tilde{\mathbb{P}}_A(B \setminus A) = c\mathbb{P}(A \cap B)$ .

Mit  $B = A$  folgt  $1 = \tilde{\mathbb{P}}_A(A) = c\mathbb{P}(A)$ , also  $\frac{1}{\mathbb{P}(A)}$ . Also ist

$$\tilde{\mathbb{P}}_A(B) = \frac{1}{\mathbb{P}(A)}\mathbb{P}(A \cap B) = \mathbb{P}_A(B)$$

□

*Bemerkung 3.3.*  $\mathbb{P}_A$  modelliert die Situation, dass wir wissen, dass  $A$  sicher eintritt.  $\mathbb{P}_A$  beschreibt das Modell welches diese Information berücksichtigt (??(i)), aber sonst möglichst wenig ändert (??(ii)).

*Beispiel.* 2-Mal Würfeln, Information ‚Summe wird 10 sein‘.

(In der Praxis: alle Würfe mit  $X_1 + X_2 \neq 10$  werden ungültig gemacht.)

$$\begin{aligned} \mathbb{P}_A(\underbrace{X_1 = 5}_B) &= \mathbb{P}(\underbrace{\{(5, x) : 1 \leq x \leq 6\}}_B \cap \underbrace{\{(x, y) : x + y = 10\}}_A) / \mathbb{P}(\underbrace{\{(x, y) : x + y = 10\}}_A) \\ &= \frac{\mathbb{P}(\{(5, 5)\})}{\mathbb{P}(\{(4, 6), (5, 5), (6, 4)\})} = \frac{1}{3} \end{aligned}$$

**Satz 3.4.** Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum,  $N \leq \infty$ ,  $(B_i)_{i=1, \dots, N}$  mit paarweise disjunkten  $B_i \in \mathcal{F}$  und  $\bigcap_{i=1}^N B_i = \Omega$  (eine abzählbare Partition von  $\Omega$ )

a) Fallunterscheidungsformel:

$$\forall A \in \mathcal{F} \text{ ist } \mathbb{P}(A) = \sum_{i=1}^N \mathbb{P}(B_i) \mathbb{P}(A|B_i)$$

(Konvention:  $\mathbb{P}(A|B_i) = 0$ , falls  $\mathbb{P}(B_i) = 0$ )

b) *Formal von Bayes*:  $\forall k \leq N$  ist

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(B_k)\mathbb{P}(A|B_k)}{\sum_{i=1}^N \mathbb{P}(B_i)\mathbb{P}(A|B_i)} \quad (\star)$$

*Beweis.* a)  $\sum_{i=1}^N \mathbb{P}(B_i)\mathbb{P}(A|B_i) = \sum_{i=1}^N \mathbb{P}(B_i \cap A) = \mathbb{P}\left(\bigcup_{i=1}^N B_i \cap A\right) = \mathbb{P}(A)$

b)  $\mathbb{P}(B_k|A) = \frac{\mathbb{P}(B_k \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B_k)\mathbb{P}(A|B_k)}{\mathbb{P}(A)} = (\star)$

□

*Beispiel 3.5* (Bayes Formel in der Medizin, False-Positive beim HIV-Test). Sei  $B_1 = \{\text{Mensch mit HIV}\}$ ,  $B_2 = \Omega \setminus B_1 = \{\text{gesunde Menschen}\}$ . Empirisch Bekannt  $\mathbb{P}(B_1) = 0.02$  (2% infizierte),  $\mathbb{P}(A|B_1) = 0.95$  (Sensitivität 95%),  $\mathbb{P}(A|B_2) = 0.1$  (Spezifität 10%).

Angenommen „Test ist Positiv“:

$$\mathbb{P}(B_1|A) = \frac{\mathbb{P}(B_1)\mathbb{P}(A|B_1)}{\mathbb{P}(B_1)\mathbb{P}(A|B_1) + \mathbb{P}(B_2)\mathbb{P}(A|B_2)} = \frac{0.02 \cdot 0.95}{0.02 \cdot 0.95 + 0.98 \cdot 0.1} \approx \frac{1}{6}$$

Viel kleiner als die naiv vermuteten 0.9.

### 3.A Galton-Watson-Prozesse und Wahrscheinlichkeitsbäume

*Beispiel 3.6* (Mehrstufiges Modell).

- a) 1 Lebewesen bekommt  $X_{1,1} \in \mathbb{N}_0$  Nachkommen und stirbt danach.  $X_{1,1}$  ist Zufallsvariable.
- b) Die  $X_{1,1}$  Nachkommen bekommen jeweils  $X_{2,1}, X_{2,2}, \dots, X_{2,X_{1,1}}$  Nachkommen und stirbt. Nun leben

$$Y_2 = \sum_{i=1}^{X_{1,1}} X_{2,i}$$

Lebewesen.

- c) Die  $Y_2$  Lebewesen bekommen jeweils  $X_{3,1}, X_{3,2}, \dots, X_{3,Y_2}$  Nachkommen.

**Proposition 3.7.** Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $A_1, \dots, A_n \in \mathcal{F}$ . Dann gilt

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdot \dots \cdot \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1})$$

*Beweis.* Falls  $\mathbb{P}(A_1 \cap \dots \cap A_n) = 0$ , dann (Konvention) ist auch  $\mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}) = 0$ . Sonst

$$\begin{aligned} & \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdot \dots \cdot \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}) \\ &= \mathbb{P}(A_1) \frac{\mathbb{P}(A_2 \cap A_1)}{\mathbb{P}(A_1)} \frac{\mathbb{P}(A_3 \cap A_1 \cap A_2)}{\mathbb{P}(A_1 \cap A_2)} \dots \frac{\mathbb{P}(A_n \cap \dots \cap A_1)}{\mathbb{P}(A_1 \cap \dots \cap A_{n-1})} \\ &= \mathbb{P}(A_1 \cap \dots \cap A_n). \end{aligned}$$

□

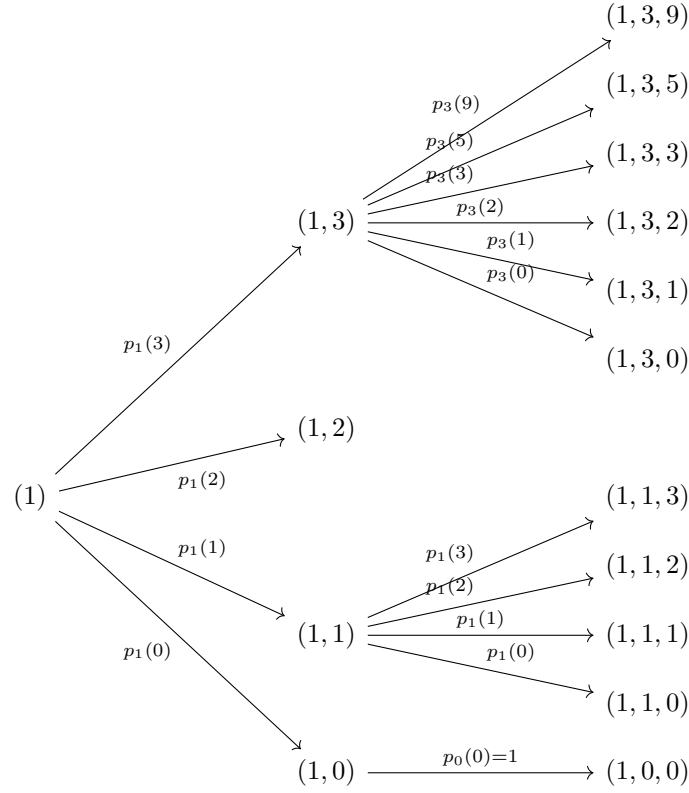


Abbildung 1: ...

**Satz 3.8.** Sei  $N \leq \infty$  und seien  $(\Omega_i, \mathcal{F}_i)$  Messräume und die  $\Omega_i$  abzählbar.

Sei  $\rho_1$  Zähldichte auf  $\Omega_1$  und

$\forall k < N, \omega_i \in \Omega_i, i \leq k$  sei  $\rho_{k+1|\omega_1, \dots, \omega_k}$  Zähldichte auf  $\Omega_{k+1}$ .

Sei  $\Omega = \prod_{i=1}^N \Omega_i$  und  $X_o : \Omega \rightarrow \Omega_i, \omega = (\omega_1, \dots) \mapsto \omega_i$  die  $i$ -te Projektion.

Dann existiert genau ein Wahrscheinlichkeitsmaß  $\mathbb{P}$  auf  $(\Omega, \bigotimes_{i=1}^N \mathcal{F}_i)$  mit den Eigenschaften

a)  $\mathbb{P}(X_1 = \omega_1) = \rho_1(\omega_1)$  für alle  $\omega_1 \in \Omega_1$ .

b)  $\forall k < N, \omega \in \Omega$  und falls  $\mathbb{P}(X_1 = \omega_1, \dots, X_k = \omega_k) \neq 0$ , dann ist

$$\mathbb{P}(X_{k+1} = \omega_{k+1} \mid X_j = \omega_j \forall j \leq k) = \rho_{\omega_{k+1}|\omega_1, \dots, \omega_k}(\omega_{k+1})$$

**Bemerkung 3.9.** a) Insbesondere osz  $\mathbb{P}(A_1 = \omega_1, \dots, X_{k+1} = \omega_{k+1}) = \rho_1(\omega_1) \cdot \dots \cdot \rho_{k+1|\omega_1, \dots, \omega_k}(\omega_{k+1})$ . („Produkt entlang der Äste“)

b) Falls  $N < \infty$  dann hat  $\mathbb{P}$  die Zähldichte  $\rho : \omega = (\omega_1, \dots, \omega_k) \mapsto \rho_1(\omega_1) \cdot \dots \cdot \rho_{N|\omega_1, \dots, \omega_{N+1}}(\omega_N)$ .

c) Falls  $N = \infty$ , dann hat  $\mathbb{P}$  im Allgemeinen keine Zähldichte

**Beweis.** a) Falls  $N < \infty$ : Nachrechnen



- b) Falls  $N = \infty$  Bilde  $[0, 1] \rightarrow \Omega$  mittels  $x \mapsto (\omega_1(x), \dots)$  mit  $\omega_i(x) =$  dasjenige  $\omega_i \in \Omega_i$ , sodass  $x$  im zu  $\omega_i$  gehörigen Intervall liegt, in Stufe  $i$ .  
 Zeige dann  $X : x \mapsto \omega(x)$  ist Zufallsvariable von  $([0, 1], \mathcal{B}([0, 1]))$  nach  $(\Omega, \mathcal{F})$ .  $\mathbb{P}$  ist dann das Bildmaß.

□

*Beispiel 3.10.1* (unendlich oft wiederholter Münzwurf). Sei  $\Omega = \{-1, 1\}^N$  und  $\mathcal{F} = \bigotimes_{i=1}^{\infty} \mathcal{P}(-1, 1)$ .  
 ( $\mathcal{F}$  wird also von den Mengen

$$\{\{\omega \in \Omega : \omega_1 = k_1, \dots, \omega_n = k_n\} : K_i \in \{-1, 1\} \forall i \leq n, n \in \mathbb{N}\}$$

erzeugt. und  $\mathbb{P}(\omega_1 = k_1, \dots, \omega_n = k_n) = 2^{-n}$ .

Im Fall von Satz 3.8 bedeutet das  $\rho_1(\omega_1) = \frac{1}{2}, \rho_{2, \omega_1}(\omega_2) = \frac{1}{2}$ .

*Beispiel 3.10.2* (Unendlich oft wiederholtes würfeln, 10-Seitiger Würfel). Sei  $\Omega = \{0, 1, \dots, 9\}$ . Die Abbildung  $X$  aus 3.8 ist hier:

$$X : [0, 1] \mapsto \Omega = \bigotimes_{i=1}^{\infty} \Omega_i, x \mapsto \omega(x)$$

**Definition 3.11.** Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  sei ein Wahrscheinlichkeitsraum,  $A, B \in \mathcal{F}$ .  
 $A, B$  heißen **unabhängig** (oder **unabhängige Ereignisse**), falls  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ . Man schreibt  $A \perp\!\!\!\perp B$ .

*Bemerkung.* a) Falls  $A \perp\!\!\!\perp B$ , dann ist

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

- b) Beachte:

$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$  („additiv“, falls  $A, B$

$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$  („multiplikativ“, falls  $A, B$  unabhängig (per Definition)).

- c) Aussage b) gilt zwar für  $A_1, A_2, A_3 \in \mathcal{F}$  falls  $\mathbb{P}(A_i \cup A_j) = \mathbb{P}(A_i) + \mathbb{P}(A_j)$  für  $i, j \in \{1, 2, 3\}$  und  $i \neq j$ . Dann gilt auch

$$\mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2)\mathbb{P}(A_3)$$

Aber: gilt nicht für Unabhängigkeit.

Unabhängigkeit ist eine algebraische Eigenschaft von Mengen und Wahrscheinlichkeitsmaßen.

Die Interpretation „ $A$  beeinflusst  $B$  nicht“ ist nicht immer richtig.

- d) Unabhängigkeit trotz Kausalität:  $\Omega = \{1, 2, 3, 4, 5, 6\}^2, \mathbb{P}(A \cap B)\mathbb{P}(\{6, 1\}) = \frac{1}{36} = \mathbb{P}(A)\mathbb{P}(B)$ . Geht nicht wenn man 7 durch 8 ersetzt!

- e)  $\mathbb{P}(A) \in \{0, 1\}$ , es folgt  $A \perp\!\!\!\perp A$

- f) Im wichtigen Fall der Produktmaße sind jedoch c) bis e) nicht relevant

**Definition 3.12.** Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum,  $I \neq \emptyset$  eine Indexmenge.  $(A_i)_{i \in I}$  heißt **unabhängige Familie** von Mengen, wenn  $\forall J \subseteq I$ , mit  $|J| < \infty$

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

(d.h. Multiplikativ für alle Kombinationen von endlich vielen Mengen) gilt.

**Definition 3.13.** Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  Wahrscheinlichkeitsraum,  $I \neq \emptyset$  sei Indexmenge und  $\forall i \in I$  seien  $(\Omega_i, \mathcal{F}_i)$  Ereignisräume. Seien  $Y_i : \Omega \rightarrow \Omega_i$  Zufallsvariablen. Die Familie  $(Y_i)_{i \in I}$  heißt **unabhängig** (oder **unabhängige Familie von Zufallsvariablen**), wenn  $\forall J \subseteq I$ , mit  $|J| < \infty$  und für alle  $(B_j)_{j \in J}$  mit  $B_j \in \mathcal{F}_j$

$$\mathbb{P} \left( \bigcap_{j \in J} Y_j^{-1}(B_j) \right) = \prod_{j \in J} \mathbb{P}(Y_j^{-1}(B_j))$$

gilt.

**Definition 3.13.** Eine Familie  $(\mathcal{A}_i)_{i \in I}$  von **Mengensystemen** heißt **unabhängig** (oder auch unabhängige Familie von Mengensystemen), wenn  $\forall J \subseteq I$ , mit  $|J| < \infty$ , für alle  $(A_j)_{j \in J}$  mit  $A_j \in \mathcal{A}_j$  sind die Mengen  $(A_j)_{j \in J}$  unabhängig.

(Dabei darf aus jedem  $\mathcal{A}_j$  höchstens ein  $A_j$  gewählt werden.)

Daher  $X \perp\!\!\!\perp Y \Leftrightarrow \sigma(X) \perp\!\!\!\perp \sigma(Y)$ .

**Satz 3.14.** In 3.13 sei  $\mathcal{G}_i$  ein Schnitt-stabiler Erzeuger von  $\mathcal{F}_i$ . Dann

$$X \perp\!\!\!\perp Y \Leftrightarrow \sigma(X) \perp\!\!\!\perp \sigma(Y)$$

*Beweis.* „ $\Leftarrow$ “ klar, da  $\mathcal{G}_i \subseteq \mathcal{F}_i$

„ $\Rightarrow$ “ Durch Induktion nach  $n := |\{i \in J : B_i \in \mathcal{F} - \mathcal{G}_i\}|$ :

$n = 0$  bedeutet  $B_i \in \mathcal{G}_i \forall i \in J$ .

$n \rightarrow n + 1$  Seien  $(B_i)_{i \in J}$  mit  $B_i \in \mathcal{F}_i \forall i \in J$  und  $B_i \in \mathcal{F}_i \setminus \mathcal{G}_i$  (n+1)-mal. Wähle  $j \in J$  und setze  $J' = J \setminus \{j\}$ , dann folgt aus der Induktionsannahme, dass für

$$A := \bigcap_{i \in J'} Y_i^{-1}(B_i)$$

$$\mathbb{P}(A) = \prod_{i \in J'} \mathbb{P}(Y_i^{-1}(B_i)) \quad (*)$$

gilt.

Falls  $\mathbb{P}(A) = 0$ , dann ist auch  $\mathbb{P}(A \cap Y_j^{-1}(B_j)) = 0$ .

Falls  $\mathbb{P}(A) > 0$ , dann definiere die Wahrscheinlichkeitsmaße

$$\mathbb{P}_i : \mathcal{F}_i \rightarrow [0, 1],$$

$$\tilde{B}_j \mapsto \mathbb{P}(Y_j^{-1}(\tilde{B}_j) | A) = \frac{\mathbb{P}(Y_j^{-1}(\tilde{B}_j) \cap A)}{\mathbb{P}(A)} \quad (**)$$

$$\mathbb{P}_2 : \mathcal{F} \rightarrow [0, 1],$$

$$\tilde{B}_j \mapsto \mathbb{P}(Y_j^{-1}(\tilde{B}_j))$$

Da ?? für  $n$  gilt ist  $\forall \tilde{B}_j \in \mathcal{G}_j$ , also

$$\mathbb{P}_1(\tilde{B}_j) = \frac{\mathbb{P}(Y_j^{-1}(\tilde{B}_j)) \cdot \prod_{i \in J'} \mathbb{P}(Y_i^{-1}(B_i))}{\mathbb{P}(A)} \stackrel{(*)}{=} \mathbb{P}(Y_j^{-1}(\tilde{B}_j)) = \mathbb{P}_2(\tilde{B}_j)$$

Es folgt, dass  $\mathbb{P}_1(\tilde{B}) = \mathbb{P}_2(\tilde{B}) \forall \tilde{B} \in \mathcal{G}_j$ , sodass aus ?? folgt, dass  $\mathbb{P}_1 = \mathbb{P}_2$ . Dann

$$\begin{aligned} \prod_{i \in J} \mathbb{P}(Y_i^{-1}(B_i)) &\stackrel{(*)}{=} \mathbb{P}(Y_j^{-1}(b_i)) \mathbb{P}(A) = \mathbb{P}_2(B_j) \mathbb{P}(A) \\ &= \mathbb{P}_1(B_j) \mathbb{P}(A) = \mathbb{P}(Y_j^{-1}(B_j) \cap A) = \mathbb{P}\left(\bigcap_{i \in J} Y_i^{-1}(B_i)\right) \end{aligned}$$

für alle  $B_i \in \mathcal{F}_i$ . Es folgt die Behauptung.  $\square$

**Korollar 3.15.** Seien  $(Y_i)_{i=1, \dots, n}$  Zufallsvariablen auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, \mathbb{P})$ , dann gilt

a) Falls  $Y_i : \Omega \rightarrow E_i$  mit  $E_i$  abzählbar, dann

$$(Y_i) \text{ ist unabhängig} \Leftrightarrow \mathbb{P}(Y_1 = e_1, \dots, Y_n = e_n) = \prod_{i=1}^n \mathbb{P}(Y_i = e_i)$$

für alle  $e_1 \in E_1, \dots, e_n \in E_n$ .

b) Falls  $Y_i : \Omega \rightarrow \mathbb{R}$ , dann ist

$$(Y_i) \text{ ist unabhängig} \Leftrightarrow \mathbb{P}(Y_1 \leq c_1, \dots, Y_n \leq c_n) = \prod_{i=1}^n \mathbb{P}(Y_i \leq c_i)$$

für alle  $c_1, \dots, c_n \in \mathbb{R}$ .

**Korollar 3.16.** Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum,  $I$  eine Menge und für alle  $i \in I$  sei  $A_i \in \mathcal{F}$ . Dann gilt

a)  $(A_i)_{i \in I}$  sind unabhängige Mengen  $\Leftrightarrow (\chi_{A_i})_{i \in I}$  unabhängige Zufallsvariablen sind.

b) Insbesondere: Falls  $(A_i)_{i \in I}$  unabhängig und  $C_i \in \{\emptyset, \Omega, A_i, A_i^C\}$  für alle  $i \in I$ , dann sind auch  $(C_i)_{i \in I}$  unabhängig.

*Beweis.* Für  $a \in \mathcal{F}$  ist  $\chi_A : \Omega \rightarrow \{0, 1\}, \omega \mapsto \chi_A(\omega)$  eine Zufallsvariable und das einelementige Mengensystem  $\{ \}$  ist ein Schnitt-stabiler Erzeuger von  $\mathcal{P}(\{0, 1\})$  und  $\chi_a^{-1}(\{1\}) = A$ .

Daher gilt, dass  $(A_i)$  unabhängig  $\Leftrightarrow ??$  gilt falls  $B_i = \{1\} \stackrel{3,14}{\Leftrightarrow} \chi_{A_i}$  ist unabhängig. Also gilt 1.

Für 2. Setze

$$B_i = \begin{cases} \{1\} & \text{falls } c_i = A \\ \{0\} & \text{falls } c_i = A^C \\ \emptyset & \text{falls } c_i = \emptyset \\ \{0, 1\} & \text{falls } c_i = \Omega \end{cases}$$

Dann folgt 2 aus der Unabhängigkeit der  $\chi_{A_i}$  und aus ?? mit den so gewählten  $B_i$ .  $\square$

Beispiel 3.17. Punkt auf Kreisscheibe: Winkel und Radius unabhängig

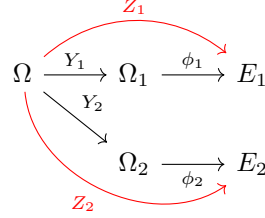


Abbildung 2: ...

**Satz 3.18.** Seien  $(Y_i)_{i=1}^\infty$  unabhängige Zufallsvariable, mit  $Y_i : (\Omega, \mathcal{F}) \rightarrow (\Omega_i, \mathcal{F}_i)$ . Dann gilt

- a) Falls  $\phi_i : (\Omega_i, \mathcal{F}_i) \rightarrow (E_i, \mathcal{E}_i)$  messbar und  $Z_i : \Omega \rightarrow E_i, \omega \mapsto \phi_i(Y_i(\omega))$ , dann sind auch  $(Z_i)_{i \in \mathbb{N}}$  unabhängig. („Stabilität unter einsetzen in Funktionen“)
- b) Seien  $J_1, J_2, \dots$ , paarweise disjunkte Teilmengen von  $\mathbb{N}$ . Für jedes  $k \in \mathbb{N}$  definiere

$$W_k(\omega) = (X_i(\omega))_{i \in J_k} \in \prod_{i \in J_k} \Omega_i$$

Dann sind die Zufallsvariablen  $(W_k)_{k \in \mathbb{N}}$  unabhängig:  
 („Stabilität gegenüber Zusammenfassen in disjunkte Blöcke“)

*Beweis.* a) Sei  $I \subset \mathbb{N}$  endlich. Dann ist (für  $A_i \in \mathcal{E}_i$ )

$$\mathbb{P} \left( \bigcap_{i \in I} Z_i^{-1}(A_i) \right) = \mathbb{P} \left( \bigcap_{i \in I} Y_i^{-1}(A_i) \right) \stackrel{Y_i \text{ unabhängig}}{=} \prod \mathbb{P}(Y_i^{-1}(\phi^{-1}(A_i)))$$

- b) Es gilt  $\times_{i=1}^n A_i = \{(\omega_1, \dots, \omega_n) : \omega_1 \in A_1, \dots, \omega_n \in A_n\} \subset \times_{i=1}^n \Omega_i$ .  
 Für  $k \in \mathbb{N}$  setze

$$\mathcal{G}_k := \left\{ \prod_{i \in J_k} A_i^{(k)} : A_i^{(k)} \in \mathcal{F}_i \forall i \in J_k, A_i^{(k)} \neq \Omega_i \text{ nur endlich oft} \right\}$$

$\mathcal{G}_k$  ist Schnitt-stabiler Erzeuger von  $\bigotimes_{i \in J_k} \mathcal{F}_i$ , also muss ?? nur auf  $(G_k)_{k \in \mathbb{N}}$  geprüft werden.

Seien also  $B_{k_1} \in \mathcal{G}_{k_1}, \dots, B_{k_n} \in \mathcal{G}_{k_n}$ . Da  $W_{k_j} \in B_{k_j}$  genau dann wenn  $X_i \in A_i^{k_j}$  ist

$$\begin{aligned}
 \mathbb{P}(W_{k_1} \in B_{k_1}, \dots, W_{k_n} \in B_{k_n}) &= \mathbb{P}(X_i \in A_i^{(k_1)} \forall i \in J_{k_1}, \dots, X_i \in A_i^{(k_n)} \forall i \in J_{k_n}) \\
 &= \prod_{l=1}^n \prod_{i \in J_{k_l}} \mathbb{P}(X_i \in A_i^{(k_l)}) = \prod_{l=1}^n \mathbb{P}(W_{k_l} \in B_{k_l})
 \end{aligned}$$

□

*Bemerkung.* Unabhängigkeit ist eng verwandt mit dem Produktmaß:

**Definition 3.21.** Sei  $(\Omega, \mathcal{F}_i, \mathbb{P}_i)$  ein Wahrscheinlichkeitsraum für  $i \in \mathbb{N}$ .  
Jedes Maß  $\mathbb{P}$  auf  $(\Omega, \mathcal{F}) := (\times_{i=1}^{\infty} \Omega_i, \bigotimes_{i=1}^{\infty} \mathcal{F}_i)$   
 $(\bigotimes_{i=1}^{\infty} \mathcal{F}_i)$  erzeugt durch Zylindermengen)  
mit der Eigenschaft

$$\mathbb{P}(\{\omega = (\omega_i)_{i \in \mathbb{N}} \in \Omega : \omega_{j_1} \in A_{j_1}, \dots, \omega_{j_m} \in A_{j_m}\}) = \prod_{k=1}^m \mathbb{P}_{j_k}(A_{j_k}) \quad (3.2)$$

für alle  $j_1 < j_2 < \dots < j_m$ , mit  $j_i \in \mathbb{N}$  und  $A_{j_i} \in \mathcal{F}_{j_i}$  heißt **Produktmaß** der Maße  $\mathbb{P}_i$ . Schreibweise  $\mathbb{P} = \bigotimes_{i=1}^{\infty} \mathbb{P}_i$

*Bemerkung 3.22.* a) In 3.21 existiert immer genau ein Produktmaß.  
(Eindeutigkeit nach Satz ??, Existenz aus dem Maßfortsetzungssatz)

b) Wichtiger Spezialfall:

- (i) Sei  $\Omega_i = \{0, 1\}$ ,  $\mathbb{P}_i(0) = \mathbb{P}_i(1) = \frac{1}{2}$  für alle  $i$  (unendlich oft wiederholter Münzwurf)
- (ii) Sei  $\Omega_i = \mathbb{R}$ ,  $\mathcal{F}_i = \mathcal{B}(\mathbb{R})$ ,  $\mathbb{P}_i = \rho(x) dx$  für alle  $i$ .  
(Zufällige Folge mit Gliedern die gemäß  $\rho(x) dx$  verteilt sind.)

**Proposition 3.23.** a) Seien  $(X_i)_{i \in \mathbb{N}}$  Zufallsvariablen,  $X_i : (\Omega, \mathcal{F}) \rightarrow (\Omega_i, \mathcal{F}_i)$ , dann gilt

$(X_i)$  ist unabhängig  $\Leftrightarrow$  Das Bildmaß der Zufallsvariablen  $X : \Omega \rightarrow \prod_{i=1}^{\infty} \Omega_i, \omega \mapsto (X_i(\omega))_{i \in \mathbb{N}}$  ist ein Produktmaß

b) Seien  $(\Omega_i, \mathcal{F}_i)_{i \in \mathbb{N}}$  seien Maßräume,  $\Omega = \times_{i=1}^{\infty} \Omega_i, \mathcal{F} = \bigotimes_{i=1}^{\infty} \mathcal{F}_i$ .  
 $\mathbb{P}$  sei ein beliebiges Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{F})$ . Dann gilt:

$\mathbb{P}$  ist Produktmaß  $\Leftrightarrow$  Die Projektionen  $X_i : \Omega \rightarrow \Omega_i, \omega = (\omega_1, \omega_2, \dots) \mapsto \omega_i$  sind unabhängige Zufallsvariablen

*Bemerkung 3.24.* In 3.23a) heißt der Wahrscheinlichkeitsraum  $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}) := (\times_{i=1}^{\infty} \Omega_i, \bigotimes_{i=1}^{\infty} \mathcal{F}_i, \mathbb{P}_X)$  mit  $X = (X_i)_{i \in \mathbb{N}}$  der kanonische Wahrscheinlichkeitsraum der Zufallsvariablen  $X_i$ . Die  $i$ -te Koordinatenabbildung  $Y_i : \tilde{\Omega} \rightarrow \Omega_i$  entspricht dann jeweils  $X_i$ , d.h.  $(Y_{i_1}, \dots, Y_{i_n})$  hat die gleiche Verteilung wie  $(X_{i_1}, \dots, X_{i_n})$ .

**Definition 3.25.** Seien  $\rho_1, \dots, \rho_n$  Wahrscheinlichkeitsmaße auf  $\mathbb{R}$  mit der Dichte  $\rho_1, \dots, \rho_n$ . Dann heißt

$$\rho : \mathbb{R}^n \rightarrow \mathbb{R}_0^+, (x_1, \dots, x_n) \mapsto \rho(x_1, \dots, x_n) = \rho_1(x_1) \dots \rho_n(x_n)$$

die **Produktdichte**.

**Satz 3.25.** Die Produktdichte ist die Dichte des Produktmaßes.

*Beweis.* Es gilt

$$\begin{aligned} \mathbb{P}((-\infty, c_1] \times (-\infty, c_2] \times \dots \times (-\infty, c_n]) &= \mathbb{P}(X_1 \leq c_1, \dots, X_n \leq c_n) \\ &\stackrel{??b)}{=} \mathbb{P}(X_1 \leq c) \dots \mathbb{P}(X_n \leq c) \\ &= \int_{-\infty}^{c_1} \rho_1(x_1) dx_1 \cdot \dots \cdot \int_{-\infty}^{c_n} \rho_n(x_n) dx_n \\ &\stackrel{\text{Satz von Fubini}}{=} \int_{(-\infty, c_1] \times \dots \times (-\infty, c_n]} \rho_1(x_1) \dots \rho_n(x_n) dx_1 \dots dx_n \\ &= \int_{(-\infty, c_1] \times (-\infty, c_2] \times \dots \times (-\infty, c_n]} \rho(x) dx \end{aligned}$$

Dann folgt mit ??, dass  $\mathbb{P}$  die Dichte  $\rho$  hat.  $\square$

*Bemerkung.* Für unabhängige Zufallsvariablen  $X, Y$  kann man die Dichte  $X+Y$  mittels Faltungsprodukt berechnen.

**Satz 3.26.** Seien  $X, Y : \Omega \rightarrow \Omega'$  unabhängige Zufallsvariablen

- a) Falls  $\Omega' = \mathbb{Z}$  und falls  $\mathbb{P}_X$  (bzw.  $\mathbb{P}_Y$ ) Zähldichte  $\rho_1$  (bzw.  $\rho_2$ ), dann hat  $\mathbb{P}_{X+Y}$  die Zähldichte

$$\rho_1 * \rho_2 : \mathbb{Z} \rightarrow [0, 1], z \mapsto \rho_1 * \rho_2(z) = \sum_{k \in \mathbb{Z}} \rho_1(k) \rho_2(z - k)$$

- b) Falls  $\Omega' = \mathbb{R}$  und falls  $\mathbb{P}_X$  (bzw.  $\mathbb{P}_Y$ ) die Dichte  $\rho_1$  (bzw.  $\rho_2$ ), dann hat  $\mathbb{P}_{X+Y}$  die Dichte

$$\rho_1 * \rho_2 : \mathbb{R} \rightarrow \mathbb{R}_0^+, x \mapsto \rho_1 * \rho_2(x) = \int_{\mathbb{R}} \rho_1(y) \rho_2(x - y) dy$$

*Beweis.* a) Da nur  $\mathbb{P}_X, \mathbb{P}_Y, \mathbb{P}_{X+Y}$  untersucht werden kann man  $\Omega = \mathbb{Z} \times \mathbb{Z}$ ,  $X((z_1, z_2)) = z_1$  und  $Y((z_1, z_2)) = z_2$  wählen und  $\mathbb{P}_{(X,Y)}$  hat die Produkt-Zähldichte  $(z_1, z_2) \mapsto \rho_1(z_1) \rho_2(z_2)$ .

Dann gilt  $X + Y = w \in \mathbb{Z}$ , genau dann wenn  $z_1 + z_2 = w$ , bzw.  $z_1 = w - z_2$ .

Also  $(X + Y)^{-1}(\{w\}) = \{(z_1, z_2) : z_1 = w - z_2\}$ , es folgt

$$\mathbb{P}_{X+Y}(\{w\}) = \mathbb{P}_{(X,Y)}((X+Y)^{-1}(\{w\})) = \sum_{\substack{z_1 \in \mathbb{Z} \\ z_2 \in \mathbb{Z} \\ z_1 + z_2 = w}} \rho(z_1) \rho(z_2) = \sum_{z \in \mathbb{Z}} \rho_1(z) \rho_2(w - z)$$

- b) Setze  $\Omega = \mathbb{R}^2$ ,  $X((x_1, x_2)) = x_1$ ,  $Y((x_1, x_2)) = x_2$ ,  $\mathbb{P}([a, b] \times [c, d]) = \int_{[a,b] \times [c,d]} \rho_1(x_1) \rho_2(x_2) dx_1 dx_2$ . Dann ist

$$\begin{aligned} \mathbb{P}(X + Y \leq c) &= \mathbb{P}(\{(x, y) \in \Omega\} : x + y \leq c) \\ &= \int_{-\infty}^{\infty} \rho_1 \int_{-\infty}^{c-x} \rho_2(y) dy dx \\ &= \int_{-\infty}^{\infty} \rho_1 \int_{-\infty}^c \rho_2(\tilde{y} - x) d\tilde{y} dx \\ &= \int_{-\infty}^c \underbrace{\int_{-\infty}^{\infty} \rho_1(x) \rho_2(\tilde{y} - x) dx}_{=\rho_1 * \rho_2(\tilde{y})} d\tilde{y} \end{aligned}$$

Dann folgt mit ??, dass  $\mathbb{P}_{X+Y}$  die Dichte  $\rho_1 * \rho_2$  hat.  $\square$

*Beispiel 3.27.1.* Sei  $X \sim \mathcal{N}_{\mu, \nu}$  und  $Y \sim \mathcal{N}_{\mu', \nu'}$ , dann ist  $X + Y \sim \mathcal{N}_{\mu + \mu', \nu + \nu'}$

*Beispiel 3.28.* Sei  $\Omega = \mathbb{R}^n$ ,  $\mathcal{F} = \otimes_{x=1}^\infty \mathcal{B}(\mathbb{R})$ ,  $X_i(\omega) = \omega_i$  Projektionen. Sei

$$\begin{aligned} A_1 &= \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = \lim_{n \rightarrow \infty} \omega_n \text{ existiert} \right\} \\ A_2 &= \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_n(\omega) = 0 \right\} \\ A_3 &= \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \sum_{i=1}^n X_n(\omega) = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \sum_{i=1}^n \omega_n \text{ existiert} \right\} \end{aligned}$$

Für  $A_1, A_2, A_3$  gilt: Sei  $\omega \in A_j$ . Wenn  $\tilde{\omega}_i = \omega_i$  für alle außer endlich viele  $i \in \mathbb{N}$  gilt, dann ist  $\tilde{\omega} \in A_j$ . Ebenso wenn  $\omega \in A_j^C$ , dann  $\omega \in A_j^C$ .

**Definition 3.29.** Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und seien  $X_i : (\Omega, \mathcal{F}) \rightarrow (\Omega_i, \mathcal{F}_i)$  seien Zufallsvariablen. Setze

$$\mathcal{F}_{\{i\}} = \sigma(X_i) = \{X_i^{-1}(B) : B \in \mathcal{F}_i\} \quad (\text{„von } X_i \text{ erzeugte } \sigma\text{-Algebra“})$$

Setze für  $K \subseteq \mathbb{N}$

$$\mathcal{F}_K := \sigma(\{X_i : i \in K\}) = \sigma(\mathcal{F}_{\{i\}} : i \in K) \quad (\text{„von } (X_i)_{i \in K} \text{ erzeugte } \sigma\text{-Algebra“})$$

Schreibe für  $n \in \mathbb{N}$

$$\begin{aligned} \mathcal{F}_n &:= \mathcal{F}_{\{1, \dots, n\}} && (\text{„}\sigma\text{-Algebra bis zur ‚Zeit‘ } n\text{“}) \\ \tau_n &:= \mathcal{F}_{n+1, n+2, \dots} && (\text{„}\sigma\text{-Algebra nach der Zeit } n\text{“}) \\ \tau &:= \bigcap_{n \in \mathbb{N}} \tau_n && (\text{„terminale/asymptotische } \sigma\text{-Algebra“}) \end{aligned}$$

*Beispiel 3.30.* Für  $\Omega = \mathbb{R}^N$ ,  $\mathcal{F} = \mathcal{B}(\mathbb{R}^N)$ ,  $X_i$  ist  $i$ -te Projektion. Dann ist

$$\mathcal{F}_{\{1, \dots, n\}} = \left\{ \{x \in \mathbb{R}^N : (x_1, \dots, x_n) \in B, x_{n+1} \text{ beliebig}, x_{n+2} \text{ beliebig}, \dots\} : B \in \mathcal{B}(\mathbb{R}^n) \right\}$$

sind Zylindermengen.

Für  $|K| = \infty$  (insbesondere für  $\tau_n$ ) kann man  $\mathcal{F}_K$  nicht explizit angeben. Stattdessen:  $\tau_n$  = kleinste  $\sigma$ -Algebra, die  $\mathcal{F}_{n+1, n+2, \dots, m}$  enthält für alle  $m > n$ .

**Definition 3.31.** Sei  $(\Omega, \mathcal{F})$  Messraum,  $A_i \in \mathcal{F}$  für alle  $i \in \mathbb{N}$ . Definiere

$$\begin{aligned} \limsup_{n \rightarrow \infty} A_n &:= \bigcap_{m=1}^\infty \bigcup_{k \geq m} A_k = \{\omega \in \Omega : \omega \in A_i \text{ für unendlich viele } i\} \\ \liminf_{n \rightarrow \infty} A_n &:= \bigcup_{m=1}^\infty \bigcap_{k \geq m} A_k = \{\omega \in \Omega : \omega \in A_i \text{ für alle außer endlich viele } i\} \end{aligned}$$

**Proposition 3.32.** a) Die Gleichheiten in 3.31 gelten

$$\text{b) } \bigcap_{n=1}^\infty A_n \subseteq \liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n \subseteq \bigcup_{n=1}^\infty A_n$$

c) Für alle  $\omega \in \Omega$  gilt

$$\begin{aligned} \liminf_{n \rightarrow \infty} \chi_{A_i}(\omega) &= \chi_{\liminf_{n \rightarrow \infty} A_i}(\omega) \\ \limsup_{n \rightarrow \infty} \chi_{A_i}(\omega) &= \chi_{\limsup_{n \rightarrow \infty} A_i}(\omega) \end{aligned}$$

- d) Falls  $X_i : (\Omega, \mathcal{F}) \rightarrow (\Omega_i, \mathcal{G}_i)$  Zufallsvariablen sind und  $A_i = X^{-1}(B_i)$  mit  $B_i \in \mathcal{G}_i$ , dann sind  $\limsup_{n \rightarrow \infty} A_n \in \tau$  und  $\liminf_{n \rightarrow \infty} A_n \in \tau$ .  
(Nicht aber  $\bigcap_{n=1}^{\infty} A_n$  oder  $\bigcup_{n=1}^{\infty} A_n$ !)

*Beispiel 3.33.* In der Situation von 3.29 mit  $\Omega_i = \mathbb{R}$  ist

$$A = \left\{ \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i \text{ existiert und liegt in } [a, b] \right\} \in \tau$$

für alle  $a \leq b$ .

**Satz 3.34.**

*Bemerkung.* In der Situation von ?? können also nur Wahrscheinlichkeiten  $\{0, 1\}$  auftreten, wenn die  $X_i$  unabhängig sind. Welche der beiden Möglichkeiten trifft zu?

Diese Frage ist i.A. schwierig zu beantworten, aber im Fall  $\limsup_{n \rightarrow \infty} A_n$  gibt es folgenden Satz:

**Satz 3.35** (Lemma von Borel-Cantelli). *Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $A_j \in \mathcal{F}$ , für alle  $j \in \mathbb{N}$ . Dann*

- a) Falls  $\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$ , dann ist  $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$   
b) Falls  $(A_k)_{k \in \mathbb{N}}$  unabhängig und  $\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \infty$ , dann ist  $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1$ .

*Beweis.* a) Für alle  $m$  ist  $A \subseteq \bigcap_{k=1}^{\infty} A_k$ . Also ist  $\mathbb{P}(A) \leq \sum_{k=m}^{\infty} \mathbb{P}(A_k)$ .  
Da  $\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$  gilt, muss  $\lim_{n \rightarrow \infty} \sum_{k=m}^{\infty} \mathbb{P}(A_k) = 0$  gelten. Da die Mengenfolge  $(\bigcup_{k=m}^{\infty} A_k)_{m \in \mathbb{N}}$  absteigend ist gilt

$$\mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k\right) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{k=m}^{\infty} A_k\right) \leq \lim_{m \rightarrow \infty} \sum_{k=m}^{\infty} \mathbb{P}(A_k) = 0$$

- b) Es gilt für das Komplement:

$$A^C = \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} A_k^C$$

Also ist

$$\begin{aligned} \mathbb{P}(A^C) &\leq \sum_{m=1}^{\infty} \mathbb{P}\left(\bigcap_{k=m}^{\infty} A_k^C\right) \stackrel{\substack{\text{Satz ?? c) \\ \text{absteigende} \\ \text{Mengenfolge}}}{=} \sum_{m=1}^{\infty} \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcap_{k=m}^n A_k^C\right) \\ &\stackrel{\text{unabh.}}{=} \sum_{m=1}^{\infty} \lim_{n \rightarrow \infty} \prod_{k=m}^n (1 - \mathbb{P}(A_k)) \\ &= \sum_{m=1}^{\infty} \lim_{n \rightarrow \infty} e^{-\sum_{k=m}^n \mathbb{P}(A_k)} \\ &= \sum_{m=1}^{\infty} 0 = 0 \end{aligned}$$

□



## 4 Erwartungswert und Varianz

**Bemerkung 4.1** (Motivation). Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $(X_i)_{i \in \mathbb{N}}$  mit  $X_i : \Omega \rightarrow \mathbb{Z}$ ,  $\mathbb{P}(X_i = k) = p_k$  seien unabhängige Zufallsvariablen. (d.h.  $(p_k)$  ist zähl-dichte für jedes  $X_i$ )  
Wenn man nun  $N$ -mal mit  $(X_i)$  „würfelt“, erhält man  $N$  zufällige ganze Zahlen  $X_1(\omega), \dots, X_n(\omega)$ .  
(„Realisierung von  $(X_1, \dots, X_n)$ “)

**Frage** Wie groß ist der (zufällige) Mittelwert für  $\frac{1}{N} \sum_{i=1}^n X_i$  typischerweise wenn  $n$  groß (unendlich groß) wird?

**Beispiel**  $\mathbb{P}(X_i = k) = \frac{1}{6}$ , falls  $k \in \{1, \dots, 6\}$ . Dann ist der Mittelwert  $\approx 3.5$ . Es ist aber auch  $(1, 1, 1, \dots, 1)$  möglich.

**Überlegung**  $p_k$  modelliert die „relative Häufigkeit“ mit der der Wert  $k$  auftritt.  
D.h. für große  $n$  sollte man cs  $p_k \cdot n$ -mal den Wert  $k$  erhalten.  
Wähle z.B.  $(-4, 2, 1, 2, -4, 0, 1, 3, 2, -3, 1, 0)$ :

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n X_i(\omega) &= \frac{1}{n} (0|\{i \leq n : X_i(\omega) = 0\}| + 1|\{i \leq n : X_i(\omega) = 1\}| + (-1)|\{i \leq n : X_i(\omega) = -1\}| + \dots) \\ &= \sum_{k \in \mathbb{Z}} \frac{k}{n} |\{i \leq n : X_i(\omega) = k\}| \\ &\stackrel{?}{\approx} \sum_{k \in \mathbb{Z}} k p_k \end{aligned} \quad (\star)$$

**Beobachtung** Das letzte  $\approx$  stimmt sicher nicht immer, sollte aber „meistens“ richtig sein.  
Siehe dazu ?? (Gesetze der großen Zahlen)

Betrachte nun die rechte Seite von  $(\star)$ :

**Definition 4.2.** Eine reelle Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  heißt **diskret** wenn die Menge  $X(\Omega)$  höchstens abzählbar ist

**Definition 4.3.** Sei  $X$  eine diskrete Zufallsvariable.  $X$  heißt **integrierbar** (bzw. „**Erwartungswert existiert**“), falls

$$\sum_{y \in X(\Omega)} |y| \mathbb{P}(X = y) < \infty$$

Man schreibt  $X \in \mathcal{L}^1$  bzw.  $X \in \mathcal{L}^1(\mathbb{P})$ .

**Definition 4.3.** Für  $X \in \mathcal{L}^1$  heißt

$$\mathbb{E}(X) = \mathbb{E}_{\mathbb{P}}(X) := \sum_{y \in X(\Omega)} \mathbb{P}(X = y) \in \mathbb{R}$$

der **Erwartungswert** von  $X$ .

**Beispiel 4.4.1.** Sei  $X : \Omega \rightarrow \{a, b\}$ ,  $a, b \in \mathbb{R}$  und  $\mathbb{P}(X = a) = p = 1 - \mathbb{P}(X = b)$ . Dann ist  $\mathbb{E}(X) = p \cdot a + (1 - p)b = b + p(a - b)$ .

$a = 1, b = 0$  Dann ist  $\mathbb{E}(X = p)$  (Münzwurf)

$a = 1, b = -1$  Dann ist  $\mathbb{E}(X = 1 - 2p)$  (Schrittweise Irrfahrt)

**Beispiel 4.4.2.** Sei  $T$  geometrisch verteilt, also  $\mathbb{P}(T = k) = p(1 - p)^{k-1}$ , dann ist

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} \overbrace{p(1-p)}^{=:q} \quad k-1 = p \underbrace{\sum_{k=1}^{\infty} k q^{k-1}}_{=:f(q)} = (\star)$$

Für die Funktion  $f(q)$  gilt

$$\begin{aligned} f(q) &= \frac{d}{dq} \sum_{k=0}^{\infty} q^k = \frac{d}{dy} \frac{1}{1-q} \\ &= \frac{1}{(1-q)^2} = \frac{1}{p^2} \end{aligned}$$

Also ist  $(\star) = \frac{p}{p^2} = p$ .

**Satz 4.5.** Seien  $X, Y \in \mathcal{L}^1$  diskrete Zufallsvariablen.

- a) Falls  $X \geq Y$ , dann ist  $\mathbb{E}(X) \geq \mathbb{E}(Y)$
- b) (i) Sei  $c \in \mathbb{R}$ , dann ist auch  $cX \in \mathcal{L}^1$  und  $\mathbb{E}(cX) = c\mathbb{E}(X)$ .  
(ii) Es ist auch  $X + Y \in \mathcal{L}^1$  und  $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

*Beweis.* Aus der Maßtheorie, da  $\mathbb{E}(X)$  das (elementare) Integral ist. □

**Definition 4.6.** Sei  $X$  reelle Zufallsvariable. Dann heißt  $X_{(n)}$  für alle  $n \in \mathbb{N}$

$$X_{(n)}(\omega) := \left\lfloor \frac{1}{n} \right\rfloor [nX(\omega)]$$

die  $\frac{1}{n}$ -Approximation.  
(Dabei ist  $\lfloor \cdot \rfloor$  die Abrunden-Funktion.)

**Satz 4.7.** Sei  $X$  reelle Zufallsvariable,  $X_{(n)}$  aus 4.6.

- a) Für alle  $n \in \mathbb{N}$  ist  $X_{(n)} \leq X \leq X_{(n)} + \frac{1}{n}$
- b) Falls  $X(N_0) \in \mathcal{L}^1$  für ein  $n_0 \in \mathbb{N}$ , dann ist  $X_{(n)} \in \mathcal{L}^1$  für alle  $n$ .
- c) Falls b) gilt, so ist  $(\mathbb{E}(X_{(n)}))_{n \in \mathbb{N}}$  eine Cauchyfolge.

*Beweis.*

- a) ist klar.
- b) Aus a) folgt, dass  $|X_{(n)}(\omega)| \leq |X_{(m)}| + \max\{\frac{1}{n}, \frac{1}{m}\}$ .  
Da insbesondere  $1 \in \mathcal{L}^1(\mathbb{P})$  folgt die Behauptung.
- c) Aus b) folgt, dass  $|\mathbb{E}(X_{(n)}) - \mathbb{E}(X_{(m)})| \leq \max\{\frac{1}{n}, \frac{1}{m}\} \xrightarrow{n, m \rightarrow \infty} 0$ . Ist Cauchyfolge.

□

**Satz 4.8.** Die Aussagen von Satz 4.5 (Linearität und Monotonie) gelten auch für Reelle Zufallsvariablen.

*Beweis.* Durch Grenzwertsätze der Maßtheorie: □

**Satz 4.10** (Monotonen Konvergenz). Seien  $(X_n)_{n \in \mathbb{N}}$ ,  $X$  reelle Zufallsvariablen mit  $X_n(\omega) \leq X_{n+1}(\omega)$  für alle  $\omega \in \Omega$  und  $\lim_{n \rightarrow \infty} X_n(\omega) = \sup_{n \in \mathbb{N}} X_n(\omega) = X(\omega)$ .

Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(X)$$

**Satz 4.11** (Majorisierte Konvergenz). Seien  $(X_n)_{n \in \mathbb{N}}$ ,  $X$  reelle Zufallsvariablen und es existiert  $Y \in \mathcal{L}^1$ , mit  $|X_n(\omega)| \leq |Y(\omega)|$  für alle  $\omega \in \Omega$  und  $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$  für alle  $\omega \in \Omega \setminus \Omega_0$ , mit  $\Omega_0$  Nullmenge. Dann ist

$$\lim_{n \rightarrow \infty} \liminf \mathbb{E}(X_n) = \mathbb{E}(X)$$

**Definition 4.12.** Sei  $A := \{\omega \in \Omega \mid \omega \text{ hat Eigenschaft } E\} \in \mathcal{F}$ .

Falls  $\mathbb{P}(A^c) = 0$ , dann sagen wir „die Eigenschaft  $E$  gilt  $\mathbb{P}$ -fast sicher“.

*Beispiel 4.13.1.* Sei  $\mathbb{P} = U[0, 1] = \lambda_{[0,1]}$ ,  $X(\omega) = \omega$ . Dann ist  $X \notin \mathbb{Q}$   $\mathbb{P}$  fast sicher, denn  $\lambda(\mathbb{Q} \cap [0, 1]) = 0$

*Beispiel 4.13.2.* Sei  $\mathbb{P} = U[-1, 1]$ ,  $X_n(y) = \sin\left(\frac{1}{|y| + \frac{1}{n}}\right)$ , dann konvergiert  $X_n(y)$   $\mathbb{P}$  fast sicher.

*Bemerkung 4.14.* Falls  $X \geq 0$ , dann erlauben wir auch  $\mathbb{E}(X) = \infty$ . Wir definieren dann  $\mathbb{E}(X) = \lim_{n \rightarrow \infty} \mathbb{E}(\min\{X, n\})$ .

Satz 4.8a), b) und 4.10 gilt weiterhin, falls  $X, Y \geq 0$  und  $c \geq 0$ .

**Satz 4.15.** Seien  $(X_n)$  reelle Zufallsvariable und  $X_n \geq 0$ , dann ist

$$\mathbb{E}(\liminf_{n \rightarrow \infty} X_n) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(X_n)$$

*Beispiel 4.16.* Sei  $(\Omega, \mathcal{F}, \mathbb{P}) = ((0, 1), \mathcal{B}((0, 1)), \lambda_{(0,1)})$  und  $X_n = n\chi_{(0, \frac{1}{n}]}$ .

Dann ist  $\mathbb{E}(X_n) = 1 \forall n$ , aber  $\liminf_{n \rightarrow \infty} X_n(\omega) = 0$ .

Also kann die Ungleichung in Satz 4.15 kann also Strikt sein.

**Satz 4.17** (Integration bezüglich des Bildmaßes). Sei  $X : \Omega \rightarrow \Omega'$  Zufallsvariable,  $f : \Omega' \rightarrow \mathbb{R}$  messbar.

Es gelte entweder  $f(\omega') \geq 0 \forall \omega' \in \Omega'$  oder

$$\mathbb{E}(|f(X)|) = \int_{\Omega} |f(X(\omega))| \mathbb{P}(d\omega) < \infty$$

Dann gilt:

$$\mathbb{E}(f(X)) = \int_{\Omega'} f(\omega') \mathbb{P}_X(d\omega') = \mathbb{E}_{\mathbb{P}_X}(f) = \int_{\Omega'} f(\omega') \mathbb{P} \circ X^{-1}(d\omega')$$

*Beweis.* (i) Sei  $f(\omega') = \chi_B(\omega')$  mit  $B \in \mathcal{F}'$ .

Dann ist  $\mathbb{E}(f(X)) = \mathbb{E}(\chi_B(X)) = \mathbb{P}(X \in B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}_X(B)$  nach Definition von  $\mathbb{P}_X$ . Dann gilt die Gleichung aus 4.17.

(ii) Sei nun  $f(\omega') = \sum_{n=1}^m c_n \chi_{B_n}(\omega')$ , mit  $C_n \in \mathbb{R}$ ,  $B_n \in \mathcal{F}'$ . Dann ist

$$\mathbb{E}(f(X)) \stackrel{4.8}{=} \sum_{n=1}^m c_n \mathbb{E}(\chi_{B_n}(X)) \stackrel{(i)}{=} \sum_{n=1}^m c_n \mathbb{E}_{\mathbb{P}_X}(\chi_{B_n}) \stackrel{4.8b)}{=} \mathbb{E}_{\mathbb{P}_X} \left( \sum_{n=1}^m c_n \chi_{B_n} \right) = \mathbb{E}_{\mathbb{P}_X}(f)$$

(iii) Sei nun  $f(\omega') \geq 0 \forall \omega' \in \Omega$ . Setze

$$f_n(\omega') := \sup \left\{ \frac{1}{n} \lfloor f(\omega') \rfloor, n \right\}$$

(Abgeschnittene  $\frac{1}{n}$  Approximation.)

Dann folgt aus (ii), dass  $\mathbb{E}(f_n(X)) = \mathbb{E}_{\mathbb{P}_X}(f_n)$ . Dann folgt aus der monotonen Konvergenz, dass im Grenzwert  $n \rightarrow \infty$  gilt:  $\mathbb{E}(f(X)) = \mathbb{E}_{\mathbb{P}_X}(f)$ . Dann gilt die Gleichung aus 4.17.

(iv) Sei  $\mathbb{E}(|f(X)|) < \infty$ . Da  $|f(X(\omega))| = f_+(X(\omega)) + f_-(\omega')$  ist auch  $\mathbb{E}(f_-(X)) < \infty$  und  $\mathbb{E}(f_+(X)) < \infty$ .

Da  $f(\omega') = f_+(\omega') - f_-(\omega')$  gilt ist

$$\mathbb{E}(f(X)) = \mathbb{E}(f_+(X)) - \mathbb{E}(f_-(X)) \stackrel{(iii)}{=} \mathbb{E}_{\mathbb{P}_X}(f_+) - \mathbb{E}_{\mathbb{P}_X}(f_-) = \mathbb{E}_{\mathbb{P}_X}(f)$$

□

*Bemerkung 4.18.* Sei  $H : [a, b] \rightarrow \mathbb{R}$  streng monoton wachsend und  $h \in C^1$ . Sei  $\Omega = [a, b]$ ,  $\mathcal{F} = \mathcal{B}([a, b])$ ,  $\mu(dx) = h'(x) dx$ .

Dann ist  $\mu([a, b]) = \int_a^b h'(x) dx = h(b) - h(a)$  das Bildmaß von  $\mu$  unter  $h$ .

$$\begin{aligned} \mu \circ h^{-1}((-\infty, x]) &= \mu(\{y \in [a, b] : -\infty \leq h(y) \leq x\}) = \mu([a, b] \cap (-\infty, h^{-1}(x))) \\ &= \int_a^{h^{-1}(x)} h'(y) dy = [h(y)]_a^{h^{-1}(x)} = h(h^{-1}(x)) - h(a) = x - h(a) \end{aligned}$$

falls  $h(a) \leq x \leq h(b)$ . Dann ist

$$\mu \circ h^{-1} = \lambda_{[h(a), h(b)]}$$

Dann folgt aus 4.17

$$\int_a^b f(h(x)) h'(x) dx = \mathbb{E}_{\mu}(f(h)) = \mathbb{E}_{\mathbb{P}_h}(f) = \int_{h(a)}^{h(b)} f(y) dy$$

**Korollar 4.19.** Seien  $X : \Omega_1 \rightarrow \Omega'$ ,  $Y : \Omega_2 \rightarrow \Omega'$  Zufallsvariablen mit  $\mathbb{P}_X = \mathbb{P}_Y$ . Dann gilt  $\mathbb{E}(f(X)) = \mathbb{E}(f(Y))$  für alle  $f : \Omega' \rightarrow \mathbb{R}$  mit  $\mathbb{E}(|f(x)|) < \infty$  oder  $f \geq 0$ .

Umgekehrt: sei  $\mathbb{E}(f(X)) = \mathbb{E}(f(Y))$  für alle  $f$  messbar, beschränkt, dann gilt  $\mathbb{P}_X = \mathbb{P}_Y$ .

*Beweis.* „ $\Rightarrow$ “ Folgt aus 4.17.

„ $\Leftarrow$ “  $\mathbb{P}_X(A) = \mathbb{E}(\chi_A(X)) = \mathbb{E}(\chi_A(Y)) = \mathbb{P}_Y(A)$ .

□

*Beispiel 4.20.1.* Sei  $X \sim \text{Exp}(\alpha)$ , d.h.  $\mathbb{P}_X(dx) = \alpha e^{-\alpha x} dx$ . Dann ist

$$\mathbb{E}(X^k) = \alpha \int_0^\infty x^k e^{-\alpha x} dx = \frac{k!}{\alpha^k}$$

*Beispiel 4.20.2.* Sei  $X \sim \mathcal{N}(m, v)$ , dann ist  $\mathbb{E}(X) = m$  und  $\mathbb{E}((X - m)^2) = v$

**Korollar 4.21.** a) Sei  $X$  reelle Zufallsvariable.  $\mathbb{P}_X$  habe dichte  $\rho$  bezüglich des Lebesgue-Maß. Dann ist

$$\mathbb{E}(f(X)) = \int_{-\infty}^{\infty} \rho(x) f(x) dx$$

(falls  $\int \rho(x) |f(x)| dx < \infty$ ). Insbesondere

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \rho(x) dx$$

b) Sei  $\Omega = \mathbb{R}$ ,  $\mathbb{P}(dy) = \rho(y) dy$  ( $\rho$  ist Dichte von  $\mathbb{P}$ ).  
Sei  $X$  reelle Zufallsvariable. Dann gilt:

$$\mathbb{E}(f(X)) = \int_{-\infty}^{\infty} f(X(y)) \rho(y) dy$$

Inbesondere  $\mathbb{E}(X) = \int_{-\infty}^{\infty} X(y) \rho(y) dy$

**Definition 4.22.** Sei  $X$  reelle Zufallsvariable

- a) Falls  $|X^n| \in \mathcal{L}^1$  für  $n \in \mathbb{N}$ , so heißt  $\mathbb{E}(X^n)$  das  **$n$ -te Moment von  $X$** .
- b) Falls  $|X^p| \in \mathcal{L}^1$ , für  $p > 0$ , so schreibt man  $X \in \mathcal{L}^p$ .
- c) Die Zahl  $\mathbb{V}(X) := \mathbb{E}((X - \mathbb{E}(X))^2)$  heißt die **Varianz** von  $X$  (bezüglich  $\mathbb{P}$ ).  
 $\sqrt{\mathbb{V}(X)}$  heißt **Standardabweichung** von  $X$ .

*Bemerkung 4.23.* a) Für die Varianz gilt

$$\mathbb{V}(X) = \mathbb{E}(X^2 - 2\mathbb{E}(X)X + \mathbb{E}(X)^2) = \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

b) Da  $|X|^p \leq 1 + |X|^q$ , falls  $p \leq q$  ist

$$\mathcal{L}^q \leq \mathcal{L}^p$$

falls  $p \leq q$ .

**Definition 4.24.** Sei  $X, Y \in \mathcal{L}^2$ , dann heißt

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

die **Kovarianz** von  $X$  und  $Y$ .

Man nennt  $X$  und  $Y$

**positiv korreliert** falls  $\text{Cov}(X, Y) > 0$

**unkorreliert** falls  $\text{Cov}(X, Y) = 0$

**negativ korreliert** falls  $\text{Cov}(X, Y) < 0$

*Beispiel 4.25.* Sei  $\Omega = [0, 1]$  mit Lebesgue-Maß, sei  $X(\omega) = \omega, Y(\omega) = \omega^2$  und  $Z(\omega) = \frac{1}{2} \sin(\pi\omega)$ . Dann ist

$$\begin{aligned}\mathbb{E}(X) &= \int_0^1 x \, dx = \frac{1}{2} \\ \mathbb{E}(Y) &= \int_0^1 x^2 \, dx = \frac{1}{3} \\ \mathbb{E}(Z) &= \int_0^1 \frac{\sin(\pi x)}{2} \, dx = \frac{1}{2\pi}\end{aligned}$$

Dann folgt für die Kovarianz:

$$\text{Cov}(X, Y) = \mathbb{E}\left(\left(X - \frac{1}{2}\right)\left(Y - \frac{1}{3}\right)\right) = \frac{1}{12}$$

$$\text{Cov}(X, Z) = \int_0^1 \left(x - \frac{1}{2}\right)\left(\sin(\pi x) - \frac{1}{2\pi}\right) \, dx = 0 \quad \text{Zum Vergleich: } \text{Cov}(X, X) = \mathbb{V}(X) = \frac{1}{12}$$

Daher eignen sich die Korrelationskoeffizienten besser als Maß

**Definition 4.26.** Für  $X, Y \in \mathcal{L}^2$  heißt

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

**Korrelationskoeffizient** von  $X$  und  $Y$ .

*Beispiel 4.27.* In Beispiel 4.25 gilt

$$\begin{aligned}\rho(X, Y) &= \frac{\sqrt{15}}{4} \approx 0.968 \\ \rho(X, X) &= 1\end{aligned}$$

**Satz 4.28** (Ungleichungen für Integrale). a) **Jensen-Ungleichung** Sei  $X$  reelle Zufallsvariable,  $\phi(\mathbb{R} \rightarrow \mathbb{R})$  sei konvex.

Sei zusätzlich noch  $\mathbb{E}(|X|) < \infty$  und  $\mathbb{E}(|\phi(X)|) < \infty$ .

Dann gilt

$$\mathbb{E}(\phi(X)) \geq \phi(\mathbb{E}(X))$$

b) **Hölder Ungleichung** Seien  $X, Y$  reelle Zufallsvariablen. Dann gilt für alle  $p, q \in [1, \infty]$  mit  $\frac{1}{p} + \frac{1}{q} = 1$ , dass

$$\mathbb{E}(|XY|) \leq \|X\|_p \|Y\|_q$$

dabei ist

$$\|A\|_p := \begin{cases} \mathbb{E}(|X|^p)^{\frac{1}{p}} & p < \infty \\ \sup\{|X(\omega)| : \omega \in \Omega\} & p = \infty \end{cases}$$

- c) **Cauchy-Schwarz-Ungleichung** Seien  $X, Y \in \mathcal{L}^2$ . Dann ist  $XY \in \mathcal{L}^2$  und

$$\mathbb{E}(|XY|)^2 \leq \mathbb{E}(|X|^2)\mathbb{E}(|Y|^2)$$

(Spezialfall von b) mit  $p = q = 2$ )

- d) **Chebyshev-Markov-Ungleichung** Sei  $X$  reelle Zufallsvariable,  $f : [0, \infty) \rightarrow [0, \infty)$  sei monoton wachsend. Dann gilt  $\forall c > 0$  mit  $f(c) > 0$ , dass

$$\mathbb{P}(|X| \geq c) \leq \frac{\mathbb{E}(f(|X|))}{f(c)}$$

Insbesondere gilt für  $f(x) = x^2$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq c) \leq \frac{\mathbb{V}(X)}{c^2}$$

*Beweis.* a) Da  $\phi$  konvex ist, ist per Definition  $A := \{(x, y) \in \mathbb{R}^2 : y \geq \phi(x)\}$  konvexe Menge.  
 $\Leftrightarrow \forall x \in \mathbb{R}$  gibt es eine Gerade durch  $(x, \phi(x))$ , die das Innere  $A^0$  von  $A$  nicht schneidet.  
 $\Leftrightarrow \forall x \in \mathbb{R} \exists a \in \mathbb{R} : l(y) := \phi(x) + a(y - x) \leq \phi(y)$ .  
Wähle nun  $x = \mathbb{E}(X)$ . Dann ist

- d) Es gilt

$$\begin{aligned} \mathbb{P}(|X| \geq c) &= \mathbb{P}(f(|X|) \geq f(c)) \\ &= \mathbb{P}\left(\frac{f(|X|)}{f(c)} \geq 1\right) = \mathbb{E}\left(\chi_{\left\{\frac{f(|X|)}{f(c)} \geq 1\right\}}\right) \\ &\leq \mathbb{E}\left(\chi_{\left\{\frac{f(|X|)}{f(c)} \geq 1\right\}} \frac{f(|X|)}{f(c)}\right) \\ &\leq \mathbb{E}\left(\frac{f(|X|)}{f(c)}\right) = \frac{1}{f(c)} \mathbb{E}(f(|X|)) \end{aligned}$$

□

*Bemerkung 4.29.* • Sei  $X, Y \in \mathcal{L}^2$ , dann gilt mit 4.28 c), dass auch  $X + \alpha Y \in \mathcal{L}^2$ . Also ist  $\mathcal{L}^2$  ein Vektorraum und  $\|X\|_2 = \mathbb{E}(|X|^2)^{\frac{1}{2}}$  ist eine Halbnorm. Dann ist  $\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$  die „Länge des optimal Vershobenen Vektors“:

$$\mathbb{V}(X) = \min\{\mathbb{E}((X - a)^2) : a \in \mathbb{R}\}$$

und

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

ist ein Skalarprodukt der beiden optimal verschobenen Vektoren.

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

gibt den Winkel zwischen optimal verschobenen Vektoren an.

**Satz 4.30.** a)  $\text{Cov}(X + a, Y + b) = \text{Cov}(X, Y)$  für alle  $a, b \in \mathbb{R}$ .

b)  $\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$

c)  $\mathbb{V}(aX) = a^2 \mathbb{V}(X)$

d) Es ist  $\sum_{i=1}^n X_i \in \mathcal{L}^2$  und

$$\mathbb{V}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{V}(X_i) + \sum_{\substack{i,j=0 \\ i \neq j}}^n \text{Cov}(X_i, X_j) \quad (\text{Binomische Formel})$$

e) Falls  $(X_i)$  paarweise unkorreliert sind, dann

$$\mathbb{V}\left(\sum_{i=1}^{\infty} X_i\right) = \sum_{i=1}^{\infty} \mathbb{V}(X_i) \quad (\text{„Gleichheit von Bienarmer“/„Pythagoras“})$$

**Satz 4.31.** Seien  $X, Y \in \mathcal{L}^1$  und  $X \perp\!\!\!\perp Y$ , dann gilt

a)  $XY \in \mathcal{L}^1$  und  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

b)  $\text{Cov}(X, Y) = 0$

*Beweis.* a) Da  $X \perp\!\!\!\perp Y$  ist folgt mit ??a), dass das Bildmaß von  $\mathbb{P}$  unter  $(X, Y) : \Omega \rightarrow \mathbb{R}^2$  ein Produktmaß, also  $\mathbb{P}_X \otimes \mathbb{P}_Y$  ist.

Nach ??a)]0417satzist

$$\mathbb{E}(|X||Y|) = \int |x||y| \mathbb{P}_{(XY)}(dx dy) = \int_{\mathbb{R}^2} |x||y| \mathbb{P}_X(dx) \mathbb{P}_Y(dy)$$

mit  $h : \mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \mapsto |x||y|$ . Dann folgt mit dem Satz von Fubini

$$= \int_{\mathbb{R}} |x| \mathbb{P}_X(dx) \int_{\mathbb{R}} |y| \mathbb{P}_Y(dy) = \mathbb{E}(|X|) \mathbb{E}(|Y|)$$

Dann folgt dass  $XY \in \mathcal{L}^1$  ist. Also darf die gleiche Rechnung auch ohne  $|\cdot|$  durchgeführt werden.

b) Es gilt

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \stackrel{a)}{=} \mathbb{E}(X - \mathbb{E}(X)) \mathbb{E}(Y - \mathbb{E}(Y)) = 0$$

□

**Bemerkung 4.32.** a) ??e)]0430rechenregeln gilt also auch wenn  $X_i \perp\!\!\!\perp X_j$  für alle  $i \neq j$ .

b) Es kann aber  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$  gelten, ohne dass  $X, Y$  unabhängig sind.

*Beispiel.* Sei  $\Omega = [0, 1]$ ,  $\mathbb{P} = \frac{1}{2} \lambda_{[-1, 1]}$ ,  $X(x) = x$ ,  $Y(x) = x^2$ . Dann ist

$$\mathbb{E}(XY) = \frac{1}{2} \int_{-1}^1 x^3 dx = 0 = \mathbb{E}(X) = \mathbb{E}(X)\mathbb{E}(Y)$$

$$\mathbb{E}(X^2 Y) = \frac{1}{2} \int_{-1}^1 x^4 dx = \frac{1}{5} \quad \text{aber} \quad \mathbb{E}(\mathbb{E}(X^2)) \mathbb{E}(Y) = \left(\frac{1}{3}\right)^3 = \frac{1}{9}$$

Wären  $X \perp\!\!\!\perp Y$ , dann müsste auch  $X^2 \perp\!\!\!\perp Y$  sein, also muss  $X \not\perp\!\!\!\perp Y$ .



*Bemerkung.* Orthogonalität in  $\mathcal{L}^2$  ist viel schwächer (und weniger stabil) als Unabhängigkeit.

## 4.A Erzeugende Funktionen

Für viele Wahrscheinlichkeitsmaße auf  $\mathbb{N}_0$  besteht die Zähldichte aus den Taylor Koeffizienten einer Funktion entwickelt bei  $x = 0$ .

*Beispiel 4.33.* Poisson Verteilung:  $e^{-\lambda} \frac{\lambda^k}{k!}$ . Dann ist

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

Geometrische Verteilung:  $(1-p)^{k-1}p$ . Dann ist

$$\sum_{k=0}^{\infty} (1-p)^{k-1} p u^k = \frac{p}{1-p} \sum_{k=1}^{\infty} ((1-p))^k = \frac{p}{1-p} \frac{1}{1-(1-p)u}$$

**Definition 4.34.** Sei  $\mathbb{P}$  ein Wahrscheinlichkeitsmaß auf  $\mathbb{N}_0$  mit Zähldichte  $(p_k)_{k \geq 0}$ . Die Funktion

$$\varphi_{\mathbb{P}} : [0, 1] \rightarrow \mathbb{R} \quad \varphi_{\mathbb{P}}(u) = \sum_{k=0}^{\infty} p_k u^k$$

heißt **Erzeugende Funktion** von  $\mathbb{P}$  (oder der Folge  $(p_k)$ ).

*Bemerkung 4.35.* a) Da  $\sum_{k=0}^{\infty} p_k = 1$  ist der Konvergenzradius von  $\phi$  größer gleich als 1.

b) Es gilt

$$\varphi_{\mathcal{B}_{n,p}}(u) = \sum_{k=0}^n p^k (1-p)^{n-k} \binom{n}{k} u^k = (pu + (1-p))^n$$

**Definition 4.36.** Sei  $X$  Zufallsvariable mit Werten in  $\mathbb{N}_0$ . Dann heißt

$$\varphi_X(u) = \mathbb{E}(u^X) = \sum_{k=0}^{\infty} \mathbb{P}(X = k) u^k$$

**erzeugende Funktion von  $X$ .**

**Satz 4.37.** Sei  $X$  eine  $\mathbb{N}_0$ -wertige Zufallsvariable. Dann

a)  $\mathbb{P}(X = k) = \frac{1}{k!} \left. \frac{d^k}{du^k} \varphi_X(u) \right|_{u=0}$  für alle  $k \in \mathbb{N}$ .

b)  $\mathbb{E}(X)$  existiert  $\Leftrightarrow \lim_{u \uparrow 1} \phi'(u)$  existiert. Dann ist  $\mathbb{E}(X) = \varphi'_X(1)$ .

c) Für alle  $p \in \mathbb{N}$  ist  $X \in \mathcal{L}^p \Leftrightarrow \lim_{u \uparrow 1} \frac{d^p}{du^p} \varphi(u)$  existiert. Dann ist

$$\frac{d^p}{du^p} \varphi_X(1) = \mathbb{E}(X(X-1) \cdot \dots \cdot (X-p+1))$$

*Beweis.* a) Satz von Taylor: Für alle  $u \leq 1$  (Konvergenzradius)

$$\partial_u^k \varphi(u) = \sum_{n=0}^{\infty} p_n (\partial^n d^k) = \sum_{n=k}^{\infty} p_n \frac{n!}{(n-k)!} u^{k-n}$$

Also gilt bei Auswertung in 0

$$\frac{1}{k!} \partial_n^k \varphi(n) \big|_{u=0} = p_k = \mathbb{P}(X = k)$$

b) Für  $u < 1$  konvergiert die Reihe absolut

$$\sum_{k=0}^{\infty} k u^k \mathbb{P}(X = k) = u \sum_{k=0}^{\infty} (\partial_u u^k) \mathbb{P}(X = k)$$

also darf man gliedweise Differenzieren

$$= u \partial_u \sum_{k=0}^{\infty} u^k \mathbb{P}(X = k) = u \varphi'_X(u)$$

Beide Seiten wachsen Monoton, wenn  $u \uparrow 1$ :

$$\text{Linke Seite } \uparrow_{u \uparrow 1} \mathbb{E}(X) \quad \text{Rechte Seite } \uparrow_{u \uparrow 1} \varphi'(1)$$

c) Anwenden des gleichen Arguments mit

$$\sum_{k \geq p} k(k-1) \dots (k-p+1) u^k \mathbb{P}(X = k) = u^p \sum_{k=0}^{\infty} \partial_u^p (u^k) \mathbb{P}(X = k)$$

gegen  $\mathbb{E}(X(X-1) \dots (X-p+1))$  bzw.  $1^p \partial_u^p \varphi_X(1)$ .

□

*Beispiel 4.38.* Sei  $X \sim \text{Poi}_\lambda$ : Dann ist

$$\mathbb{E}(X) = \partial_u (e^{\lambda u - \lambda}) \big|_{u=1} = \lambda e^0 = \lambda$$

$$\mathbb{E}(X(X-1)) = \partial_u^2 (e^{\lambda u - \lambda}) = \lambda^2$$

Also gilt für die Varianz:

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X(X-1) + \mathbb{E}(X) - \mathbb{E}(X)^2) = \lambda$$

**Satz 4.39.** Seien  $X, Y$  unabhängige  $\mathbb{N}_0$ -verteilte Zufallsvariablen. Dann gilt

$$\varphi_{X+Y}(u) = \varphi_X(u) \varphi_Y(u)$$

*Beweis.* Sei  $X \sim \text{Poi}_\lambda, Y \sim \text{Poi}_\mu, Y \perp\!\!\!\perp X$ . Dann ist

$$\varphi_{X+Y}(u) = \varphi_X(u) \varphi_Y(u) = e^{\lambda u - \lambda} e^{\mu u - \mu} = e^{(\mu + \lambda)u - (\mu + \lambda)}$$

Also  $X + Y \sim \text{Poi}_{\lambda + \mu}$ .

□

*Bemerkung 4.40* (Ausblick). a) Die rechte Seite in 4.37c) ist etwas unhandlich. Daher definiere für  $t > 0$

$$L_X(t) := \varphi_X(e^{-t}) = \mathbb{E}((e^{-t})^X) = \mathbb{E}(e^{-tX})$$

(„Laplace Transformierte von  $\mathbb{P}_X$ “)

Dann ist

$$\mathbb{E}(X^p) = \left. \frac{d^p}{dt^p} L_X \right|_{t=0}$$

falls  $\mathbb{E}(X^p) < \infty$ .

- b) Das geht auch für reelle Zufallsvariablen. Sei  $X : \Omega \rightarrow \mathbb{R}$ . Es sei  $\mathbb{E}(e^{-tX}) < \infty$  für ein  $c > 0$ . Dann ist

$$L_X(t) = \mathbb{E}(e^{-tX}) = \int_{-\infty}^{\infty} e^{-tX} \mathbb{P}_X(dy)$$

für alle  $t \in \{\tilde{t} \in \mathbb{R} \mid \mathbb{E}(e^{-\tilde{t}X}) < \infty\}$ .

Wieder gilt

$$\mathbb{E}(X^p) = (-1)^p \left. \partial_t^p L_X(t) \right|_{t=0}$$

und es gilt auch  $L_X$  bestimmt  $P_X$  eindeutig.

Satz 4.39 gilt mit  $L_{X+Y} = L_X L_Y$ . Außerdem ist  $L_{\alpha X}(t) = \mathbb{E}(e^{-t\alpha X}) = L_X(\alpha t)$ , falls alles  $< \infty$ .

- c) Für unabhängige reelle Zufallsvariablen  $X_i$  mit  $\mathbb{P}_{X_i} = \mathbb{P}_{X_j}$  gilt

$$(**) = L_{\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i}(t) = \prod_{i=1}^n L_{\frac{X_i}{\sqrt{n}}}(t) = \prod_{i=1}^n L_{X_i}\left(\frac{t}{\sqrt{n}}\right) = \left(L_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

Es gilt immer  $L_X(0) = 1$  und  $L'_X(0) = \mathbb{E}(X)$ .

Angenommen  $\mathbb{E}(X) = 0$  und  $L''_X(0) = \mathbb{E}(X^2) < \infty$ , dann gilt die Taylor-Gleichung:

$$\ln(L_X(s)) = \ln(L_X(0) + sL'_X(0) + \mathcal{O}(s^3)) = \frac{1}{2}s^2L''_X(0) + \mathcal{O}(s^3)$$

Es folgt, dass

$$(**) = e^{n \ln L_X\left(\frac{t}{\sqrt{n}}\right)} = e^{n\left(\frac{1}{2}\frac{t^2}{n}L''_X(0) + \mathcal{O}\left(\frac{t^3}{n^{3/2}}\right)\right)} \xrightarrow{n \rightarrow \infty} e^{\frac{t^2}{2}\mathbb{E}(X^2)}$$

## 5 Grenzwertsätze

Wir haben  $\mathbb{E}(X) = \sum_{k \in \mathbb{Z}} k \mathbb{P}(X = k)$  über die relative Häufigkeit motiviert:

$$\mathbb{E}(X) \approx \frac{1}{N} \sum_{i=1}^N X_i(\omega) \approx \sum_{k \in \mathbb{Z}} \frac{|\{i : X_i(\omega) = k\}|}{N} k$$

Wir werden nun begründen, dass dies Intuition richtig ist.

**Definition 5.1.** Eine Folge  $(X_n)$  von Zufallsvariablen heißt **unabhängig identisch Verteilt (uiv)**, wenn

(i) die  $X_n$  sind paarweise unabhängig

(ii)  $\mathbb{P}_{X_i} = \mathbb{P}_{X_j}$  für alle  $i, j \in \mathbb{N}$ .

**Satz 5.2** (Schwache Gesetz der großen Zahlen (GGZ), Basisversion). *Seien  $(X_i)$  i.i.v., reelle Zufallsvariable mit  $\mathbb{V}(X_i) < \infty$ . Dann ist für alle  $\epsilon > 0$*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_1) \right| \geq \epsilon \right) = 0$$

*Beweis.* Setze  $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$ , dann ist

$$\mathbb{E}(Y_n) = \mathbb{E} \left( \frac{1}{n} \sum_{i=1}^n X_i \right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \mathbb{E}(X_1)$$

$$\mathbb{V}(Y_n) = \mathbb{V} \left( \frac{1}{n} \sum_{i=1}^n X_i \right) \stackrel{4.30}{=} \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(X_i) = \frac{1}{n} \mathbb{V}(X_1)$$

Mit 4.28d) (Chebyshev-Markov-Ungleichung) folgt

$$\mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_1) \right| > \epsilon \right) = \mathbb{P}(|Y_n - \mathbb{E}(Y_n)| > \epsilon) \leq \frac{1}{\epsilon^2} \mathbb{V}(Y_n) = \mathbb{V}(X_1) \frac{1}{n\epsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

□

**Satz 5.3** (Schwache Gesetz der großen Zahlen,  $\mathcal{L}^2$ -Version). *Seien  $(X_i)$  (nicht notwendigerweise i.i.v.) reelle Zufallsvariablen für die gilt:*

(i) **Paarweise unkorreliert**  $\text{Cov}(X_i, X_j) = 0$  für alle  $i \neq j$ .

(ii) **Beschränkte Varianzen**  $v := \sup \{\mathbb{V}(X_i), i \in \mathbb{N}\} < \infty$

Dann gilt für alle  $n \in \mathbb{N}$  und alle  $\epsilon_n > 0$

$$\mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}(X_i)) \right| \geq \epsilon_n \right) \leq \frac{v}{n\epsilon_n^2}$$

*Beweis.* Analog, Übung

□

**Satz 5.4** (Schwache Gesetz der großen Zahlen,  $\mathcal{L}^1$ -Version). *Seien  $(X_i)$  paarweise unabhängige, identisch verteilte Zufallsvariablen und  $\mathbb{E}(|X_1|) < \infty$ . Dann gilt für alle  $\epsilon > 0$*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_1) \right| \geq \epsilon \right) = 0$$

*Beweis.* Zerlege

$$X_i(\omega) = \underbrace{X_i(\omega) \chi_{\{|X_i(\omega)| < \sqrt[4]{i}\}}(\omega)}_{Y_i(\omega)} + \underbrace{X_i(\omega) \chi_{\{|X_i(\omega)| \geq \sqrt[4]{i}\}}(\omega)}_{Z_i(\omega)}$$

Dann gilt

a)  $\mathbb{E}(X_i) = \mathbb{E}(Y_i) + \mathbb{E}(Z_i)$

b) mit der Dreiecksungleichung

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^n X_i(\omega) - \mathbb{E}(X_1) \right| &= \frac{1}{n} \left| \sum_{i=1}^n (Y_i(\omega) - \mathbb{E}(Y_i) + Z_i(\omega) - \mathbb{E}(Z_i)) \right| \\ &\leq \frac{1}{n} \left| \sum_{i=1}^n (Y_i(\omega) - \mathbb{E}(Y_i)) \right| + \frac{1}{n} \left| \sum_{i=1}^n (Z_i(\omega) - \mathbb{E}(Z_i)) \right| \end{aligned}$$

Daher ist

$$\underbrace{\left\{ \omega \in \Omega : \left| \frac{1}{n} \sum_{i=1}^n X_i(\omega) - \mathbb{E}(X_i) \right| \geq \epsilon \right\}}_{A_n} \subset \underbrace{\left\{ \omega \in \Omega : \left| \frac{1}{n} \sum_{i=1}^n Y_i(\omega) - \mathbb{E}(Y_i) \right| \geq \frac{\epsilon}{2} \right\}}_{B_n} \cup \underbrace{\left\{ \omega \in \Omega : \left| \frac{1}{n} \sum_{i=1}^n Z_i(\omega) - \mathbb{E}(Z_i) \right| \geq \frac{\epsilon}{2} \right\}}_{C_n}$$

und damit  $\mathbb{P}(A_n) \leq \mathbb{P}(B_n) + \mathbb{P}(C_n)$ .

Wegen ??a)]0318satz sind auch die  $Y_i$  paarweise unabhängig. Sei nun  $\bar{Y}_n := \frac{1}{n} \sum_{i=1}^n Y_i$ , dann ist

$$\begin{aligned} \mathbb{P}(B_n) &= \mathbb{P} \left( |\bar{Y}_n - \mathbb{E}(\bar{Y}_n)| \geq \frac{\epsilon}{2} \right) \stackrel{4.28d)}{\leq} \frac{4}{\epsilon^2} \mathbb{V}(\bar{Y}_n) = \frac{4}{\epsilon^2} \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(Y_i) \\ &\leq \frac{4}{n^2 \epsilon^2} \sum_{i=1}^n \mathbb{E}(Y_i^2) \leq \frac{4}{n^2 \epsilon^2} \sum_{i=1}^n \sqrt{i} \leq \frac{4}{\epsilon^2 n^2} n \sqrt{n} = 4 \frac{\epsilon^2}{\sqrt{n}} \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

Analog für  $C_n$  mit Sei  $\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i$ . Dann ist

$$\begin{aligned} \mathbb{P}(C_n) &= \mathbb{P} \left( |\bar{Z}_n - \mathbb{E}(\bar{Z}_n)| \geq \frac{\epsilon}{2} \right) \stackrel{4.28d)}{\leq} \frac{f(x)=|x|}{\epsilon} \mathbb{E}(|\bar{Z}_n - \mathbb{E}(\bar{Z}_n)|) = \frac{2}{n\epsilon} \mathbb{E} \left( \left| \sum_{i=1}^n Z_i - \mathbb{E}(Z_i) \right| \right) \\ &\leq \frac{2}{n\epsilon} \sum_{i=1}^n \mathbb{E}(|Z_i - \mathbb{E}(Z_i)|) \leq \frac{2}{n\epsilon} \sum_{i=1}^n 2\mathbb{E}(Z_i) \end{aligned}$$

Da aber falls  $a_i \xrightarrow{i \rightarrow \infty} 0$  geht auch  $\frac{1}{n} \sum_{i=1}^n a_i \xrightarrow{in \rightarrow \infty} 0$ . Also

$$\mathbb{P}(C_n) \leq \frac{2}{n\epsilon} \sum_{i=1}^n 2\mathbb{E}(Z_i) \xrightarrow{n \rightarrow \infty} 0$$

□

*Beispiel 5.5* (Satz von Weierstraß). Sei  $f : [0, 1] \rightarrow \mathbb{R}$  stetig.

Zu  $p \in [0, 1]$  seien  $(X_i)_{i \in \mathbb{N}}$  uiv Zufallsvariablen mit  $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$ . Dann ist  $\sum_{i=1}^n X_i \sim \mathcal{B}_{n,p}$  und

$$f_n(p) := \mathbb{E}_p \left( f \left( \frac{1}{n} \sum_{i=1}^n X_i \right) \right) = \sum_{k=0}^n f \left( \frac{k}{n} \right) \mathbb{P}_p \left( \sum_{i=1}^n X_i = k \right) = \sum_{k=0}^n f \left( \frac{k}{n} \right) \binom{n}{k} p^k (1-p)^{n-k}$$

ist Polynom  $n$ -ten Grades in  $p$  („Bernstein-Polynome“).

$f$  ist gleichmäßig stetig in  $[0, 1]$ , also  $\forall \epsilon > 0 \exists \delta > 0$ , sodass für alle  $|x - y| < \delta$  immer  $|f(x) - f(y)| < \epsilon$  gilt.

Außerdem ist

$$f_n(p) - f(p) = \mathbb{E}_p \left( f \left( \frac{1}{n} \sum_{i=1}^n X_i \right) - f(p) \right)$$

Also kann man abschätzen:

$$\begin{aligned} |f_n(p) - f(p)| &\leq \left| \mathbb{E}_p \left( f \left( \frac{1}{n} \sum_{i=1}^n X_i \right) - f(p) \right) \right| \leq \mathbb{E}_p \left( \left| f \left( \frac{1}{n} \sum_{i=1}^n X_i \right) - f(p) \right| \right) \\ &\leq \epsilon + 2 \|f\|_\infty \mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n X_i - p \right| \geq \delta \right) \stackrel{5.3}{\leq} \epsilon + 2 \|f\|_\infty \frac{\mathbb{V}(X_i)}{n\delta^2} \end{aligned}$$

Also ist  $\limsup_{n \rightarrow \infty} |f_n(p) - f(p)| \leq \epsilon$  für alle  $\epsilon > 0$ .

Also ist  $\lim_{n \rightarrow \infty} |f_n(p) - f(p)| = 0$ .

*Bemerkung.*

*Beispiel 5.6.* Sei  $\Omega = [0, 1]$  mit dem Lebesgue-Maß, Sei  $X_i(\omega) = \chi_{I_i}(\omega)$ . für  $i = 1, \dots, 15$  mit

$$\begin{array}{lll} I_1 = [0, \frac{1}{2}) & & I_1 = [\frac{1}{2}, 1) \\ I_3 = [0, \frac{1}{4}) & \dots & I_6 = [\frac{3}{4}, 1) \\ I_7 = [0, \frac{1}{8}) & \dots & I_{15} = [\frac{7}{8}, 1) \end{array}$$

Dann ist

$$\mathbb{P}(|X_n - 0| > \epsilon) = \begin{cases} \frac{1}{2} & n = 1, 2 \\ \frac{1}{4} & n = 3, 4, 5, 6 \\ \frac{1}{8} & n = 7, \dots, 15 \end{cases} \leq \frac{2}{n} \xrightarrow{n \rightarrow \infty} 0$$

Aber  $\limsup_{n \rightarrow \infty} |X_n(\omega) - 0| = 1 \quad \forall \omega \in \Omega$ .

Also  $X_n(\omega) \not\xrightarrow{n \rightarrow \infty} 0$ .

**Definition 5.7.** Seien  $(S_n)$  Zufallsvariablen. Man sagt für  $c \in \mathbb{R}$ , dass  $S_n$  konvergiert  $\mathbb{P}$  fast sicher ( $S_n \xrightarrow{n \rightarrow \infty} c$ ), wenn

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} S_n = c \right) = 1 \quad (\text{Gl. 5.2})$$

*Bemerkung 5.8.* Setze  $\Omega_m := \{\omega \in \Omega : \limsup_{n \rightarrow \infty} |S_n - c| < \frac{1}{m}\}$ .

Da  $\Omega_{m+1} \subseteq \Omega_m$ , also

$$\Omega_\infty := \left\{ \omega \in \Omega : \limsup_{n \rightarrow \infty} S_n = c \right\}$$

mit ??e). Dann gilt

$$\mathbb{P} \left( \limsup_{n \rightarrow \infty} S_n = c \right) = 1 \Leftrightarrow \forall \epsilon > 0: \mathbb{P} \left( \limsup_{n \rightarrow \infty} |S_n - c| < \epsilon \right) = 1 \quad (\star\star\star)$$

Daher besteht der Unterschied zwischen ?? und Gl. 5.2 in der Platzierung der Grenzwerts:  $\forall \epsilon, c > 0$

- ?? besagt  $\lim_{n \rightarrow \infty} \mathbb{P}(|S_n - c| < \epsilon) = 1$
- Gl. 5.2 besagt  $\mathbb{P}(\limsup_{n \rightarrow \infty} |S_n - c| < \epsilon) = 1$

Außerdem gilt das Lemma von Fatou 4.15:  $\mathbb{E}(\liminf_{n \rightarrow \infty} X_n) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(X_n)$ , falls  $X_n \geq 0$ .

Setze nun  $X_n = \chi_{\{|S_n - c| < \epsilon\}}$ . Dann ist

$$\begin{aligned} \liminf_{n \rightarrow \infty} (X_n \omega) = 1 &\Leftrightarrow \exists m \in \mathbb{N} : \forall n > m \chi_{\{|S_n - c| < \epsilon\}}(\omega) = 1 \\ &\Leftrightarrow |S_n - c| < \epsilon \text{ unendlich oft} \Leftrightarrow \limsup_{n \rightarrow \infty} |S_n - c| < \epsilon \end{aligned}$$

Dann folgt (Fatou)

$$\liminf_{n \rightarrow \infty} \mathbb{P}(|S_n - c| < \epsilon) = \liminf_{n \rightarrow \infty} \mathbb{E}(X_n) \geq \mathbb{E}(\liminf_{n \rightarrow \infty} X_n) = \mathbb{E}(\limsup_{n \rightarrow \infty} |S_n - c| < \epsilon)$$

Daher folgt aus  $\mathbb{P}(\lim_{n \rightarrow \infty} |S_n - c| = 0)$  mit  $\star\star\star$ , dass  $\forall \epsilon > 0. \mathbb{P}(\limsup_{n \rightarrow \infty} |S_n - c| < \epsilon)$ .

Dann folgt mit Fatou:  $\forall \epsilon > 0 \lim_{n \rightarrow \infty} \mathbb{P}(|S_n - c| < \epsilon) = 1$ .

Also ist Gl. 5.2 stärker als Gl. 5.2.

**Satz 5.9** (Starkes Gesetz der großen Zahlen,  $\mathcal{L}^4$ -Version). *Seien  $X_n$  Zufallsvariablen mit  $\mathbb{E}(|X_n|^4) < \infty$ , dann ist*

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mathbb{E}(X_1)\right) = 1$$

*Beweis.* Wegen  $\frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_1) = \frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}(X_i))$  kann man  $\mathbb{E}(X_1) = 0$  annehmen.

(i) Sei

$$A_n = \left\{ \omega \in \Omega \mid \left| \frac{1}{n} \sum_{i=1}^n X_i(\omega) \right| \geq n^{-\frac{1}{8}} \right\}$$

Es gilt

$$\begin{aligned} \mathbb{P}(A_n) &= \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n X_i \right| \geq n^{-\frac{1}{8}}\right) \\ &\stackrel{4.28}{\text{„Chebychev } f(x) = x^4\text{“}} \leq \frac{\mathbb{E}\left(\left(\frac{1}{n} \sum_{i=1}^n X_i\right)^4\right)}{\left(n^{-\frac{1}{8}}\right)^4} = n^{\frac{1}{2}} n^{-4} \mathbb{E}\left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n X_i X_j X_k X_l\right) \\ &= n^{\frac{1}{2}-4} \sum_{i,j,k,l=1}^n \mathbb{E}(X_i X_j X_k X_l) \end{aligned}$$

$\mathbb{E}(\dots) \neq 0$ , nur wenn je zwei oder alle Indizes gleich sind.

$$\begin{aligned}
&= n^{\frac{1}{2}-4} \left( \sum_{i=1}^n \mathbb{E}(X_i^4) + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \mathbb{E}(X_i^2 X_j^2) \right) \\
&= n^{\frac{1}{2}-4} (n \mathbb{E}(X_1^4) + n(n-1) \mathbb{E}(X_1^2)^2) \leq cn^{-\frac{3}{2}}
\end{aligned}$$

Also ist  $\sum_{i=1}^{\infty} \mathbb{P}(A_n) \leq \infty$ .

(ii) Dann folgt mit dem Lemma von Borel-Cantelli (3.35):

$$\begin{aligned}
0 &= \mathbb{P} \left( \limsup_{n \rightarrow \infty} A_n \right) = \mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n X_i \right| \geq n^{-\frac{1}{8}} \text{ unendlich oft} \right) \\
&= 1 - \underbrace{\mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n X_i \right| \geq n^{-\frac{1}{8}} \text{ nur endlich oft} \right)}_{\geq \mathbb{P} \left( \lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{i=1}^n X_i \right| = 0 \right)} \\
&\geq 1 - \mathbb{P} \left( \lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{i=1}^n X_i \right| = 0 \right)
\end{aligned}$$

Dann ist also  $\mathbb{P} \left( \lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{i=1}^n X_i \right| = 0 \right) \geq 1$ , also  $= 1$ .

□

*Bemerkung 5.10.* Die aussage von 5.9 gilt auch falls  $X \in \mathcal{L}^1$ .

*Bemerkung 5.11.* Rückblick auf 5.3:

Falls  $X_i$  i.i.d., dann ist

$$\mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_1) \right| \geq \epsilon_n \right) \leq \frac{V}{n \epsilon_n^2}$$

Also gilt  $\mathbb{P}(\dots) \xrightarrow{n \rightarrow \infty} 0$ , falls  $\epsilon_n \gg \frac{1}{\sqrt{n}}$  (Also  $\limsup_{n \rightarrow \infty} \sqrt{n} \epsilon_n = \infty$ ).

Das bedeutet: Der Abstand zwischen  $\frac{1}{n} \sum_{i=1}^n X_i$  und  $\mathbb{E}(X_1)$  schrumpft „fast“ wie  $\frac{1}{\sqrt{n}}$ .

Frage: Was ist  $\mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}(X_i)) \right| \geq \frac{1}{\sqrt{n}} \right)$ ?

Für  $\alpha > \frac{1}{2}$  gilt

$$\mathbb{V} \left( n^{\alpha-1} \sum_{i=1}^n (X_i - \mathbb{E}(X_i)) \right) = n^{2\alpha-2} \sum_{i=1}^n \mathbb{V}(X_1)$$

Sei  $v = n^{2\alpha-1}$ , dann  $v \xrightarrow{n \rightarrow \infty}$ , falls  $\alpha > \frac{1}{2}$  und  $n > 0$ . Also ist zu erwarten, dass für  $\alpha > \frac{1}{2}$  gilt

$$\mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_i) \right| \leq n^{-\alpha} \right) \not\rightarrow 0$$

Tatsächlich gilt



**Satz 5.12** (zentraler Grenzwertsatz). Sei  $X_n$  eine Folge von unabhängigen, identisch verteilten Zufallsvariablen,  $\mathbb{E}(X_i) = m$ ,  $\mathbb{V}(X_i) = v \leq \infty$  für alle  $i \in \mathbb{N}$ .

Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - m) \leq c \right) = \int_{-\infty}^c \frac{1}{\sqrt{2\pi v}} e^{-\frac{x^2}{2v}} dx$$

*Bemerkung 5.13.* Die Gleichung aus 5.12 bedeutet, dass die Folge der Bildmaße  $\mathbb{P}_{\overline{S}_n(\omega)}$  von  $\mathbb{P}$  unter  $\omega \mapsto \overline{S}_n(\omega) := \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i(\omega) - m)$  konvergiert gegen  $\mathcal{N}_{0,v}$ . Im Sinne der Maßkonvergenz

$$\mathbb{P}_{\overline{S}_n}((-\infty, c]) \xrightarrow{n \rightarrow \infty} \mathbb{P}_{\mathcal{N}_{0,v}}((-\infty, c]) \forall c$$

**Definition 5.14.** Eine Folge  $\mu_n$  von endlichen Maßen auf  $\mathbb{R}$  heißt **schwach konvergent** gegen ein Maß  $\mu$ , falls  $\forall f \in C_0(\mathbb{R})$

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mu_n}(f) = \mathbb{E}_{\mu}(f) \quad \text{bzw.} \quad \lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu$$

*Bemerkung.* In 5.12 verwenden wir  $f(x) = \chi_{(-\infty, c]}(x)$ . Da  $f$  unstetig ist, gilt z.B. mit  $\mu_n = \delta_{\frac{1}{n}}$ ,  $g = \chi_{(-\infty, 0]}$ ,  $\mu = \delta_0$ . Dann ist

$$\int f d\mu_n = f\left(\frac{1}{n}\right) \xrightarrow{n \rightarrow \infty} f(0)$$

falls  $f$  stetig differenzierbar bei 0, aber  $\int g d\mu_n = 0$ ,  $\int g d\mu = 1$ .

**Proposition 5.15.** Seien  $\mu_n$  Wahrscheinlichkeitsmaße auf  $\mathbb{R}$ ,  $\mu$  Wahrscheinlichkeitsmaß auf  $\mathbb{R}$  und  $\mu$  habe Lebesgue-Dichte  $\rho$  und sei für alle  $k \in \mathbb{N}$

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mu_n}(f) = \mathbb{E}_{\mu}(f) \forall f \in C_b^k$$

Dann ist

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\mu_n}((-\infty, c]) = \mathbb{P}_{\mu}((-\infty, c]) \forall c \in \mathbb{R}$$

*Beweis.* (i) Zu zeigen:  $\forall x \in \mathbb{R} \forall \epsilon > 0 \exists \delta > 0$

$$\limsup_{n \rightarrow \infty} \mathbb{P}_{\mu_n}([x - \delta, x + \delta]) < \epsilon$$

Wähle dazu  $f \in C_b^k$  mit  $0 \leq f \leq 1$ ,  $f(y) = 1 \forall y \in [x - \delta, x + \delta]$  und  $f(y) = 0 \forall y \notin [x - 2\delta, x + 2\delta]$ .

Dann gilt

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{P}_{\mu_n}([x - \delta, x + \delta]) &\leq \limsup_{n \rightarrow \infty} \mathbb{E}_{\mu_n}(f) = \mathbb{E}_{\mathbb{P}_X} \mu(f) \\ &= \int_{x-2\delta}^{x+2\delta} f(y) \rho(y) dy \leq \int_{x-2\delta}^{x+2\delta} \rho(y) dy \xrightarrow{\delta \rightarrow 0} 0 \end{aligned}$$

mit 4.11, da  $\rho \in \mathcal{L}^1$ .

- (ii) Zu  $\epsilon > 0$ ,  $c \in \mathbb{R}$  wähle  $f(x) = 1 \forall x \leq c$ .  
Dann ist

$$\mathbb{P}_{\mu_n}((-\infty, c]) - \mathbb{P}_{\mu}((-\infty, c]) = \mathbb{E}_{\mu_n}(\chi_{(-\infty, c]} - f) + \underbrace{\mathbb{E}_{\mu_n}(f) - \mathbb{E}_{\mu}(f) - \mathbb{E}_{\mu}(\chi_{(-\infty, c]} - f)}_{|\cdot| \leq \chi_{(-\infty, c]} \quad (\star)}$$

Mit (i) folgt,  $\delta$  existiert, sodass  $|(\star)| \leq 2\epsilon + |\mathbb{E}_{\mu_n}(f) - \mathbb{E}_{\mathbb{P}_X \mu}(f)|$ . Also ist  
 $\limsup_{n \rightarrow \infty} |(\star)| \leq 2\epsilon \forall \epsilon > 0$ . □

*Beweis von 5.12.* Setze  $\tilde{X}_i = \frac{1}{\sqrt{v}}(X_i - m)$ . Dann ist  $\mathbb{E}(\tilde{X}_i) = 0$  und  $\mathbb{V}(\tilde{X}_i) = 1$  und

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - m) \geq c\right) = \mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{X}_i \geq c\sqrt{v}\right)$$

Also kann oBdA angenommen werden, dass bereits  $\mathbb{E}(X - i) = 0$  und  $\mathbb{V}(X_i) = 1$  gilt.

Sei nun  $\bar{S}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ .

Mit 5.15 folgt, dass es bereits ausreicht, dass zu prüfen, ob (mit  $k = 3$ )  $\lim_{n \rightarrow \infty} \mathbb{E}(f(\bar{S}_n)) = \mathbb{E}_{\mathcal{N}_{0,1}}(f)$  für alle  $f \in C_b^3(\mathbb{R})$  gilt.

Sei dazu  $(Y_n)$  eine weitere Folge von unabhängig identisch verteilten Zufallsvariablen mit

- a)  $(Y_n)$  unabhängig von  $(X_n)$
- b)  $(Y_n \sim \mathcal{N}_{0,1}) \forall n$

Dann gilt für  $\bar{T}_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \sim \mathcal{N}_{0,1}$ , dass

$$\mathbb{E}(f(\bar{T}_n)) = \mathbb{E}_{\mathcal{N}_{0,1}}(f)$$

Also gilt es zu zeigen, dass

$$\forall f \in C_b^3 \quad \mathbb{E}(f(\bar{S}_n) - f(\bar{T}_n)) \xrightarrow{n \rightarrow \infty} 0$$

Dabei gilt

$$\begin{aligned} & -f(\bar{S}_n(\omega)) - f(\bar{T}_n) \\ &= \sum_{i=1}^n \left[ f\left(\frac{1}{\sqrt{n}} \left( \sum_{j=1}^i X_j(\omega) + \sum_{j=i+1}^n Y_j(\omega) \right)\right) - f\left(\frac{1}{\sqrt{n}} \left( \sum_{j=1}^{i-1} X_j(\omega) + \sum_{j=i}^n Y_j(\omega) \right)\right) \right] \\ &= \sum_{i=1}^n f \end{aligned}$$

□

## 6 Grundzüge der Statistik

### 6.A Fragestellung und Modellbildung

*Bemerkung 6.1.* a) Ausgangspunkt: Daten  $x = (x_1, \dots, x_n) \in \mathbb{R}$  oder allgemein  $x \in \Omega$ . Diese Daten sind fest vorgegeben, nicht zufällig und nicht verhandelbar (z.B. Messwerte).

b) Modellannahme:  $x_1, \dots, x_n$  sind das Ergebnis eines (unbekannten) zufälligen Geschehens. Also gibt es einen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, \mathbb{P})$  mit  $x \in \Omega$ , sodass  $x$  das Ergebnis von „einmal würfeln mit  $\mathbb{P}$ “ ist.

c) Ziel: Finde heraus welches  $\mathbb{P}$  (äquivalent: welche Zufallsvariable  $X$ ) die Daten  $x$  erzeugt haben könnte.

d) Häufte zusätzliche Annahme: Es kommt in b) nur eine Familie  $(P_\theta)_{\theta \in \Theta}$  mit  $\mathbb{P}_\theta = \mathcal{N}(\mu, \sigma^2)$  in Frage.

$\theta$  heißt Parameter. Daher spricht man von „parametrische Statistik“.

*Beispiel 6.2* (Feuerwerk). Großpackung  $10^6$  Raketen. Davon werden 100 getestet und 5 davon sind defekt.

Wie viele der  $10^6$  Raketen sind voraussichtlich Defekt (Ausschussquote)?

a) Daten:  $x = 5$

b) Modell (Approximativ) ziehen mit zurücklegen (eigentlich: ohne) mit Erfolgswahrscheinlichkeit  $\theta = \frac{|\{\text{defekte}\}|}{10^6}$ .

d)  $\Theta = [0, 1]$ ,  $\mathbb{P}_\theta = \mathcal{B}_{N, \theta}$  binomialverteilt  $N = 100$ , Erfolgswahrscheinlichkeit  $\theta$ .

c) Welches  $\theta$  hat die Daten erzeugt.

*Beispiel 6.3* („6.2“ abstrahiert). Ein Ratespiel. Der Spielleiter hat für jedes  $\theta \in \Theta$  einen Würfel mit der Verteilung  $\mathbb{P}_\theta$ . Er nimmt einen davon, würfelt einmal und sagt das Ergebnis.

Nun muss man möglichst gut Raten welcher Würfel genommen wurde.

**Definition 6.4.** Ein **statistisches Modell** (sM) ist ein Tripel  $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$ , wobei  $(\Omega, \mathcal{F})$  ein Messraum ist und  $\Theta$  die sog. „**Parametermenge**“ und  $\mathbb{P}_\theta$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{F})$  für alle  $\theta \in \Theta$  ist.

**Definition 6.5.** Ein statistisches Modell heißt **Produktmodell**, falls  $n \in \mathbb{N}$  existiert, sodass für  $(\tilde{\Omega}, \tilde{\mathcal{F}})$  Messraum,  $\tilde{\mathbb{P}}_\theta$  Familie von Wahrscheinlichkeitsmaßen auf  $\tilde{\Omega}$

$$(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta}) = (\tilde{\Omega}^n, \tilde{\mathcal{F}}^n, (\tilde{\mathbb{P}}_\theta^{\otimes n})_{\theta \in \Theta})$$

ein  $n$ -faches unabhängiges Würfeln Modelliert.

*Beispiel 6.6.* In Beispiel 6.2 alternativ:

$$\Omega = \{0, 1\}^{100}, \quad \mathcal{F} = \mathcal{P}(\{0, 1\})^{\otimes 100}, \quad \mathbb{P}_\theta = (\text{Ber}_{\{0, 1\}, \theta})$$

Typische Daten:  $x = (0, 0, 1, 0, \dots, 1)$ . Interessant ist die Anzahl der Einsen.

**Definition 6.7.**  $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$  sei statistisches Modell. Eine messbare Abbildung  $T : \Omega \rightarrow \Theta$  heißt **Schätzer** für den Parameter  $\theta \in \Theta$ .

Eine messbare Abbildung  $\Omega \rightarrow \Omega'$  heißt  $\Omega'$ -wertiger Schätzer.

Bedeutung: Bei Daten  $x \in \Omega$  „schätzt“ man dass  $\theta = T(x)$ . Ein  $\Omega'$ -wertiger Schätzer ist sinnvoll, falls z.B.  $\Theta = \mathbb{R} \times \mathbb{R}^+$ ,  $\theta = (\mu, \sigma^2)$ ,  $\mathbb{P}_\theta = \mathcal{N}(\mu, \sigma^2)$  man aber nur  $\mu$  wissen will. Dann ist  $\Omega' = \mathbb{R}$ .

*Bemerkung.* 6.7 verlangt nicht, dass der geschätzte Wert  $T(x)$  etwas mit dem wahren  $\theta$  zu tun hat. Im Falle eines Produktmodells hat man aber folgendes Gütekriterium:

**Definition 6.8.** Zu jedem  $n \in \mathbb{N}$  sei  $(\Omega^n, \mathcal{F}^{\otimes n}, (\mathbb{P}_{(X,Y)\theta}^{\otimes n})_{\theta \in \Theta})$  ein Produktmodell  $T_N : \Omega \rightarrow \Omega'$  ein Schätzer,  $f : \Theta \rightarrow \Omega'$  eine Funktion:  $(T_n)_{n \in \mathbb{N}}$  heißt **schwach konsistente Familie von Schätzern**, falls

$$\forall \theta \in \Theta, \forall \epsilon > 0 : \lim_{n \rightarrow \infty} \mathbb{P}_{\theta^{\otimes n}}(|T_n - f(\theta)| > \epsilon) = 0$$

*Bemerkung* 6.9. 6.8 bedeutet nicht, dass für gegebene Daten  $x$  und  $\epsilon > 0$  und großem  $n$  der Wert  $T(x)$  mit hoher Wahrscheinlichkeit höchstens Abstand  $\epsilon$  von  $f(\theta)$  hat.

Denn  $x$  und  $\theta$  sind fest (obwohl  $\theta$  unbekannt ist). Hier kommt kein Zufall in Spiel.

Stattdessen: Egal welches  $\theta$  das richtige ist liegt  $T(x)$  nur dann weit von  $f(\theta)$  weg (für große  $n$ ) wenn wir beim Erzeugen der Daten „Pech hatten“.

Mit anderen Worten: Würden wir die Messung häufig wiederholen und Datensätze  $x^{(1)}, \dots, x^{(m)}$  erzeugen, so wäre für die meisten dieser Datensätze die Ungleichung  $|T(x^{(k)}) - f(\theta)| \leq \epsilon$  wahr.

*Beispiel* 6.10. In 6.2 wählt man  $T(x) = \frac{x}{100}$  bzw in der Form ?? :  $T(x) = \frac{1}{100} \sum_{i=1}^{100} x_i$ .

Speziell für  $x = 5$  ist  $T(x) = \frac{1}{20}$ , d.h. man schätzt 5% Ausschuss.

**Definition 6.11.** Sei  $(\Omega^n, \mathcal{F}^{\otimes n}, (\mathbb{P}_\theta)_{\theta \in \Theta}^{\otimes n})$  für alle  $n$  ein Produktmodell. Sei  $\mathbb{P}_\theta$  die Verteilung einer reellen Zufallsvariable  $X$  mit  $\mathbb{E}(|X|) < \infty$  und  $\mathbb{E}(X) = \theta$ . Dann heißt  $T_n : \Omega^n \rightarrow \mathbb{R}, x \mapsto \frac{1}{n} \sum_{i=1}^n x_i$  heißt **Mittelwertschätzer**.

**Satz 6.11.** Der Mittelwertschätzer ist schwach konsistent

*Beweis.* Durch GGZ. □

**Definition 6.12.** Sei  $T$  Schätzer für eine Größe  $f(\theta)$  dann

a) ist der sog **Bias** definiert als

$$\text{Bias}_\theta := \mathbb{E}_\theta(T) - f(\theta)$$

b) heißt  $T$  **Erwartungstreu**, falls  $\text{Bias}_\theta = 0$ .

c) der sog **mean square error** definiert als

$$\text{MSE}_\theta(T) := \mathbb{E}_{\mathbb{P}_\theta^{\otimes n}} \left( (T - \mathbb{E}_{\mathbb{P}_\theta^{\otimes n}}(T))^2 \right) = \mathbb{V}_{\mathbb{P}_\theta^{\otimes n}}(T)$$

d) heißt  $T$  **konsistent im quadratischen Mittel** (im Produktmodell), falls  $\text{MSE}_\theta \xrightarrow{n \rightarrow \infty} 0$ .

**Lemma 6.13.** Sei  $T$  erwartungstreu und  $\mathbb{P}_{(X,Y)}\theta^{\otimes n}(T)$  konsistent im quadratischen Mittel, dann ist  $T$  schwach konsistent.

**Definition 6.14.** Sei  $(\Omega^n, \mathcal{F}^{\otimes n}, (\mathbb{P}_\theta)_{\theta \in \Theta}^{\otimes n})$  für alle  $n$  ein Produktmodell, sei  $\mathbb{P}_{(X,Y)}\theta$  die Verteilung einer reellen Zufallsvariable  $X$  mit  $\mathbb{E}(|X|) < \infty$ ,  $\mathbb{V}_\theta(X) = \theta$ . Dann ist der **Varianzschätzer** gegeben durch

$$x \mapsto \frac{1}{n-1} \sum_{i=1}^n (x_i - \frac{1}{n} \sum_{i=1}^n x_i)^2 \quad (\text{„Empirische Varianz“})$$

**Korollar 6.14.**  $T$  ist erwartungstreu

**Korollar 6.14.** Wenn  $(X_i)$  Gaußverteilt ist, dann ist  $T$  konsistent im quadratischen Mittel.

*Bemerkung.* In der Theorie der Statistik untersucht man „optimale“ Schätzer (z.B. mit minimalem MSE). Sie zu finden ist nicht immer einfach.

*Beispiel 6.15.* Sei  $\Omega = [0, \infty)$ ,  $\mathbb{P}_{(X,Y)}\theta = \frac{1}{\theta} \chi_{[0,\theta]} \cdot \lambda$ . (Gleichverteilung auf  $[0, \theta]$ ). Modell:  $(\Omega^n, \mathcal{F}^{\otimes n}, \mathbb{P}_\theta^{\otimes n})$ . Schätze  $\theta$  aus Daten  $x_1, \dots, x_n$ . Mögliche Sätze

- a)  $T_1(x) = 2 \frac{1}{n} \sum_{i=1}^n x_i$
- b)  $T_2(x) = \frac{n+1}{n} \max\{x_1, \dots, x_n\}$
- c)  $T_3(x) = \max\{x_1, \dots, x_n\} + \min\{x_1, \dots, x_n\}$

## 6.B Maximum-Likelihood-Schätzer

**Definition 6.16.** a) Auf diskreten Wahrscheinlichkeitsräumen: Sei  $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$  statistisches Modell,  $\Omega$  abzählbar,  $\rho_\theta(x) := \mathbb{P}(\{x\})$  sei Zähldichte von  $\mathbb{P}_\theta$ .

**Simplified Idee** (i) Zu den Daten  $x$  Berechne  $\mathbb{P}_\theta(\{x\})$   
(ii) Rak-Vorschrift: das „richtige“  $\theta$  ist jenes für das  $\mathbb{P}_\theta(\{x\})$  maximal ist.

**Formal** Zu  $x \in \Omega$  heißt die Abbildung  $\theta \mapsto \rho_\theta(x)$  Likelihood- (oder Plausibilitäts-)Funktion zum Beobachtungswert  $x$ .

Eine Abbildung  $\Omega \rightarrow \Theta, x \mapsto \arg \max \rho_x$  heißt **maximum-Likelihood Schätzer**.

- b) Stetige Wahrscheinlichkeitsräume mit Dichte: Sei  $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$  statistisches Modell,  $\Omega = \mathbb{R}^d$ ,  $\rho_\theta$  Dichtefunktion zu  $\mathbb{P}_\theta$ . Dann heißt die Abbildung  $\Omega \rightarrow \Theta, x \mapsto \arg \max \rho_\theta$  **ML-Schätzer**.

*Bemerkung 6.17.* Berechnen von ML Schätzern ist meist durch Ableiten von  $\theta \mapsto \rho_\theta(x)$  und Nullsetzen möglich. Da Schätzer häufig Produktstruktur haben ist es oft einfacher  $\theta \mapsto \ln \rho_\theta(x)$ .

**Definition 6.18.** Sei  $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$  statistisches Modell,  $f : \Theta \rightarrow \Omega'$  messbar,  $\alpha > 0$ .

Eine Abbildung  $K : \Omega \rightarrow \mathcal{P}(\Omega')$  heißt **Konfidenzbereich für  $f$  zum Irrtumsniveau  $\alpha$  (Sicherheitsniveau  $1 - \alpha$ )**, wenn gilt

$$(i) \quad \{\omega \in \Omega : f(\theta) \in K(\omega)\} \in \mathcal{F}$$

$$(ii) \quad \mathbb{P}_\theta(f(\theta) \in K(\cdot)) \geq 1 - \alpha$$

**Bemerkung 6.19.** es wird zu jedem Datensatz  $x$  eine Menge  $K(x) \subset \Omega'$  bestimmt, sodass  $f(\theta)$  nur mit sehr geringer Wahrscheinlichkeit (beim Erzeugern der Daten) nicht in  $K(x)$  liegt, falls  $\theta$  der wahre Wert war.

Genauer: Macht man das Experiment

- a) Würfele  $x$  mit  $\mathbb{P}_\theta$  aus
- b) rate, dass  $\theta \in K(x)$
- c) Schaue nach ob richtig geraten wurde

Häufiges unabhängiges durchführen: Man hat in mehr als  $(1 - \alpha)N$  Fällen recht.

**Definition 6.20.** Sei (vgl ??)  $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$  statistisches Modell,  $f : \Theta \rightarrow \Omega'$  messbar,  $\Omega' = \mathbb{R}$  und  $K : \Omega \rightarrow \mathcal{P}(\Omega')$ . Dann heißt die Abbildung  $x \mapsto K(x)$  **Konfidenzintervall**.

**Satz 6.21** (Berechnung von Konfidenzintervallen). *Sei  $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$  statistisches Modell,  $\Omega' = \mathbb{R}$ ,  $f : \Theta \mapsto \mathbb{R}$  Zielgröße,  $x \in \Omega$  Daten.*

*Gesucht:  $KI : [R^-, R^+]$ .*

**Einzigste Möglichkeit** Sage  $R_-, R_+$  aus den Daten vorher. *Gesucht sind Schätzer  $x \mapsto R_-(x), x \mapsto R_+(x)$  für die Intervallenden.*

**Benötigte Eigenschaft**

$$\forall \theta : \mathbb{P}_\theta(f(\theta) \in [R_-, R_+]) \geq 1 - \alpha \Leftrightarrow \mathbb{P}_\theta(R_- > f(\theta)) \text{ oder } \mathbb{P}_\theta(R_+ < f(\theta)) < \alpha$$

**Hinreichend** (und in der Praxis auch immer Benutzt)

$$\mathbb{P}_\theta(R_- > f(\theta)) \leq \alpha/2 \quad \text{und} \quad \mathbb{P}_\theta(R_+ < f(\theta)) \leq \alpha/2$$

Finde also  $\forall \theta \in \Theta$  eine Funktion  $\omega \mapsto R_{-, \theta}(\omega)$ , sodass

$$\mathbb{P}_\theta(R_{-, \theta} > f(\theta)) \leq \alpha/2 \Leftrightarrow \mathbb{P}_\theta \circ R_{-, \theta}^{-1}([f(\theta), \infty)) \leq \alpha/2$$

Da man  $\theta$  nicht kennt muss man annehmen  $R_-(x) = \inf\{R_{-, \theta(x)} | \theta \in \Theta\}$  und  $R_+(x) = \sup\{R_{+, \theta(x)} | \theta \in \Theta\}$ .

**Beispiel 6.22.** Sei  $\mathbb{P}_\theta = \mathcal{N}_{\theta, \sigma^2}$ ,  $\sigma^2$  bekannt und fest.

Finde  $f(\theta) = \theta$ , d.h. schätze den Erwartungswert.

$(\Omega^n, \mathcal{F}^{\otimes n}, (\mathbb{P}_\theta^{\otimes n})_{\theta \in \Theta})$  Wähle dann

$$R_-(x) = \frac{1}{n} \sum_{i=1}^n x_i - s \quad R_+ = \frac{1}{n} \sum_{i=1}^n x_i + s$$

Dann ist, da  $X_i \sim \theta, \sigma^2$ ,  $Y_i \sim \mathcal{N}_{0,\sigma}$ ,  $Z \sim \mathcal{N}_{0,\sigma^2}$ ,  $W \sim \mathcal{N}_{0,1}$ .

$$\begin{aligned}\mathbb{P}^{\otimes n}(R_- \geq \theta) &= \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n (X_i - \theta) \geq s\right) \\ &= \mathbb{P}\left(\frac{1}{\sqrt{n}} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i\right) \geq s\right) \\ &= \mathbb{P}\left(\frac{1}{\sqrt{n}} Z \geq s\right) \\ &= \mathbb{P}(Z \geq s\sqrt{n}) = \mathbb{P}\left(W \geq \frac{s}{\sigma} \sqrt{n}\right)\end{aligned}$$

Also muss  $s$  (mindestens) als Lösung der Gleichung

$$\int_{\frac{s}{\sigma}\sqrt{n}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \alpha/2$$

wählen.

**Definition 6.23.** Sei  $\mathbb{P}$  Wahrscheinlichkeitsmaß auf  $\mathbb{R}$ ,  $0 < \alpha < 1$ .

Eine Zahl  $q_\alpha \in \mathbb{R}$  heißt  **$\alpha$ -quantil** von  $\mathbb{P}$ , falls  $\mathbb{P}((-\infty, q_\alpha)) \leq \alpha$  und  $\mathbb{P}((-\infty, q_\alpha]) \geq \alpha$ .

**Lemma 6.24.** In 6.23 sei  $F$  die Verteilungsfunktion von  $\mathbb{P}$ . Dann gilt

- a)  $\lim_{x \uparrow q_\alpha} \leq \alpha$  und  $F(q_\alpha) \geq \alpha$  für jedes  $\alpha$ -Quantil  $q_\alpha$ .
- b) Jede Zahl im Intervall  $(\sup\{y \in \mathbb{R} : F(y) \leq \alpha\}, \inf\{y \in \mathbb{R} : F(y) \geq \alpha\})$  ist ein  $\alpha$ -Quantil
- c) Falls  $F$  stetig und streng monoton ist, dann ist  $q_\alpha$  die eindeutige Zahl mit  $F(q_\alpha) = \alpha$ .

**Proposition 6.25.** Sei  $(\Omega^n, \mathcal{F}^{\otimes n}, (\mathbb{P}_\theta^{\otimes n})_{\theta \in \mathbb{R}})$  statistisches Modell mit  $\mathbb{P}_\theta = \mathcal{N}(\theta, \sigma^2)$ , wobei  $\sigma^2$  bekannt und fest. Sei  $q_{1-\frac{\alpha}{2}}$  das  $1 - \frac{\alpha}{2}$ -Quantil von  $\mathcal{N}(0, 1)$ . Dann ist die Abbildung

$$K : \Omega^n \rightarrow \mathcal{P}(\mathbb{R}) : x \mapsto \left[ M(x) - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}, M(x) + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \right] = \left[ M(x) + \frac{\sigma}{\sqrt{n}} q_{\alpha/2}, M(x) + \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2} \right]$$

(Dabei ist  $M(x) := \frac{1}{n} \sum_{i=1}^n x_i$  der Mittelwert) ein Konfidenzintervall zum Irrtumsniveau  $\alpha$ .

*Beweis.* Wähle  $s = \frac{\sigma}{\sqrt{n}} q_{1-\alpha}$  in 6.22 □

*Beispiel 6.26.* Sei  $\Theta = \mathbb{R} \times \mathbb{R}^+$ ,  $\theta = (\mu, \sigma^2)$ ,  $\mathbb{P}_\theta = \mathcal{N}(\mu, \sigma^2)$  und  $(\Omega^n, \mathcal{F}^{\otimes n}, (\mathbb{P}_\theta^{\otimes n})_{\theta \in \Theta})$  statistisches Modell.

Gesucht wird nun das Konfidenzintervall für die Varianz  $\sigma^2$ .

Der einzige vernünftige Schätzer ist dann

$$S^2 : \Omega^n \rightarrow \mathbb{R}^+, S^2(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - M(x))^2$$

Für  $X_i \sim \mathcal{N}(\mu, \sigma^2)$  ist dann

$$\begin{aligned} S^2((X_1, \dots, X_n)) &= \frac{1}{n} \sum_{i=1}^n \left[ (X_i - \mu) - \frac{1}{n} \sum_{j=1}^n (X_j - \mu) \right]^2 \\ &= \frac{\sigma^2}{n-1} \sum_{i=1}^n \left[ \frac{X_i - \mu}{\sigma} - \frac{1}{n} \sum_{j=1}^n \frac{X_j - \mu}{\sigma} \right]^2 \end{aligned}$$

Dabei ist  $\frac{X_i - \mu}{\sigma} \sim \mathcal{N}(0, 1)$ , also mit  $Y_1, \dots, Y_n \sim \mathcal{N}(0, 1)$  uiv

$$\sim \sigma^2 S^2((Y_1, \dots, Y_n))$$

Da Bildmaß von  $\frac{S^2(x)}{\sigma^2}$  ist unabhängig von  $\theta$ ! Also folgt

$$\begin{aligned} \mathbb{P}_\theta(S^2 > a\sigma^2) &= \mathbb{P}_\theta \circ \left( \frac{S^2}{\sigma^2} \right)^{-1} ((a, \infty]) =: \nu(a, \infty) \\ \mathbb{P}_\theta(S^2 < b\sigma^2) &= \mathbb{P}_\theta \circ \left( \frac{S^2}{\sigma^2} \right)^{-1} ((-\infty, b)) =: \nu(-\infty, b) \end{aligned}$$

Falls man  $a$  als  $(1 - \frac{\alpha}{2})$ -Quantil von  $\nu$  und  $b$  als  $\alpha/2$ -Quantil von  $\mu$  wählt, dann gilt

$$\mathbb{P}_\theta(\sigma^2 \in [\frac{1}{a}S^2, \frac{1}{b}S^2]) \geq 1 - \alpha$$

**Definition 6.27.** Inhalt...

**Satz 6.28.** Seien  $X_1, \dots, X_n$  uiv und  $X_1 \sim \mathcal{N}(\mu, \sigma)$ . Dann ist  $M(x) = \frac{1}{n} \sum_{i=1}^n X_i(\omega)$ ,  $S^2(\omega) = \frac{1}{1-n} \sum_{i=1}^n (X_i(\omega) - M(\omega))^2$ .

*Bemerkung.*  $t_{n-1}$  ist ähnlich wie  $\mathcal{N}(0, 1)$  und für große  $n$  gilt  $\mathbb{P}_{t_n} \rightarrow \mathcal{N}(0, 1)$

**Proposition 6.29.** Sei  $(\Omega^n, \mathcal{F}^{\otimes n}, (\mathbb{P}_\theta^{\otimes n}))$  statistisches Modell mit  $\mathbb{P}_\theta = \mathcal{N}(\mu, \sigma^2)$ ,  $f(\theta) = f((\mu, \sigma^2)) = \sigma^2$ .

Seien  $q_\alpha$  das  $\alpha$ -Quantil der  $\chi_{n-1}^2$ -Verteilung. Dann ist die Abbildung

$$K : \Omega^n \rightarrow \mathcal{B}([0, \infty]), \quad x \mapsto \left[ (n-1) \frac{S^2(x)}{q_{1-\alpha/2}}, (n-1) \frac{S^2(x)}{q_{\alpha/2}} \right]$$

ein Konfidenzintervall von  $f(\theta)$  zum Irrtumsniveau  $\alpha$ .

*Beweis.* Es gilt

$$\begin{aligned} \mathbb{P}_\theta \left( f(\theta) < \frac{(n-1)S^2}{q_{1-\alpha/2}} \right) &= \mathbb{P}((n-1)S^2 > \sigma^2 q_{1-\alpha/2}) \\ &= \mathbb{P}_\theta \left( \frac{n-1}{\sigma^2} S^2 > q_{1-\alpha/2} \right) \\ &= \mathbb{P}_{\chi_{n-1}^2}(Y > q_{1-\alpha/2}) = \alpha/2 \end{aligned}$$

$$\text{und } \mathbb{P}_\theta(f(\theta) > (n-1)^2 \frac{S^2}{q_{\alpha/2}}) = \dots = \mathbb{P}_{\chi_{n-1}^2}(Y < q_{\alpha/2}) = \frac{\alpha}{2}.$$

□



*Bemerkung 6.30.* 6.29 erscheint so also ob das Konfidenzintervall mit wachsendem  $n$  größer wird. Kontraintuitiv!  
Es gilt

$$\mathbb{P}_{\chi_n^2}(\text{id} \leq c) = \mathbb{P}\left(\sum_{j=1}^n X_j \leq c\right)$$

mit  $X - i \sim \Gamma_{\frac{1}{2}, \frac{1}{2}}$  (siehe ??c).

Dann gilt  $\mathbb{E}(X_i) = 1$ ,  $\mathbb{E}(X_i^2) = 3$ , also  $\mathbb{V}(X_i) = 2$ . Somit ist

$$\begin{aligned}\mathbb{P}_{\chi_{n-1}^2}((-\infty, c]) &= \mathbb{P}\left(\sum_{j=1}^n (X_j - 1) \leq c - (n - 1)\right) \\ &= \mathbb{P}\left(\frac{1}{\sqrt{n-1}} \sum_{j=1}^n (X_j) \leq \frac{c}{\sqrt{n-1}} - \sqrt{n-1}\right)\end{aligned}$$

ist nach dem Gesetz der Großen Zahlen  $\sim \mathcal{N}(0, 2)$  für große  $n$ .

*Beispiel 6.31.* Sei  $\Theta = \mathbb{R} \times \mathbb{R}^+$ ,  $\theta = (\mu, \sigma^2)$ ,  $\mathbb{P}_\theta = \mathcal{N}(\mu, \sigma^2)$ . Sei  $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$  statistisches Modell. Nun sind  $\mu$  und  $\sigma^2$  unbekannt, gesucht ist ein Konfidenzintervall für  $\mu$ :

Naiver Versuch: Berechne das Konfidenzintervall gemäß 6.22 für alle  $\sigma^2$  und maximiere über  $\sigma^2$ . Dann ergibt sich

$$K_{\sigma^2}(x) = \left[ M(x) + \frac{\sigma}{\sqrt{n}} q_{\alpha/2}, M(x) + \frac{\sigma}{\sqrt{n}} q_{(1-\alpha)/2} \right]$$

Dann ergibt das Maximieren über  $\sigma \in \mathbb{R}$  ergibt  $K(x) = \mathbb{R}$ .

*Beweis.* Für alle  $\theta \in \Theta$  gilt

$$\mathbb{P}_\theta \left( M - \frac{\sqrt{S^2}}{\sqrt{n}} q_{1-\alpha/2} \geq \mu \right) = \mathbb{P}_\theta \left( M - \mu \geq \frac{\sqrt{S^2}}{\sqrt{n}} q_{1-\alpha/2} \right) = \mathbb{P}_\theta \left( \sqrt{n} \frac{M - \mu}{\sqrt{S^2}} \geq q_{1-\alpha/2} \right)$$

□

## 6.C Tests

*Bemerkung 6.32.* Wichtigstes Werkzeug der Statistik: wissenschaftliche Studien, Medikamentenzulassung, Nachweis von Produktmängeln, ect. Dient zum Treffen von Ja-Nein-Entscheidungen anhand „verrauschter“ Daten.

Sei  $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$  statistisches Modell. Wir teilen  $\Theta$  in zwei Bereiche:  $A_0, A_1 \subset \Theta$ , mit  $A_0 \cup A_1 = \Theta$ . Dann beschreibt die

- **Nullhypothese:**  $\theta \in A_0$  (Schreibe  $H_0$  gilt)
- **Alternative:**  $\theta \in A_1$  (Schreibe  $H_1$  gilt)

Testproblem: Kann man aufgrund der Daten  $x \in \Omega$  bestimmen, ob  $\sigma \in A_0$  oder  $\in A_1$  war?

Dabei soll  $\theta \in A_1$  nur sehr unwahrscheinlich (bzw. bezüglich der Daten) fälschlicherweise angenommen werden.

**Definition 6.33.** Eine Abbildung  $T : \Omega \rightarrow \{0, 1\}$  heißt **Test**. Dabei bedeute

$T(x) = 0$  als „ $\theta$  liegt in  $A_0$ “.

Wir sagen **Nullhypothese wird nicht abgelehnt**

$T(x) = 1$  als „ $\theta$  liegt in  $A_1$ “.

Wir sagen **Nullhypothese wird abgelehnt**

Falls  $\theta \in A_0$  aber  $T(x) = 1$  so sagen wir die Daten führen zu einem **Fehler 1. Art**.

Falls  $\theta \in A_1$  aber  $T(x) = 0$  so sagen wir die Daten führen zu einem **Fehler 2. Art**.

*Beispiel 6.34.1* (Feuerwerkskörper). Behauptung des Herstellers: höchstens 3% Ausschuss,

$H_1: \theta > 0.03$ ,  $H_0: \theta \leq 0.03$ .

Falls die Daten suggerieren, dass  $H_1$  gilt, dann gibt es Ärger für alle Beteiligten (Reklamation, Untersuchung, Prozess). Falls  $H_1$  dagegen nicht erkannt wird, sind die Verluste gering.

*Beispiel 6.34.2* (Wissenschaftlicher Durchbruch).  $H_0$  entspricht der „Lehrmeinung“,  $H_1$  entspricht „Skandal, Sensation“ (z.B. „Kalte Kernfusion“, „Stammzellentherapie“, ...).

Ein falsches Behaupten von  $H_1$  kostet Reputation, Karriere.

Nicht erkennen von  $H_1$  ist eine verpasste Chance aber nicht fatal.

**Definition 6.35.** Ein Test  $T$  hält das **Irrtumsniveau**  $\alpha > 0$  ein, falls gilt:

$$\forall \theta \in A_0 : \mathbb{P}_\theta(T = 1) < \alpha$$

Zu  $\theta \in A_1$  heißt die Zahl  $G_T(\theta) := \mathbb{P}_\theta(T = 1)$  die **Güte** (oder Schärfe, Mächte) des Test bei  $\theta$ .

*Bemerkung 6.36* (Konstruktion von Tests). mit Konfidenzintervallen

- a) Zweiseitiger Test: angenommen man hat ein Konfidenzintervall  $K : \Omega \rightarrow \Omega'$  für die Größe  $f(\theta) \in \Omega'$  zum Niveau  $\alpha$  gegeben. Dann setzte man

$$T : \Omega \rightarrow \{0, 1\}, T(x) = \begin{cases} 1 & \text{falls } K(x) \cap \{f(\theta) : \theta \in A_0\} = \emptyset \\ 0 & \text{sonst} \end{cases}$$

Dann gilt für alle  $\theta_0 \in A_0$

$$\mathbb{P}_\theta(T = 1) \mathbb{P}_{\theta_0}(\forall \theta \in A_0 : f(\theta) \notin K) \leq \mathbb{P}_{\theta_0}(f(\theta_0) \notin K) = 1 - \mathbb{P}_{\theta_0}(f(\theta_0) \in K) \leq \alpha$$

Interpretation, falls  $T(x) = 1$ : Die Daten  $x$  können nur mit sehr geringer Wahrscheinlichkeit von  $\theta_0 \in A_0$  erzeugt werden, sodass wir  $\theta_0 \in A_0$  ausschließen.

- b) einseitige Tests: mit Asymmetrischen Konfidenzintervallen. Oft ist  $\Omega' = \mathbb{R}$ ,  $A_0 = (-\infty, c]$  oder  $A_0 = [c, \infty)$ . Dann ist es besser statt wie in a) symmetrischer auch asymmetrische Konfidenzintervalle der Form  $K(x) = [S(x), \infty)$  oder  $K(x) = (-\infty, S(x)]$  zu wählen.

*Beispiel 6.37.1* (Gaußtest, Zweiseitig). Inhalt...

## 7 Markovketten

*Bemerkung* (Motivation). Sei  $E$  eine Menge („Spielfeld“) mit abzählbar vielen Feldern („Zuständen“) = Elemente von  $E$ .

Auf jedem Feld  $x \in E$  liegt ein Würfel bereit, der  $N_x \leq |E|$  Seiten hat und mit Wahrscheinlichkeit  $p(x, y)$  die Seite  $y$  würfelt.

Man geht dann auf das Feld  $y$ .

*Beispiel 7.1.1* (Irrfahrt).  $E = \mathbb{Z}$ ,  $p(x, x+1) = p(x, x-1) = \frac{1}{2}$ .

*Beispiel 7.1.2* (Galton-Watson-Populationsmodell).  $E = \mathbb{N}_0$ ,  $p(x, y) = \mathbb{P}\left(\sum_{j=1}^x Z_j = y\right)$ , wobei  $Z_j$  eine Zufallsvariable ist, die die Nachkommen des  $j$ -ten Individuum modelliert. Normalerweise sind die  $Z_j$  i.i.d.

*Beispiel 7.1.3* (Sammelbilder). Sei  $E = \{1, \dots, N\}$  („Anzahl der noch fehlenden Bilder“) und

$$p(x, y) = \begin{cases} x/N & \text{falls } y = x-1 \\ 1 - \frac{x}{N} & \text{falls } y = x \\ 0 & \text{sonst} \end{cases}$$

*Bemerkung.* Wie hoch ist die Wahrscheinlichkeit in 2 Schritten von  $x$  nach  $z$  zu gehen?

$$\begin{aligned} \mathbb{P}(x \xrightarrow{2 \text{ Schritte}} z) &= \mathbb{P}(x \rightarrow y, y \rightarrow z) \\ &= \sum_{y \in E} \mathbb{P}(x \rightarrow y) \mathbb{P}(y \rightarrow z) \\ &= \sum_{y \in E} p(x, y) p(y, z) \end{aligned}$$

Vergleiche: Matrixmultiplikation

$$= P^2(x, z)$$

mit  $P^2 = P \cdot P$ ,  $P = (p(x, y))_{x, y \in E}$  (Matrix)

Beachte dabei:

$$\sum_{y \in E} p(x, y) = \sum_{y \in E} \mathbb{P}(x \rightarrow y) = \mathbb{P}(x \rightarrow E) = 1$$

(Zeilensumme = 1)

**Definition 7.2.** Eine Matrix  $P = (p(x, y))_{x, y \in E}$  heißt **stochastisch**, wenn gilt

- (i)  $p(x, y) \geq 0 \forall x, y \in E$
- (ii)  $\sum_{y \in E} p(x, y) = 1 \forall x \in E$

**Definition 7.3.** Sei  $E$  höchstens abzählbar,  $P$  stochastische Matrix.

Eine Folge  $(X_n)$  von  $E$ -wertigen Zufallsvariablen heißt **Markovkette** (Mk) mit Übergangsmatrix  $P$  (ÜM), falls für alle  $n \in \mathbb{N}$  und für alle  $x_1, \dots, x_n \in E$  gilt

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) = p(x_n, x_{n+1})$$

*Bemerkung.* Falls  $\mathbb{P}(X_0 = x) = \mu(x)$  für ein Wahrscheinlichkeitsmaß  $\mu$ , dann schreibt man auch  $\mathbb{P}^\mu$  statt  $\mathbb{P}$ .

Falls

$$\mathbb{P}(X_0 = x) = \delta_y(x) = \begin{cases} 1 & \text{falls } y = x \\ 0 & \text{sonst} \end{cases}$$

schreibt man  $\mathbb{P}^y$  statt  $\mathbb{P}$ .

**Proposition 7.4.** Sei  $(X_n)$  Markovkette mit Übergangsmatrix  $P$ . Dann gilt für alle  $x \in \mathbb{N}$ ,  $x, y \in E$

$$\mathbb{P}^x(X_n = y) = P^n(x, y)$$

wobei  $P^n$  die  $n$ -te Potenz von  $P$  ist.

*Beweis.* Induktiv □

*Bemerkung.* Es besteht eine starke Verbindung zur linearen Algebra. Falls  $|E| = \infty$  wird  $P$  trotzdem Matrix genannt.

**Proposition 7.5.** Sei  $P$  stochastische Matrix. Dann gilt

- a)  $P1 = 1$ , wobei  $1(x) = 1 \forall x$ .  
Das heißt,  $1$  ist Rechts-Eigenvektor von  $P_y$  zum Eigenwert  $1$ .
- b) Falls  $|E| < \infty$ , dann Existiert ein Zeilenvektor  $\mu \neq 0$  mit  $\mu P = \mu$ . D.h.  $\mu$  ist Links-Eigenvektor zu  $P$ , zum Eigenwert  $1$ .

*Beweis.* a)  $P1(x) = \sum_{y \in E} p(x, y)1(y) = 1$  (Zeilensumme)

- b)  $P1 = 1$  genau dann wenn  $\langle w, P1 \rangle = \langle w, 1 \rangle$  für alle  $w \in \mathbb{R}^E$ .  
Genau dann wenn  $\langle P^*w, w, 1 \rangle = 0$  für alle  $w \in \mathbb{R}^E$ .  
Genau dann wenn  $\text{Im}(P^*w - 1) \perp \langle \{1\} \rangle$ .  
Dann ist also  $\dim(\text{Kern}(P^* - 1)) > 0$ .  
Aus dem Eigenwert  $1$  folgt, dass  $P$  Zeileneigenwerte zum Eigenwert  $1$ . □

**Definition 7.6.** Sei  $(X_i)$  Markovkette auf  $E$  mit Übergangsmatrix  $P$ . Ein Wahrscheinlichkeitsmaß  $\hat{\mu}$  auf  $E$  heißt **invariante Verteilung** von  $(X_n)$ , falls  $\mathbb{P}^{\hat{\mu}}(X_1 = y) = \hat{\mu}(\{y\})$ . Dabei ist

$$\begin{aligned} \mathbb{P}^{\hat{\mu}}(X_1 = y) &= \sum_{x \in E} \hat{\mu}(\{x\}) \mathbb{P}^x(X_1 = y) \\ &= \sum_{x \in E} \hat{\mu}(\{x\}) p(xy) = (\mu P)(y) \end{aligned}$$

(und  $\mu$  die Zähldichte von  $\hat{\mu}$ ).

*Bemerkung 7.7.* Invarianz von Verteilungen bedeutet, dass die Zähldichte  $\mu$  von  $\hat{\mu}$  ein Linkseigenvektor von  $P$  ist.

Nach 7.5 existiert für  $|E| < \infty$  immer irgendein Linkseigenvektor.

Ist dies auch die Zähldichte einer Verteilung?

Dazu muss der Linkseigenvektor zusätzlich nur nichtnegative Einträge haben und Zeilensumme  $1$  haben.

Der Satz von Perron-Frobenius („Sei  $A$  eine Matrix mit nichtnegativen Einträgen,  $\|A\| = \sup\{|Av| : |v| = 1\}$ , dann ist  $w$  ein Eigenvektor von  $A$  zum Eigenwert  $\|A\|$ . Dann kann  $w$  nicht-negativ gewählt werden.“)

*Bemerkung.* Markovketten sind „gedächtnislose“ Prozesse.

Ist man zur Zeit  $x \in X$ , dann spielt es für die weitere Entwicklung keine Rolle, wie man dort hin gekommen ist.

**Satz 7.8** (Markov-Eigenschaft). *Sei  $X_n$  Markovkette mit Übergangsmatrix  $P$ . Dann gilt für alle  $x, z \in E$ ,  $n \in \mathbb{N}$ ,  $A \in \mathcal{P}(E)^{\otimes \mathbb{N}_0}$  und  $B \in E^n$ , falls  $\mathbb{P}((X_1, \dots, X_n) \in B, X_n = x) > 0$ , dann ist*

$$\begin{aligned} \mathbb{P}^z((X_n, X_{n+1}, X_{n+2}, \dots) \in A | (X_0, \dots, X_{n-1}) \in B, X_n = x) &= \mathbb{P}^z((X_n, X_{n+1}, \dots) \in A | X_n = x) \\ &= \mathbb{P}((X_1, X_2, \dots) \in A) \end{aligned}$$

*Beispiel 7.9* (Ruin des Spielers). Sei  $X_n$  = Gewinn nach dem  $n$ -ten Spiel,  $E = \{-a, \dots, 0, \dots, b\}$ ,  $p(n, n+1) = p$ ,  $p(n, n-1) = 1-p$ , falls  $-a < n < b$ .

- $p(-a, -a) = 1$  („Ruin“)
- $p(b, b) = 1$  („Gewinnmitnahme“)

**Definition 7.10.**  $z \in E$  heißt **absorbierend** bezüglich  $\mathbb{P}$ , falls  $p(z, z) = 1$ . Dann heißt

$$h_z(x) := \mathbb{P}^x(\exists N \in \mathbb{N} : \forall n \geq N : X_n = z) = \mathbb{P}\left(\bigcup_{N \geq 0} \bigcap_{n \geq N} \{X_n = z\}\right)$$

die **Absorptionswahrscheinlichkeit in  $z$  bei Startwert  $x$** .

**Definition 7.11.** Sei  $(X_n)$  eine Markovkette,  $z \in E$ . Die Zufallsvariable

$$\tau_z : \Omega \rightarrow \mathbb{N} \cup \{\infty\}, \omega \mapsto \inf\{n \geq 1 : X_n(\omega) = z\}$$

heißt **Trefferzeit** (oder **Eintrittszeit**) bei  $z$ .

**Lemma 7.12.** *Sei  $(X_n)$  Markovkette,  $\tau_z$  Trefferzeit. Dann ist*

$$\{\omega \in \Omega : \tau_z = n\} \in \sigma(X_1, \dots, X_n) \quad (\text{Gl. 7.4})$$

*Beweis.* Es gilt

$$\begin{aligned} \{\omega \in \Omega : \tau_z(\omega) = n\} &= \{\omega \in \Omega : X_1(\omega) \neq z, \dots, X_{n-1}(\omega) \neq z, X_n(\omega) = z\} \\ &= \bigcap_{k=1}^{n-1} X_k^{-1}(E \setminus \{z\}) \cap X_n^{-1}(\{z\}) \in \sigma(X_1, \dots, X_n) \end{aligned}$$

□

*Bemerkung.* Die Messbarkeitseigenschaft Gl. 7.4 macht  $\tau_z$  zu einer sogenannten **Stoppzeit**. Um zu entscheiden, dass „gestoppt“ wird muss man nicht „in die Zukunft Blicken“.

**Satz 7.13.** *Sei  $(X_n)$  Markovkette,  $z \in E$  absorbierend,  $x \in E$  beliebig. Dann ist*

$$h_z(x) = \mathbb{P}^x(\tau_z < \infty) = \mathbb{P}^x\left(\lim_{n \rightarrow \infty} X_n = z\right) \quad (\text{Gl. 7.5})$$

und

$$h_z = \min \left\{ h : E \rightarrow \mathbb{R}^+ \text{ mit } h(z) = 1 \text{ und } \sum_{y \in E} p(x, y)h(y) = h(x) \right\} \quad (\text{Gl. 7.6})$$

(wobei  $h \leq g \Leftrightarrow h(x) \leq g(x) \forall x \in E$ )

Da Minimum ist eindeutig.

*Bemerkung.*  $h_z$  ist also Rechts-Eigenvektor von  $P$ , mit  $h_z(z) = 1$ . Auch  $1$  mit  $1(x) = 1$  ist Rechts-Eigenvektor, aber  $h_z$  kann kleiner sein.

*Beweis.* Zu Gl. 7.5:

$$\mathbb{P}^z(X_n = z \forall n \geq 1) = \lim_{k \rightarrow \infty} \mathbb{P}^z(X_n = z \forall n \leq k)$$

Da  $\{X_n = z \forall n \leq k\}_k$  eine Folge absteigender Mengen ist

$$= \lim_{k \rightarrow \infty} P^k(z, z) = \lim_{k \rightarrow \infty} p(z, z)^k = 1$$

Also gilt für alle  $x \in E$

$$\begin{aligned} \mathbb{P}^x(X_n = z) &= \mathbb{P}^x(X_n = z) \mathbb{P}^z(X_i = z \forall i \geq 1) \\ &\stackrel{7.8}{=} \mathbb{P}^x(X_n = z, X_{n+1} = z \forall i \geq 1) \\ &\xrightarrow{n \rightarrow \infty} \mathbb{P} \left( \bigcup_{n=1}^{\infty} \bigcap_{j=n}^{\infty} \{X_j = z\} \right) = h_z(x) \end{aligned}$$

Weiter gilt für alle  $k \leq n$

$$\begin{aligned} \mathbb{P}^x(X_n = z | \tau_z = k) &= \mathbb{P}^x(X_n = z | \tau_z = k, X_k = z) \\ &\stackrel{7.8}{=} \mathbb{P}^x(X_n = z | X_k = z) = \mathbb{P}^z(X_{n-k} = z) = 1 \end{aligned}$$

Dann folgt mit 7.12, dass

$$\begin{aligned} \underbrace{\mathbb{P}^x(X_n = z)}_{\xrightarrow{n \rightarrow \infty} h_z(x)} &= \sum_{k=1}^n \mathbb{P}^x(\tau_z = k) \overbrace{\mathbb{P}^x(X - n = z | \tau_z = k)}^{=1} \\ &= \mathbb{P}(\tau_n \leq n) \xrightarrow{n \rightarrow \infty} \mathbb{P}(\tau_z < \infty) \end{aligned}$$

Zu Gl. 7.6:

- a)  $h_z \in \{h : E \rightarrow \mathbb{R}^+, h(z) = 1, Ph = h\}$ , denn  $h_z(x) \geq 0$  als Wahrscheinlichkeit,  $h(z) = 1$  und

$$\sum_{y \in E} p(x, y)h_z(y) = \sum_{y \in E} \mathbb{P}^x(X_1 = y) \mathbb{P}^y(\{\exists N \in \mathbb{N} : \forall n \geq N : X_n = z\})$$

$$= \sum_{y \in E} \mathbb{P}^x(X_n = y) \mathbb{P}(\exists N \in \mathbb{N}_0 : \forall n \geq N : X_n = z)$$

$$\stackrel{3.4a)}{=} \mathbb{P}^x(\exists N \in \mathbb{N} : \forall n \geq N : X_{n+1} = z)$$

b)  $h_z$  ist maximale. Sei  $\tilde{h} \geq 0$ ,  $\tilde{h}(x) = 1$  und  $P\tilde{h} = \tilde{h}$ , dann ist

$$\tilde{h}(x) = P^n \tilde{h}(x) = \sum_{x \in E} P^n(x, y) \tilde{h}(y) \geq \dots$$

□

*Bemerkung 7.14.* für  $z \in E_0 \subset E$  ( $\phi$  beliebige Funktion) und  $Ph(y) = h(y) \forall y \notin E_0$ .

*Beispiel 7.15* (Galton-Watson-Verzweigungsprozess). Sei  $n$  die Anzahl der Individuen, jedes davon bekommt  $k$  Kinder mit Wahrscheinlichkeit  $\rho(k)$  und stirbt dann. Dann ist also  $E = \mathbb{N}_0$  und

$$p(n, m) = \sum_{\substack{k_1, \dots, k_n=1 \\ \sum_{i=1}^n k_i = m}}^{\infty} \rho(k_1) \dots \rho(k_n) = \sum_{k_1=0}^m \sum_{k_2=0}^{m-k_1} \dots \sum_{k_n=0}^{m-\sum_{i=1}^{n-1} k_i} \rho(k_1) \dots \rho(k_n)$$

Da  $\rho(0, 0) = 1$  ist 0 absorbierend.

Frage: Wie groß ist  $\mathbb{P}(\text{Population stirbt aus} \mid \text{Anfangspopulation}=1)$ . Also was ist  $\mathbb{P}(\tau_0 < \infty)$ ?

Triviale Fälle:

$$\rho(0) = \begin{cases} 1 & \text{dann ist } \mathbb{P}^1(\tau_0 < \infty) = 1 \\ 0 & \text{dann ist } \mathbb{P}^1(\tau_0 < \infty) = 0 \end{cases}$$

Für  $0 < \rho(0) < 1$  gilt:

a) Für alle  $k < n \geq 0$  ist  $\mathbb{P}^k(X_n = 0) = [\mathbb{P}^1(X_n = 0)]^k$ .

*Beweis.* Induktiv

$n = 0$  stimmt

**Induktionsschritt :**

$$\begin{aligned} &= \sum_{\substack{l_1, \dots, l_k=0 \\ \sum_{i=1}^k l_i = l}}^{\infty} \rho(l_1) \dots \rho(l_k) \underbrace{\mathbb{P}^l(X_n = 0)}_{\text{Ind V.} = \mathbb{P}^1(X_n = 0)^l} \\ &= \sum_{l_1, \dots, l_k=0}^{\infty} \rho(l_1) \dots \rho(l_k) (\mathbb{P}^1(X_n = 0))^{l_1 + l_2 + \dots + l_k} \\ &= \left( \sum_{l=0}^{\infty} p(1, l) \mathbb{P}^l(X_n = 0) \right)^k \\ &= (\mathbb{P}(X_{n+1} = 0))^k \end{aligned}$$

□

b) Sei  $q(k) := \mathbb{P}^k(\tau_0 < \infty)$ .

$$\begin{aligned} &= \lim_{n \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}^k(\tau_0 \leq n) = \lim_{n \rightarrow \infty} \mathbb{P}^k(X_n = 0) \\ &\stackrel{a)}{=} \lim_{n \rightarrow \infty} (\mathbb{P}^1(X_n = 0))^k = \left[ \lim_{n \rightarrow \infty} \mathbb{P}^1(X_n = 0) \right]^k = q(1)^k \end{aligned}$$

c)  $q(1) = h_0(1)$  ist harmonisch. d.h.

$$\begin{aligned} q(1) &= Pq(1) = \sum_{k=0}^{\infty} p(1, k)q(k) \\ &\stackrel{b)}{=} \sum_{k=0}^{\infty} \rho(k)q(1)^k \\ &= \varphi_{\rho}(q(1)) \end{aligned}$$

Dabei ist  $\varphi$  die erzeugenden Funktion von  $\rho$  (??).

Also ist  $q(1)$  ein **Fixpunkt** von  $\varphi_{\rho}$  und zwar der kleinste. Denn falls  $s$  ein beliebiger Fixpunkt von  $\varphi_{\rho}$  ist, dann gilt  $\tilde{h}(k) := s^k : P(\tilde{h}) = \tilde{h}$ .

Also ist  $\tilde{h}$  harmonisch bezüglich  $P$ . Da aber  $q(1)$  die minimal harmonische Funktion ist gilt  $s \geq q(1)$ .

Die Funktion  $s \mapsto \varphi_{\rho}(s) = \sum_{k=0}^{\infty} \rho(k)s^k$  hat immer den Fixpunkt  $s = 1$ .

Falls  $\rho(0) + \rho(1) = 1$ , dann ist  $\varphi_{\rho}''(x) = 0$ . Also