

# Stochastik für Studierende der Informatik

## Inhaltsverzeichnis

<b>1</b>	<b>Mathematische Modelle von Zufallsexperimenten</b>	<b>2</b>
1.1	Grundbegriffe . . . . .	2
1.2	Erste Beispiele endlicher Wahrscheinlichkeitsverteilungen . . . . .	5
1.3	Diskrete Wahrscheinlichkeitsräume . . . . .	9
<b>2</b>	<b>Bedingte Wahrscheinlichkeit und Unabhängigkeit</b>	<b>14</b>
<b>3</b>	<b>Zufallsvariablen und ihre Verteilungen</b>	<b>20</b>
3.1	Unabhängige Zufallsvariablen . . . . .	21
3.2	Markovketten . . . . .	32
<b>4</b>	<b>Erwartungswerte und Momente von Zufallsvariablen</b>	<b>37</b>
4.1	Erwartungswerte und ihre Eigenschaften . . . . .	37
4.2	Momente von Zufallsvariablen . . . . .	42
4.3	Anwendung auf Schätzprobleme . . . . .	50
<b>5</b>	<b>Grenzwertsätze</b>	<b>52</b>
<b>6</b>	<b>Wahrscheinlichkeitsmaße auf <math>\mathbb{R}</math></b>	<b>59</b>
6.1	Grundbegriffe reloaded . . . . .	59
6.2	Zufallsvariablen und ihre Erwartungswerte . . . . .	63
6.3	Unabhängigkeit . . . . .	65

# 1 Mathematische Modelle von Zufallsexperimenten

Wir alle haben eine intuitive Vorstellung von Zufall und zufälligen Ereignissen. Zufällige Ereignisse in der Natur und in unserem Alltag sind allgegenwärtig. Viele Begriffe aus der Stochastik sind fester Bestandteil unseres Wortschatzes, sie werden aber meist anders benutzt als wir es hier machen werden. So werden im Alltag die beiden Wörter *wahrscheinlich* und *zufällig* typischerweise benutzt, um Ereignisse zu beschreiben, die eigentlich *ziemlich sicher* beziehungsweise *eher selten und unerwartet* sind. Dieser Gegensatz zwischen den mathematischen Begriffen und dem alltäglichen Sprachgebrauch macht den Einstieg in die *elementare Stochastik* eventuell etwas schwieriger als in manche andere Teilgebiete der Mathematik. Man sollte sich also von Anfang mit den Begriffen streng mathematisch auseinandersetzen und sollte sich nicht zu sehr auf die vermeintliche Intuition aus dem Alltag verlassen.

In Stochastik soll die Intuition aus dem Alltag mit einem mathematischen Fundament untermauert werden. Während zufällige Ereignisse gerade dadurch gekennzeichnet sind, dass deren Ausgang nicht vorherbestimmt ist, erlaubt uns eine mathematische Analyse tiefere Einblicke und Gesetzmäßigkeiten des Zufalls zu verstehen und herzuleiten.

## Was ist Zufall?

*Vorschulgruppe in der Kita:* “Wenn Bjarne die gleiche Hose an hat, wie ich.”

*Wikipedia:* “Von Zufall spricht man, wenn für ein einzelnes Ereignis oder das Zusammentreffen mehrerer Ereignisse keine kausale Erklärung gegeben werden kann. Als kausale Erklärungen für Ereignisse kommen je nach Kontext eher Absichten handelnder Personen oder auch naturwissenschaftliche deterministische Abläufe in Frage.”

*Georgii (2009):* “Eine möglicherweise naturinhärente Indeterminiertheit, als auch unsere (eventuell prinzipielle) Unkenntnis über die genauen Rahmenbedingungen der Situation.”

► Die Antwort auf diese Frage ist eher philosophischer Natur für die mathematische Modellierung und Analyse nebensächlich.

**Was ist Stochastik?** ► Die Lehre von den Gesetzmäßigkeiten des Zufalls.

*Resultate* (die uns im Verlauf der Vorlesung begegnen werden):

- Was ist der erwartete/mittlere Ausgang eines Zufallsexperiments?
- Das Gesetz der großen Zahlen *Beispiel:* Münzexperiment mit fairer Münze. Für eine große Anzahl von Versuchen kommen Anteile von Kopf und Zahl nah an 50%
- Der zentrale Grenzwertsatz.

Wir müssen zunächst „den Zufall“ bzw. Zufallsexperimente mathematisch beschreiben (*Wahrscheinlichkeitstheorie*). In einem zweiten Schritt kann man aufgrund von Beobachtungen Rückschlüsse auf das mathematische Modell ziehen (*Statistik*, nicht im Fokus dieser Vorlesung. (Stochastik 2)).

## 1.1 Grundbegriffe

Wir benötigen zunächst einige Grundbegriffe:

- Der **Grundraum** ist eine nicht leere Menge  $\Omega \neq \emptyset$ . Sie enthält alle möglichen **Ergebnisse** eines Zufallsexperimentes.
- **Ereignisse** sind Teilmengen  $A \subseteq \Omega$ , denen prinzipiell eine Wahrscheinlichkeit zugeordnet werden kann.

- Nicht immer wird jede Teilmenge als Ereignis bezeichnet, sondern nur Mengen aus einem Mengensystem  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ , wobei

$$\mathcal{P}(\Omega) := \{A : A \text{ ist Teilmenge von } \Omega\} \quad \text{die Potenzmenge bezeichnet.}$$

- Ist  $\Omega$  endlich (oder allgemeiner höchstens abzählbar), dann wählt man typischerweise  $\mathcal{A} = \mathcal{P}(\Omega)$ , d.h. jede mögliche Teilmenge des Grundraums soll eine Wahrscheinlichkeit erhalten.

► *Sprechweise:* Ereignis  $A$  tritt ein  $\Leftrightarrow$  Ergebnis  $\omega$  liegt in  $A$

- **Vorsicht:**  $\omega$  ist kein Ereignis! Aber  $\{\omega\}$  ist ein (Elementar-)Ereignis, falls  $\mathcal{A}$  geeignet gewählt ist.

*Beispiel 1.1.*

1. Werfen einer Münze:

$$\Omega = \{K, Z\}$$

Das Ereignis “Es fällt Kopf” entspricht  $A = \{K\}$ .

2. Werfen eines Würfels:

$$\Omega = \{1, \dots, 6\}$$

Das Ereignis “Augenzahl ist gerade” ist gegeben durch  $A = \{2, 4, 6\}$ .

3. In einem Netzwerk werden Längen (in Byte) der ersten  $n = 10^5$  Datenpakete beobachtet, die an einem Router ankommen:

$$\Omega = \mathbb{N}^n = \{(\omega_1, \dots, \omega_n) \mid \omega_i \in \mathbb{N} \text{ für alle } 1 \leq i \leq n\}$$

Interpretation:  $\omega_i =$  Länge des  $i$ -ten Paketes

Das Ereignis “Das größte Paket umfasst maximal  $10^7$  Byte” entspricht

$$A := \{(\omega_1, \dots, \omega_n) \mid \omega_i \leq 10^7 \text{ für alle } 1 \leq i \leq n\}.$$

◇

Logische Verknüpfungen zwischen Bedingungen, die Ereignisse definieren, lassen sich durch mengentheoretische Operationen zwischen diesen Ereignissen beschreiben.

Seien  $A, B \subseteq \Omega$  Ereignisse. Dann ist:

- $A \cup B = \{\omega \in \Omega \mid \omega \in A \text{ oder } \omega \in B\}$  das Ereignis, dass  $A$  eintritt oder  $B$  eintritt (nicht exklusiv, d.h. es können auch beide Ereignisse eintreten),
- $A \cap B = \{\omega \in \Omega \mid \omega \in A \text{ und } \omega \in B\}$  das Ereignis, dass  $A$  und  $B$  eintritt,
- $A \setminus B = \{\omega \in \Omega \mid \omega \in A, \omega \notin B\}$  das Ereignis, dass  $A$  eintritt, aber nicht  $B$  eintritt,
- $B^c$  das Ereignis, dass  $B$  nicht eintritt.
- $A \subseteq B$  bedeutet: Wenn  $A$  eintritt, dann tritt auch  $B$  ein.

Im Wahrscheinlichkeitsmodell soll **jedem** Ereignis eine Wahrscheinlichkeit zugeordnet werden. Diese müssen "zueinander passen". Bspw. muss bei  $A \subseteq B$  gelten:

$$(\text{Wahrscheinlichkeit von } A) \leq (\text{Wahrscheinlichkeit von } B).$$

Man spricht hier von der **Monotonie** von Wahrscheinlichkeiten.

Mathematisch werden alle Wahrscheinlichkeiten durch eine *Abbildung*

$$P: \mathcal{A} \rightarrow \mathbb{R}$$

beschrieben, die wie folgt normiert ist:

- $P(\emptyset) = 0$ , d.h. das Ereignis, das **nie** eintritt, hat W'keit 0.
- $P(\Omega) = 1$ , d.h. das Ereignis, das **immer** eintritt, hat W'keit 1.

Zusammen mit der Monotonie folgt, dass  $P$  nur Werte in  $[0, 1]$  annehmen kann, also

$$P: \mathcal{A} \rightarrow [0, 1].$$

► Welche weiteren Eigenschaften sollte so eine Abbildungen  $P$  erfüllen?

Im einfachsten Fall besteht der Grundraum  $\Omega$  (nur) aus endlich vielen Elementen.

**Definition 1.2.** Ein **endlicher Wahrscheinlichkeitsraum** ist ein Paar  $(\Omega, P)$  bestehend aus einem endlichen Grundraum  $\Omega \neq \emptyset$  und einer Abbildung  $P: \mathcal{P}(\Omega) \rightarrow [0, 1]$  mit den Eigenschaften:

- $P(\Omega) = 1$
- $A, B \subseteq \Omega$  disjunkt  $\implies P(A \cup B) = P(A) + P(B)$  (Additivität)

$P$  heißt dann **Wahrscheinlichkeitsmaß**.

Diese axiomatische Definition von Wahrscheinlichkeitsräumen geht auf Kolmogorov (1933) zurück. Wir werden im Verlauf der Vorlesung auch größere Grundräume zulassen.

*Beispiel 1.3.* 1. Einmaliges Werfen eines fairen Würfels:  
 $\Omega = \{1, \dots, 6\}$ ,  
 $A \subseteq \Omega: P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{6}$ ,  
 insbesondere  $P(\{\omega\}) = \frac{1}{6}$  für alle  $\omega \in \Omega$ .

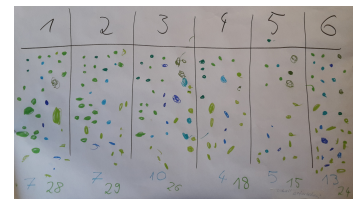


Abb.: Würfelexperiment in der Vorschulgruppe

2. Werfen zweier ununterscheidbarer fairer Würfel:

$$\Omega = \{(\omega_1, \omega_2) \mid \omega_1, \omega_2 \in \{1, \dots, 6\}, \omega_1 \leq \omega_2\}$$

Hier ist es **nicht** sinnvoll, jeder Menge  $\{(\omega_1, \omega_2)\}$  die gleiche Wahrscheinlichkeit zuzuweisen, denn es gibt z.B. 2 Möglichkeiten (1,2) zu erhalten, aber nur eine für (1,1).

$$\rightsquigarrow P(\{(\omega_1, \omega_2)\}) = \begin{cases} \frac{1}{36}, & \text{falls } \omega_1 = \omega_2 \\ \frac{2}{36}, & \text{falls } \omega_1 \neq \omega_2 \end{cases}$$

Für  $A \subseteq \Omega$  folgt  $P(A) = \sum_{\omega \in A} P(\{\omega\})$ .

◇

Wir fassen die wichtigsten elementaren Eigenschaften von Wahrscheinlichkeitsmaßen auf endlichen Räumen in folgendem Satz zusammen.

**Satz 1.4.** Sei  $(\Omega, P)$  ein endlicher Wahrscheinlichkeitsraum. Dann gilt

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

$$(ii) \forall A_1, \dots, A_n \subseteq \Omega \text{ disjunkt, } n \in \mathbb{N}: P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i),$$

$$(iii) \forall A \subseteq \Omega: P(A^c) = 1 - P(A),$$

$$(iv) \forall A, B \subseteq \Omega, A \subseteq B: P(B \setminus A) = P(B) - P(A), \text{ insbes. } P(A) \leq P(B),$$

$$(v) \forall A, B \subseteq \Omega: P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

$$(vi) \forall A_1, \dots, A_n \subseteq \Omega, n \in \mathbb{N}: P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i) \quad (\text{“Subadditivität”}).$$

*Beweis.* (i) bis (v): Übungsaufgaben.

(vi) Für  $A_1, \dots, A_n \subseteq \Omega$  betrachten wir  $B_1 := A_1$  und  $B_i := A_i \setminus (\bigcup_{k=1}^{i-1} A_k) \subseteq A_i$  für  $i = 2, \dots, n$ . Dann sind  $B_1, \dots, B_n$  paarweise disjunkt und es gilt

$$\bigcup_{i=1}^n A_i = \bigcup_{i=1}^n B_i.$$

Wir folgern:

$$P\left(\bigcup_{i=1}^n A_i\right) = P\left(\bigcup_{i=1}^n B_i\right) \stackrel{(ii)}{=} \sum_{i=1}^n P(B_i) \stackrel{(iv)}{\leq} \sum_{i=1}^n P(A_i). \quad \square$$

## 1.2 Erste Beispiele endlicher Wahrscheinlichkeitsverteilungen

Wir werden nun einige wichtige Wahrscheinlichkeitsverteilung auf endlichen Wahrscheinlichkeitsräumen diskutieren lernen.

Betrachten wir den kleinsten nichttrivialen Wahrscheinlichkeitsraum.

**Definition 1.5.** Auf dem Grundraum  $\Omega = \{0, 1\}$  wird durch

$$P(\{1\}) = p \quad \text{für ein } p \in [0, 1]$$

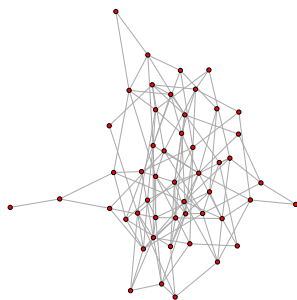
eine Wahrscheinlichkeitsverteilung festgelegt, welche als **Bernoulliverteilung mit Erfolgswahrscheinlichkeit  $p$**  bezeichnet wird.

Es gilt

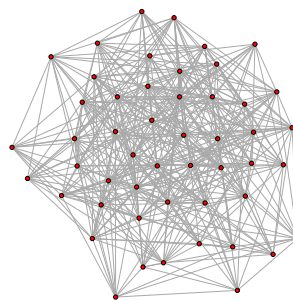
$$P(\{0\}) = 1 - P(\{1\}) = 1 - p,$$

sodass tatsächlich für alle Elemente der Potenzmenge  $\mathcal{P}(\Omega) = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$  Wahrscheinlichkeiten eindeutig festgelegt sind.

*Münzwurfexperiment:* Mit der Kodierung  $0 \hat{=} \text{„Zahl“}$  und  $1 \hat{=} \text{„Kopf“}$  kann das Münzwurfexperiment mit einer Bernoulliverteilung modelliert werden, wobei die Münze fair ist, wenn  $p = 1/2$ .



ER-Graph mit  $n=50$ ,  $p=1/10$



ER-Graph mit  $n=50$ ,  $p=1/3$

Abb.: Realisierungen von Erdős-Rényi-Graphen mit Parametern  $n = 50$  und  $p = 1/10$  (links) und  $p = 1/3$  (rechts).

*Beispiel 1.6.* Ein sehr einfaches Modell für (soziale) Netzwerke ist ein *Erdős-Rényi-Graph*: Wir betrachten  $n \in \mathbb{N}$  Individuen, wobei sich zwei Personen  $(i, j) \in \{i, j \in \{1, \dots, n\} : i \neq j\}$  mit Wahrscheinlichkeit  $p \in (0, 1)$  gegenseitig mögen/folgen/likern/zitieren/... Wir stellen die Individuen als Knoten eines (ungerichteten) Graphen dar. Für jedes Paar  $(i, j)$  führen wir ein Bernoulliexperiment mit Erfolgswahrscheinlichkeit  $p$  durch. Falls das Ereignis  $\{1\}$  eintritt, verbinden wir die Knoten  $i, j$  mit einer Kante.  $\diamond$

Das Modell für den fairen Würfel ist ein Spezialfall einer wichtigen Klasse von Wahrscheinlichkeitsräumen:

**Definition 1.7.** Ist  $\Omega \neq \emptyset$  endlich, so heißt das durch

$$P(A) := \frac{|A|}{|\Omega|}, \quad A \subseteq \Omega,$$

definierte Wahrscheinlichkeitsmaß die **Gleichverteilung** oder **Laplace-Verteilung** auf  $\Omega$ . Hierbei bezeichnet  $|A|$  die Anzahl der Elemente der Menge  $A$ .

Bei Gleichverteilungen sind also nur Mächtigkeiten von Mengen zu bestimmen. Dafür sind oft sog. Urnenmodelle und Kombinatorik hilfreich.

### Urnenmodelle

Es werden  $k$  Ziehungen aus einer Urne mit Kugeln, die mit  $1, \dots, n$  durchnummeriert sind, durchgeführt.

Dabei betrachten wir wahlweise

- Ziehen **mit** oder **ohne Zurücklegen**,
- Ziehen **mit** oder **ohne Berücksichtigung der Reihenfolge**.

Im Folgenden werden geeignete Grundräume für diese Zufallsexperimente und deren Mächtigkeiten angegeben.

### Ziehen mit Zurücklegen und mit Berücksichtigung der Reihenfolge:

Möglicher *Grundraum*:

$$\Omega_{mZ,mR} = \{(\omega_1, \dots, \omega_k) \mid \omega_i \in \{1, \dots, n\} \forall 1 \leq i \leq k\} = \{1, \dots, n\}^k$$

*Interpretation:*  $\omega = (\omega_1, \dots, \omega_k) \in \Omega_{mZ, mR}$  bedeutet, dass für alle  $1 \leq i \leq k$  im  $i$ -ten Zug die Kugel mit der Nummer  $\omega_i$  gezogen worden ist.

► Es gilt  $|\Omega_{mZ, mR}| = n^k$ .

### Ziehen ohne Zurücklegen und mit Berücksichtigung der Reihenfolge:

Möglicher Grundraum:

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

*Interpretation:*  $\omega = (\omega_1, \dots, \omega_k) \in \Omega_{oZ, mR}$  bedeutet, dass für alle  $1 \leq i \leq k$  im  $i$ -ten Zug die Kugel mit der Nummer  $\omega_i$  gezogen worden ist.

► Es gilt  $|\Omega_{oZ, mR}| = n \cdot (n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!}$ .

*Begründung:*  $n$  Möglichkeiten für  $\omega_1$ ,  
dann  $n-1$  Möglichkeiten für  $\omega_2$  bei gegebenem  $\omega_1$ ,  
dann  $n-2$  Möglichkeiten für  $\omega_3$  bei gegebenen  $\omega_1, \omega_2$ ,  
...  
dann  $n-k+1$  Möglichkeiten für  $\omega_k$  bei gegebenen  $\omega_1, \dots, \omega_{k-1}$ .

### Ziehen ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge:

Möglicher Grundraum:

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

*Interpretation:*  $\omega = (\omega_1, \dots, \omega_k) \in \Omega_{oZ, oR}$  bedeutet, dass die **der Größe nach geordneten** gezogenen Kugelnummern gerade  $\omega_1, \dots, \omega_k$  sind.

► Es gilt  $|\Omega_{oZ, oR}| = \frac{|\Omega_{oZ, mR}|}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$ .

*Begründung:* Es gibt zu jedem  $(\omega_1, \dots, \omega_k) \in \Omega_{oZ, oR}$  genau  $k!$  verschiedene Möglichkeiten, die  $k$  verschiedenen Kugelnummern anzuordnen, d.h. jedem  $\omega \in \Omega_{oZ, oR}$  entsprechen  $k!$  Elemente aus  $\Omega_{oZ, mR}$ .

*Alternatives Modell:*  $\tilde{\Omega}_{oZ, oR} := \{S \subseteq \{1, \dots, n\} \mid |S| = k\}$ , wobei  $S \in \tilde{\Omega}_{oZ, oR}$  als die Menge der gezogenen Kugelnummern zu interpretieren ist.

### Ziehen mit Zurücklegen und ohne Berücksichtigung der Reihenfolge:

Möglicher Grundraum:

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

*Interpretation:*  $\omega = (\omega_1, \dots, \omega_k) \in \Omega_{mZ, oR}$  bedeutet, dass die der Größe nach geordneten gezogenen Kugelnummern gerade  $\omega_1, \dots, \omega_k$  sind.

Um  $|\Omega_{mZ, oR}|$  zu bestimmen betrachte die bijektive Abbildung

$$S : \Omega_{mZ, oR} \rightarrow \Omega^* := \{(\tilde{\omega}_1, \dots, \tilde{\omega}_k) \in \{1, \dots, n+k-1\}^k \mid \tilde{\omega}_1 < \tilde{\omega}_2 < \dots < \tilde{\omega}_k\}$$

$$S((\omega_1, \dots, \omega_k)) := (\omega_1, \omega_2 + 1, \omega_3 + 2, \dots, \omega_k + k - 1)$$

$\Omega^*$  ist vom gleichen Typ wie  $\Omega_{oZ, oR}$  mit  $n$  ersetzt durch  $n+k-1$

► Daher gilt  $|\Omega_{mZ, oR}| = |\Omega^*| = \binom{n+k-1}{k}$ .

*Beispiel 1.8.* 6 Kugeln werden ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge aus 49 Kugeln gezogen.

⇒ Es gibt  $\binom{49}{6} = 13.983.816$  Möglichkeiten

Da alle Kombinationen gleich wahrscheinlich sind, ist die W'keit für 6 Richtige

$$\frac{1}{\binom{49}{6}} \approx 7,15 \cdot 10^{-8}. \quad \diamond$$

*Beispiel 1.9.* Sie haben am Geldautomaten die PIN für Ihre EC-Karte vergessen, wissen aber noch, dass es der Geburtstag einer Ihrer 10 Cousinen war. Sie probieren die 10 möglichen PINs in zufälliger Reihenfolge am Geldautomanten durch, bis dieser evtl. die Karte nach drei falschen Versuchen einzieht.

► Wie groß ist die W'keit, dass Sie genau 3 Versuche benötigen?

*Situation:* 3 mal Ziehen ohne Zurücklegen und mit Berücks. der Reihenfolge

$$\leadsto \Omega = \{(\omega_1, \omega_2, \omega_3) \in \{1, \dots, 10\}^3 \mid \omega_i \neq \omega_j \forall i \neq j\} \Rightarrow |\Omega| = \frac{10!}{7!} = 720$$

Es kann *Gleichverteilung* angenommen werden.

Ist  $r \in \{1, \dots, 10\}$  die richtige PIN, entspricht „Erfolg im dritten Versuch“ dem Ereignis

$$A = \{(\omega_1, \omega_2, \omega_3) \in \Omega \mid \omega_3 = r\} \Rightarrow |A| = (10 - 1) \cdot (9 - 1) \cdot 1 = 72.$$

► Bei Gleichverteilung gilt also  $P(A) = \frac{|A|}{|\Omega|} = \frac{72}{720} = 0,1.$  ◇

Den Urnenmodellen sehr ähnlich sind **Fächermodelle:**

Wir haben  $k$  Murmeln und wollen diese auf  $n$  Fächer verteilen. Dabei gelte wahlweise:

- Mehrfachbelegung der Fächer **erlaubt** oder **nicht erlaubt**
- **unterscheidbare** Murmeln oder **nicht unterscheidbare** Murmeln.

► Wie viele verschiedene Ergebnisse gibt es jeweils?

► Die resultierenden 4 möglichen Situationen entsprechen den 4 Urnenmodellen.

*Begründung:* Nummeriere die Fächer von 1 bis  $n$  durch.

Nimm eine Urne mit  $n$  durchnummerierten Kugeln.

Für jede Murmel ziehen wir jetzt eine Kugel aus der Urne und legen die Murmel in das Fach mit der Nummer, die auf der Kugel steht.

*Mehrfachbelegung* möglich  $\Leftrightarrow$  *Ziehung mit Zurücklegen*

*Murmeln unterscheidbar*  $\Leftrightarrow$  *Reihenfolge berücksichtigt.*

Es ergibt sich folgender Zusammenhang zwischen Urnen- und Fächermodellen.

Angegeben ist jeweils die Mächtigkeit des Grundraums

Urnenmodell	mit Zurücklegen	ohne Zurücklegen	
mit $n$ Kugeln und $k$ Ziehungen			
mit Reihenfolge	$n^k$	$\frac{n!}{(n-k)!}$	unterscheidbare Murmeln
ohne Reihenfolge	$\binom{n+k-1}{k}$	$\binom{n}{k}$	unterscheidbare Murmeln
	mit Mehrfachbelegung	ohne Mehrfachbelegung	Verteilung von $k$ Murmeln auf $n$ Fächer

**Warnhinweis:**

Urnenmodelle ermöglichen es, Anzahlen vieler Mengen zu bestimmen, und damit die auch Wahrscheinlichkeiten, **falls** von einer Laplace-Verteilung/Gleichverteilung ausgegangen werden kann. Allerdings ist die Annahme einer Gleichverteilung in vielen Fällen **nicht** gerechtfertigt, insbes. oft nicht im Modell mit Zurücklegen und ohne Berücksichtigung der Reihenfolge!

**Erinnerung Beispiel 1.3(2):** Werfen zweier ununterscheidbarer fairer Würfel. Der Grundraum

$$\Omega = \{(\omega_1, \omega_2) \mid \omega_1, \omega_2 \in \{1, \dots, 6\}, \omega_1 \leq \omega_2\}$$

ist vom Typ  $\Omega_{mZ, oR}$  mit  $k = 2, n = 6$ , **aber** Zufallsmechanismus wird **nicht** durch Gleichverteilung beschrieben, da z.B.

$$P(\{(1, 2)\}) = 2/36 \neq 1/36 = P(\{(1, 1)\}).$$

**1.3 Diskrete Wahrscheinlichkeitsräume**

Im Beispiel zur Größe von Datenpaketen liegt kein **endlicher**, sondern ein **abzählbarer** Grundraum vor. Die Additivität ist dann nicht mehr ausreichend, um eine vielseitig einsetzbare Theorie zu entwickeln.

**Definition 1.10.** Sei  $\Omega \neq \emptyset$  eine beliebige nicht-leere Menge. Eine Abbildung  $P: \mathcal{P}(\Omega) \rightarrow [0, 1]$  heißt **diskretes Wahrscheinlichkeitsmaß**, falls

- (a)  $P(\Omega) = 1$ ,
- (b)  $\forall A_n \subseteq \Omega, n \in \mathbb{N}, \text{disjunkt: } P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n)$  ( $\sigma$ -Additivität),
- (c) es existiert eine (höchstens) abzählbare Menge  $\Omega_0 \subseteq \Omega$  mit  $P(\Omega_0) = 1$ .

Dann heißt  $(\Omega, P)$  **diskreter Wahrscheinlichkeitsraum**.

Man beachte, dass wir nun in (b) *abzählbar unendlich viele* disjunkte Mengen zulassen.

Satz 1.4 gilt analog auch für diskrete Wahrscheinlichkeitsräume:

**Satz 1.11.** Sei  $(\Omega, P)$  ein diskreter Wahrscheinlichkeitsraum. Dann gilt

- (i)  $P(\emptyset) = 0$
- (ii)  $\forall A_1, \dots, A_k \subseteq \Omega \text{ disjunkt, } k \in \mathbb{N}: P\left(\bigcup_{i=1}^k A_i\right) = \sum_{i=1}^k P(A_i)$
- (iii)  $\forall A \subseteq \Omega: P(A^c) = 1 - P(A)$
- (iv)  $\forall A, B \subseteq \Omega, A \subseteq B: P(B \setminus A) = P(B) - P(A)$  Insbes.  $P(A) \leq P(B)$
- (v)  $\forall A, B \subseteq \Omega: P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- (vi) Für alle Folgen  $A_n \subseteq \Omega, n \in \mathbb{N}$  gilt  $P\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} P(A_n)$  (" $\sigma$ -Subadditivität")

Man beachte, dass wir die Subadditivität in (vi) von endlich vielen Mengen auf abzählbar viele Mengen erweitert haben.

*Beweis.* Da  $\sigma$ -Additivität insbesondere Additivität impliziert (setze  $A_n = \emptyset$  für alle  $n > k$ ), kann man (i) bis (v) exakt wie in Satz 1.4 zeigen. Für (vi) gehen wir ebenfalls analog vor, wobei wir statt (ii) die  $\sigma$ -Additivität von  $P$  verwenden.  $\square$

Diskrete W'maße lassen sich durch eine Funktion  $f: \Omega \rightarrow [0, 1]$  eindeutig beschreiben:

**Satz 1.12.**

(i) Sei  $(\Omega, P)$  ein diskreter Wahrscheinlichkeitsraum. Dann wird die Funktion  $f: \Omega \rightarrow [0, 1], f(\omega) = P(\{\omega\})$  **Wahrscheinlichkeitsfunktion** oder **Zähldichte** von  $P$  genannt und besitzt folgende Eigenschaften:

(a)  $\Omega_T := \{\omega \in \Omega \mid f(\omega) > 0\}$  ist abzählbar (und heißt **Träger von  $P$**  bzw.  $f$ ),

(b)  $\sum_{\omega \in \Omega} f(\omega) = 1$ .

Für alle  $A \subseteq \Omega$  gilt dann:

$$P(A) = \sum_{\omega \in A} f(\omega) = \sum_{\omega \in A \cap \Omega_T} f(\omega).$$

(ii) Ist umgekehrt  $\Omega \neq \emptyset$  und  $f: \Omega \rightarrow [0, 1]$  eine Funktion, die (a) und (b) erfüllt, so existiert genau ein diskretes Wahrscheinlichkeitsmaß auf  $\Omega$ , das  $f$  zur Zähldichte hat.

*Beweis.* (i) Wegen Eigenschaft (c) aus der Definition gilt für alle  $\omega \notin \Omega_0$ , dass  $0 \leq f(\omega) = P(\{\omega\}) \leq P(\Omega_0^c) = 1 - P(\Omega_0) = 0$ . Also ist  $\Omega_T \subseteq \Omega_0$  abzählbar. Damit haben wir (a) gezeigt. (b) folgt aus

$$\sum_{\omega \in \Omega} f(\omega) \stackrel{\Omega_T \subseteq \Omega_0}{=} \sum_{\omega \in \Omega_0} f(\omega) = \sum_{\omega \in \Omega_0} P(\{\omega\}) \stackrel{\sigma\text{-additiv}}{=} P(\Omega_0) = 1.$$

Weiterhin gilt für alle  $A \subseteq \Omega$ , dass  $P(A \setminus \Omega_0) \leq P(\Omega_0^c) = 0$  gilt, sodass

$$\begin{aligned} P(A) &= P(A \cap \Omega_0) + P(A \setminus \Omega_0) \\ &= P(A \cap \Omega_0) \\ &\stackrel{\sigma\text{-additiv}}{=} \sum_{\omega \in A \cap \Omega_0} P(\{\omega\}) \\ &= \sum_{\omega \in A \cap \Omega_0} f(\omega) \end{aligned}$$

folgt. □

*Beweis.* Hieraus folgt einerseits

$$P(A) = \sum_{\omega \in A \cap \Omega_0} f(\omega) = \sum_{\omega \in A \cap \Omega_T} f(\omega) - \sum_{\omega \in (A \cap \Omega_0) \setminus \Omega_T} \underbrace{f(\omega)}_{=0 \forall \omega \in \Omega_T^c} = \sum_{\omega \in A \cap \Omega_T} f(\omega)$$

und analog

$$P(A) = \sum_{\omega \in A \cap \Omega_T} f(\omega) = \sum_{\omega \in A} f(\omega) - \sum_{\omega \in A \setminus \Omega_T} f(\omega) = \sum_{\omega \in A} f(\omega).$$

Man beachte, dass in der letzten Summe nur höchstens abzählbar viele Summanden von null verschieden sind. □

*Beweis. (ii) Eindeutigkeit:* Ist  $f$  die Zähldichte zweier Wahrscheinlichkeitsmaße  $P$  und  $Q$  auf  $\Omega$ , dann gilt:

$$P(A) = \sum_{\omega \in A} f(\omega) = Q(A) \quad \forall A \subseteq \Omega.$$

Also stimmen die beiden Abbildungen  $P: \mathcal{P}(\Omega) \rightarrow [0, 1]$  und  $Q: \mathcal{P}(\Omega) \rightarrow [0, 1]$  überein.  $\square$

*Beweis. Existenz:* Wir definieren eine Abbildung  $P: \mathcal{P}(\Omega) \rightarrow [0, 1]$  auf  $\Omega$  mittels

$$P(A) := \sum_{\omega \in A} f(\omega) \quad \forall A \subseteq \Omega.$$

Dann gilt  $P(\Omega) = \sum_{\omega \in \Omega} f(\omega) = 1$ , für alle disjunkten  $A_n \subseteq \Omega, n \in \mathbb{N}$  gilt

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{\omega \in \bigcup_{n \in \mathbb{N}} A_n} f(\omega) = \sum_{n \in \mathbb{N}} \left( \sum_{\omega \in A_n} f(\omega) \right) = \sum_{n \in \mathbb{N}} P(A_n),$$

und für  $\Omega_0 := \Omega_T = \{\omega \in \Omega : f(\omega) > 0\}$  gilt

$$P(\Omega_0) = \sum_{\omega \in \Omega_0} f(\omega) = \sum_{\omega \in \Omega} f(\omega) - \sum_{\omega \in \Omega \setminus \Omega_0} f(\omega) = 1.$$

Somit ist  $P$  ein diskretes Wahrscheinlichkeitsmaß auf  $\Omega$  mit der Eigenschaft  $P(\{\omega\}) = f(\omega)$  für alle  $\omega \in \Omega$ .  $\square$

Für  $p \in [0, 1]$  liefert der Binomische Lehrsatz:

$$1 = (p + (1 - p))^n = \sum_{k=0}^n \underbrace{\binom{n}{k} p^k \cdot (1 - p)^{n-k}}_{=: f(k)}$$

►  $f: \{0, \dots, n\} \rightarrow [0, 1]$ , erfüllt die Bedingungen (a) und (b) aus Satz 1.12, ist also die Zähldichte eines diskreten Wahrscheinlichkeitsmaßes auf  $\Omega$ .

**Definition 1.13.** Das Wahrscheinlichkeitsmaß  $P = \mathcal{B}_{(n,p)}$  auf  $\{0, \dots, n\}$  mit der Zähldichte

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

heißt **Binomialverteilung** mit Parametern  $n \in \mathbb{N}$  und  $p \in [0, 1]$ .

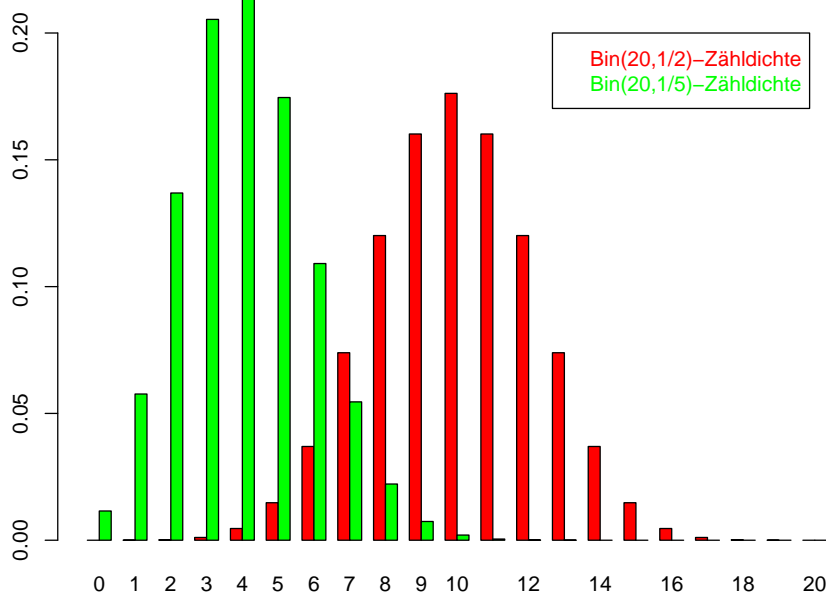


Abb.: Zähldichten der Binomialverteilung mit  $n = 20$  und Erfolgswahrscheinlichkeiten  $p = \{1/2, 1/5\}$ .

Später:  $\mathcal{B}_{(n,p)}$  beschreibt zufällige Anzahl der Erfolge bei  $n$ -maliger unabhängiger Durchführung eines Zufallsexperiments mit Erfolgswahrscheinlichkeit  $p$ . Wir sehen aber bereits hier, dass wir für  $n = 1$  gerade die Bernoulliverteilung erhalten.

Auf eine weitere wichtige Verteilungsfamilie führt folgende Frage:

Beispiel: Wie groß ist die W'keit, dass man genau  $k$ -mal (unabh. voneinander) einen fairen Würfel werfen muss, **bevor** das erste Mal 6 gewürfelt wird?

W'keit,  $k$ -mal **keine** 6 zu würfeln:  $(\frac{5}{6})^k$ , W'keit, beim  $(k + 1)$ . Mal 6 zu würfeln:  $\frac{1}{6}$ .

► Gesuchte Wahrscheinlichkeit =  $(\frac{5}{6})^k \cdot \frac{1}{6} \quad \forall k \in \mathbb{N}_0$

**Definition 1.14.** Das Wahrscheinlichkeitsmaß  $P = \mathcal{G}_p$  auf  $\mathbb{N}_0$  mit der Zähldichte

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

heißt **geometrische Verteilung** mit Parameter  $p \in (0, 1]$ .

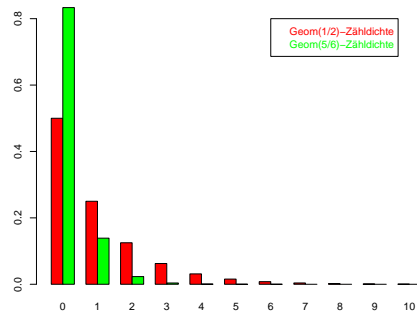


Abb.: Zähldichte der geometrischen Verteilung mit  $p \in \{1/2, 5/6\}$ .

$f$  ist tatsächlich eine Zähldichte auf  $\mathbb{N}_0$ , denn  $\mathbb{N}_0$  ist abzählbar und

$$\sum_{k \in \mathbb{N}_0} f(k) = p \cdot \sum_{k \in \mathbb{N}_0} (1 - p)^k \stackrel{\text{geometr. Reihe}}{=} p \cdot \frac{1}{1 - (1 - p)} = 1.$$

$\mathcal{G}_p$  beschreibt Anzahl der Misserfolge bis zum ersten Erfolg bei unabhängiger Durchführung eines Zufallsexperiments mit Erfolgsw'keit  $p$ .

## 2 Bedingte Wahrscheinlichkeit und Unabhängigkeit

*Frage:* Wie kann man den Einfluss eines Ereignisses  $B$  auf die Eintrittsw'keit eines anderen Ereignisses  $A$  beschreiben?

*Idee:* Vergleiche die W'keit, dass beide Ereignisse eintreten, mit der W'keit, dass das Ereignis  $B$  eintritt.

*Motivationsbeispiel:* Zwei W'rfel werden geworfen, wobei einer offen liegt und einer verdeckt ist. Die Augenzahl des offenen W'rfels ist also bekannt. Was ist die Wahrscheinlichkeit, dass die Summe der beiden Augenzahlen  $\geq 9$  ist?

Ist die bekannte Augenzahl in  $\{1, 2\}$ , dann ist die gesuchte W'keit = 0.

Ist die bekannte Augenzahl in  $\{3, \dots, 6\}$ , so ist es m'glich mit dem zweiten W'rfel auf die Augensumme  $\geq 9$  zu kommen. Wie bestimmt man die W'keit?

**Definition 2.1.** Sei  $(\Omega, P)$  ein diskreter Wahrscheinlichkeitsraum,  $B \subseteq \Omega$  mit  $P(B) > 0$  und  $A \subseteq \Omega$ . Dann hei'ft

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

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

*Bemerkung:* Man beachte, dass f'ur festes  $B \subseteq \Omega$  mit  $P(B) > 0$ , die Abbildung  $A \mapsto P(A|B)$  wieder ein diskretes Wahrscheinlichkeitsma' auf  $\Omega$  ist.

*Rechnung im Motivationsbeispiel:*  $\Omega = \{\omega_1, \omega_2\} : \omega_i \in \{1, \dots, 6\}, i \in \{1, 2\}$  ausgestattet mit Gleichverteilung auf  $\Omega$ . Sei  $B_i = \{\omega_1 = i\}$ ,  $i \in \{1, 2\}$  und  $A = \{\omega_1 + \omega_2 \geq 9\}$ .

Dann ist  $P(A|B_1) = P(A|B_2) = 0$  und z.B.  $P(A|B_3) = \frac{P(\{\omega_1=3, \omega_2=6\})}{P(\{\omega_1=3\})} = \frac{1/36}{1/6} = \frac{1}{6}$ .

*Beispiel 2.2.*

Beim Skat erh'lt jeder der 3 Spieler 10 Karten, 2 Karten kommen in den Stock.

*Modell:*  $\Omega = \{\omega = (\omega_1, \dots, \omega_{32}) \in K^{32} \mid \omega_i \neq \omega_j \forall i \neq j\}$   
mit Menge aller Karten  $K = \{\diamond A, \spadesuit A, \heartsuit A, \clubsuit A, \diamond K, \dots, \clubsuit 7\}$ ,  
 $P$  Gleichverteilung auf  $\Omega$ .

*Interpretation:*  $\omega_1, \dots, \omega_{10}$ : Karten von Spieler 1,  
 $\omega_{11}, \dots, \omega_{20}$ : Karten von Spieler 2,  
 $\omega_{21}, \dots, \omega_{30}$ : Karten von Spieler 3,  
 $\omega_{31}, \omega_{32}$ : Stock

Spieler 1 erh'lt keinen Buben und m'chte nun die W'keit bestimmen, dass einer der anderen Spieler alle 4 Buben h'lt.

Sei

$B := \{\omega \in \Omega \mid \{\omega_1, \dots, \omega_{10}\} \cap \{\diamond B, \dots, \clubsuit B\} = \emptyset\} \hat{=} \text{„Spieler 1 hat keinen Buben“},$

$A_i := \{\omega \in \Omega \mid \{\diamond B, \dots, \clubsuit B\} \subseteq \{\omega_{10i-9}, \dots, \omega_{10i}\}\} \hat{=} \text{„Spieler } i \text{ hat alle Buben“}.$

► Gesucht ist  $P(A_2 \cup A_3 | B)$ .

Es gilt

$$\begin{aligned}
 P(A_2 \cup A_3 | B) &= \frac{P((A_2 \cup A_3) \cap B)}{P(B)} \\
 &\stackrel{A_2, A_3 \subseteq B}{=} \frac{P(A_2 \cup A_3)}{P(B)} \\
 &\stackrel{A_2, A_3 \text{ disjunkt}}{=} \frac{P(A_2) + P(A_3)}{P(B)} \\
 &= \frac{|A_2|/|\Omega| + |A_3|/|\Omega|}{|B|/|\Omega|} = \frac{2|A_2|}{|B|},
 \end{aligned}$$

wobei

$|B| = \frac{28!}{18!} \cdot 22! =$  Anz. Mögl., 10 Karten von Spieler 1 ohne Zurückl. aus 28 Nicht-Buben zu ziehen und übrige 22 Karten bel. zu verteilen

$|A_2| = \frac{10!}{6!} \cdot 28! =$  Anz. Mögl., 4 Buben auf 10 Karten von Spieler 2 zu verteilen und die 28 anderen Karten beliebig

$$\blacktriangleright P(A_2 \cup A_3 | B) = 2 \frac{|A_2|}{|B|} = 2 \frac{10! \cdot 28! \cdot 18!}{6! \cdot 28! \cdot 22!} = \frac{12}{209} \approx 0,057.$$

Im Vergleich die W'keit vor Verteilen der Karten, dass irgendein Spieler alle 4 Buben erhält

$$3P(A_2) = 3 \frac{|A_2|}{|\Omega|} = 3 \frac{10! \cdot 28!}{6! \cdot 32!} = \frac{63}{3596} \approx 0,0175. \quad \diamond$$

**Satz 2.3.** Sei  $(\Omega, P)$  ein diskreter Wahrscheinlichkeitsraum,  $I$  eine (höchstens) abzählbare Indexmenge,  $B_i \subseteq \Omega, i \in I$ , disjunkt mit  $P(B_i) > 0$  und  $\bigcup_{i \in I} B_i = \Omega$  (d.h.  $B_i, i \in I$ , bildet eine disjunkte Zerlegung von  $\Omega$ ). Weiter sei  $A \subseteq \Omega$  beliebig.

(i) Es gilt der Satz von der totalen Wahrscheinlichkeit:

$$P(A) = \sum_{i \in I} P(A | B_i) \cdot P(B_i).$$

(ii) Falls  $P(A) > 0$  und  $k \in I$ , dann gilt der Satz von Bayes:

$$P(B_k | A) = \frac{P(A | B_k) \cdot P(B_k)}{\sum_{i \in I} P(A | B_i) \cdot P(B_i)}.$$

*Beweis.* (i) Für  $(B_i)_{i \in I} \subseteq \Omega$  und  $A \subseteq \Omega$  wie angegeben gilt

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

Nun sind alle  $A \cap B_i, i \in I$ , disjunkt, sodass aus der  $\sigma$ -Additivität und der Definition der bedingten Erwartungen folgt:

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

(ii) Gilt zusätzlich  $P(A) > 0$ , erhalten wir für jedes  $k \in I$ :

$$\begin{aligned}
 P(B_k | A) &= \frac{P(B_k \cap A)}{P(A)} \stackrel{(i)}{=} \frac{P(B_k \cap A)}{\sum_{i \in I} P(A | B_i) \cdot P(B_i)} \\
 &= \frac{\frac{P(B_k \cap A)}{P(B_k)} \cdot P(B_k)}{\sum_{i \in I} P(A | B_i) \cdot P(B_i)} = \frac{P(A | B_k) \cdot P(B_k)}{\sum_{i \in I} P(A | B_i) \cdot P(B_i)}. \quad \square
 \end{aligned}$$

*Beispiel 2.4* (Brustkrebsvorsorge).

Röntgen-Mammographie ist eine Möglichkeit, Brustkrebs vorzeitig zu erkennen.

In ca. 83% der Fälle, in denen ein Tumor (in irgendeinem Stadium) vorhanden ist, ist der Test positiv, d.h. der Tumor wird erkannt.

In 97% der Fälle, in denen kein Tumor vorliegt, fällt der Test negativ aus.

Wir betrachten nun die Ereignisse:

$K$  : Es ist ein Tumor vorhanden.

$T$  : Der Test ist positiv

Dann gilt:

$$\underbrace{P(T | K) = 0,83}_{\text{Sensitivität des Tests}} \quad \text{und} \quad \underbrace{P(T^c | K^c) = 0,97}_{\text{Spezifität des Tests}} \implies P(T | K^c) = 0,03$$

► Was bedeutet es nun, wenn der Test positiv ausfällt?

Wir wenden den Satz von Bayes auf die Zerlegung  $\Omega = K \cup K^c$  und  $A = T$  an:

$$P(K | T) \stackrel{\text{Bayes}}{=} \frac{P(T | K) \cdot P(K)}{P(T | K) \cdot P(K) + P(T | K^c) \cdot P(K^c)} = \frac{1}{1 + \frac{P(T | K^c) \cdot P(K^c)}{P(T | K) \cdot P(K)}}$$

Wird der Test ohne besondere Indikation angewendet, so entspricht  $P(K)$  der relativen Häufigkeit, mit der Brustkrebs in der betrachteten Bevölkerungsgruppe auftritt.

Die Werte liegen dann etwa bei (Quelle: Zentrum für Krebsregisterdaten, Datenbasis 2014)

$$35 - 45\text{-Jährige: } P(K) \approx \frac{1}{110} \implies P(K | T) \approx 0,202$$

$$55 - 65\text{-Jährige: } P(K) \approx \frac{1}{34} \implies P(K | T) \approx 0,456$$

*Folgerung:*

Während bei einer um die 60-Jährigen also etwa mit einer Wahrscheinlichkeit von über 45% tatsächlich ein Tumor vorhanden ist, wenn der Test positiv ausfällt, ist dies bei um die 40-Jährigen nur mit einer Wahrscheinlichkeit von 20% der Fall (immer immernoch deutlich über der W'keit von  $\frac{1}{110} \approx 0,009$ ).

Generell ist die Aussagekraft von derartigen medizinischen Tests selbst bei hoher Spezifität sehr beschränkt, wenn die Erkrankungswahrscheinlichkeit sehr gering ist. (Die Sensitivität spielt bei der Berechnung der Erkrankungsw'keit im Fall seltener Erkrankungen eine deutlich geringere Rolle, so lange sie merklich größer als 50% ist.)

◇

*Beispiel 2.5* (Simpson-Paradoxon).

Bei den 6 Fächern mit den höchsten Bewerberzahlen wurden 1973 an der Universität Berkeley ca. 44,5% der männlichen, aber nur etwa 30,3% der weiblichen Bewerber zugelassen.

Fach	Männer		Frauen	
	Anzahl der Bewerber	Zulassungsquote (in %)	Anzahl der Bewerberinnen	Zulassungsquote (in %)
1	825	62	108	82
2	560	63	25	68
3	325	37	593	34
4	417	33	375	35
5	191	28	393	24
6	373	6	341	7
insg.	2691	<b>44,5</b>	1835	<b>30,3</b>

► Ist die geringere Annahmequote bei Frauen ein Zeichen für Geschlechterdiskriminierung? Wir setzen natürlich voraus, dass bei beiden Geschlechtern der Anteil Qualifizierter gleich hoch ist. In 4 der 6 Fächer ist die Annahmequote der Frauen höher, in den beiden anderen nur geringfügig niedriger. Dennoch ist die Gesamtannahmequote bei den Frauen wesentlich niedriger.

*Grund:* Die Frauen haben sich bevorzugt in Fächern mit hoher Ablehnungsquote beworben (> 51% der Männer haben sich bei den beiden Fächern mit der höchsten Zulassungsquote beworben, aber nur  $\approx 7\%$  der Frauen).

*Was ist der Zusammenhang zur bedingten Wahrscheinlichkeit?*

Wir modellieren der Einfachheit halber die Situation für Frauen und Männer getrennt (aber analog). Hier sei das Modell für eine zufällig unter allen Bewerberinnen ausgewählte Frau vorgestellt.

Setze  $\Omega = \{1, \dots, 6\} \times \{1, 0\}$ , wobei für  $(f, b) \in \Omega$   $f$  das gewählte Fach angibt,  $b = 1$  bedeutet, dass Bewerberin zugelassen, und  $b = 0$ , dass sie abgelehnt wurde.

Dann ist

$$F_i := \{(i, 1), (i, 0)\} \hat{=} \text{„sie hat sich auf Fach } i \text{ beworben“}, 1 \leq i \leq 6,$$

$$Z := \{(i, 1) \mid 1 \leq i \leq 6\} \hat{=} \text{„sie wurde zugelassen“}.$$

Das  $W$ -maß  $P_w$ , das die Situation der Frauen beschreibt, ist vollständig bestimmt durch die Angaben

$$P_w(F_i) = \text{Anteil der Frauen, die sich auf Fach } i \text{ beworben haben,}$$

$$P_w(Z|F_i) = \text{Annahmequote für Frauen im Fach } i.$$

(Warum?)

Sei  $(\Omega, P_m)$  das analoge Modell für die Männer.

► Die Annahmew'keit beträgt (Satz von der totalen W'keit) für Frauen/Männer:

$$P_w(Z) = \sum_{i=1}^6 P_w(Z|F_i) \cdot P_w(F_i),$$

$$P_m(Z) = \sum_{i=1}^6 P_m(Z|F_i) \cdot P_m(F_i).$$

►  $P_w(Z) < P_m(Z)$  kann auftreten obwohl  $P_w(Z|F_i) \geq P_m(Z|F_i)$ , wenn nur die kleinen bedingten W'keiten  $P_w(Z|F_i)$  mit großen Gewichten  $P_w(F_i)$  in die Summe eingehen.

► Das „Paradoxon“ besteht also i.W. nur in einer unzulässigen Gleichsetzung von bedingten und „normalen/unbedingten“ Wahrscheinlichkeiten!

► Alle relevanten Einflussfaktoren (hier: die Fächerwahl) müssen berücksichtigt werden, wenn man den Einfluss eines Merkmals (hier: das Geschlecht) auf eine Zielgröße (hier: die Zulassungsquote) untersuchen möchte, es sei denn, man kann sicherstellen, dass diese weiteren Einflussfaktoren für alle Ausprägungen des eigentlich interessierenden Merkmals gleich sind (im vorliegenden Fall also  $P_w(F_i) = P_m(F_i)$  gelten würde).

*Anwendung in medizinischen Studien:* Beim Vergleich der Wirksamkeit eines neuen Medikaments mit der eines anderen Medikaments oder eines Placebos werden die teilnehmenden Patienten zufällig auf die beiden Gruppen (also diejenigen, die das neue Medikament erhalten, und die Kontrollgruppe derjenigen, die das andere Medikament oder das Placebo erhalten) aufgeteilt („randomized trial“).

◇

Bevor wir weitermachen listen wir stichwortartig Begriffe, die wir uns bis jetzt angesehen haben.

1. Wir wissen was W-Maße und diskrete W-Räume sind. (Definition und elementare Eigenschaften.)
2. Urnenmodelle. Vielfach anwendbar und oft mit Gleichverteilung modellierbar.
3. Diskrete W-Maße können eindeutig durch Zähldichten charakterisiert wird.
4. Wir haben einige wichtige Verteilungen kennengelernt.
5. Wir haben bedingte Wahrscheinlichkeiten definiert, Satz von totaler W-keit und Satz von Bayes bewiesen und Beispiele betrachtet.

Folgende Fragen sollten alle mit dem Wissen der Vorlesung beantworten können. Wer die Fragen nicht richtig beantworten kann oder Schwierigkeiten damit hat, sollte unbedingt die entsprechenden Stellen im Skript nacharbeiten.

**Welche der folgenden Aussagen sind wahr?**

- (a) Auf  $\{0, 1\}$  ist die Bernoulli-Verteilung ein diskretes Wahrscheinlichkeitsmaß.
- (b) Auf  $\mathbb{R}$  sei ein W'maß definiert via  $P(1) = p$  und  $P(0) = 1 - p$  für ein  $p \in [0, 1]$ . Dann ist  $(\mathbb{R}, P)$  ein diskreter Wahrscheinlichkeitsraum.
- (c) Die Zähldichte der Gleichverteilung auf  $\{1, 2, \dots, n\}$  für ein  $n \in \mathbb{N}$  ist gegeben durch  $f(k) = 1/k$  für  $k = 1, \dots, n$ .
- (d) Jedes Wahrscheinlichkeitsmaß ist  $\sigma$ -additiv.
- (e) Wenn Mengen  $A, B \subseteq \Omega$  nicht disjunkt sind, d.h. wenn  $A \cap B \neq \emptyset$ , dann gilt  $P(A \cup B) < P(A) + P(B)$ .

Wir kommen nun zu einer sehr zentralen Definition in der Stochastik.

Wird  $A$  nicht von  $B$  beeinflusst, so sollte  $P(A) = P(A | B) = \frac{P(A \cap B)}{P(B)}$  gelten.

**Definition 2.6.** Sei  $(\Omega, P)$  ein diskreter Wahrscheinlichkeitsraum. Zwei Ereignisse  $A, B \subseteq \Omega$  heißen **(P-)stochastisch unabhängig**, falls

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

Ereignisse  $A_1, \dots, A_n \subseteq \Omega$  in einem diskreten Wahrscheinlichkeitsraum  $(\Omega, P)$  heißen **(P-)stochastisch unabhängig**, wenn für **jede** Indexmenge  $I \subseteq \{1, \dots, n\}$ ,  $I \neq \emptyset$ , gilt

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

*Beispiel 2.7.*

1. Wir betrachten zweimaliges Werfen eines fairen Würfels:  $\Omega = \{1, \dots, 6\}^2$  mit Gleichverteilung  $P$ .

$$A = \{1\} \times \{1, \dots, 6\} \quad (\text{beim ersten Wurf fällt 1})$$

$$B = \{1, \dots, 6\} \times \{1\} \quad (\text{beim zweiten Wurf fällt 1})$$

sollten unabhängig sein. In der Tat gilt

$$P(A \cap B) = P(\{1\} \times \{1\}) = \frac{1}{36} = \frac{6}{36} \cdot \frac{6}{36} = P(A) \cdot P(B).$$

2. Wir betrachten wieder das Skatspiel wie oben.  $C = \{\diamond B, \spadesuit B, \heartsuit B, \clubsuit B\}$ .

$$A = \{\omega \in \Omega : |\{\omega_1, \dots, \omega_{10}\} \cap C| = 1\} \quad (\text{Spieler 1 hat einen Buben})$$

$$B = \{\omega \in \Omega : |\{\omega_{11}, \dots, \omega_{20}\} \cap C| = 1\} \quad (\text{Spieler 2 hat einen Buben})$$

sollten nicht unabhängig sein:  $P(A) = P(B) = 4 \cdot 10 \cdot \frac{28!}{19!} \cdot 22!/32! \approx 0,4283$ , aber

$$P(A) \cdot P(B) \neq P(A \cap B) = 4 \cdot 10 \cdot 3 \cdot 10 \cdot \frac{28!}{(28-18)!} \cdot 12!/32! \approx 0,1835.$$

◇

*Bemerkung 2.8.*

1. Definition 2.6 stellt sicher, dass jede beliebige Auswahl  $A_i, i \in I \subseteq \{1, \dots, n\}$  aus unabhängigen Ereignissen  $A_1, \dots, A_n$  auch wieder unabhängig ist.
2. Mehr als zwei Ereignisse  $A_1, \dots, A_n$  sind im allgemeinen nicht stochastisch unabhängig, wenn nur

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

gilt. *Gegenbeispiel:*

$$\underbrace{A_1 = \emptyset}_{P(A_1)=0} \quad \text{und} \quad \underbrace{A_2 = A_3 = A}_{P(A) \in (0,1)}$$

So gilt zwar

$$P\left(\bigcap_{i=1}^n A_i\right) = P(\emptyset) = 0 = \prod_{i=1}^n P(A_i),$$

aber  $A_1, A_2, A_3$  sind nicht stochastisch unabhängig, da

$$P(A_2 \cap A_3) = P(A) \neq P(A_2) \cdot P(A_3) = (P(A))^2.$$

3. Mehr als zwei Ereignisse sind in der Regel nicht stochastisch unabhängig, wenn jeweils zwei der Ereignisse stochastisch unabhängig sind. (**Paarweise Unabhängigkeit**)

**Gegenbeispiel:** Zweimaliges Werfen eines fairen Würfels

Modell:  $\Omega = \{1, \dots, 6\}^2$ ,  $P$  Gleichverteilung

Betrachte die Ereignisse

$$A_1 = \{1, 3, 5\} \times \{1, \dots, 6\} \quad (\text{erste Augenzahl ist ungerade}),$$

$$A_2 = \{1, \dots, 6\} \times \{1, 3, 5\} \quad (\text{zweite Augenzahl ist ungerade}),$$

$$A_3 = \{(\omega_1, \omega_2) \in \Omega \mid \omega_1 + \omega_2 \text{ ungerade}\} \quad (\text{Augensumme ungerade}) \\ = (\{1, 3, 5\} \times \{2, 4, 6\}) \cup (\{2, 4, 6\} \times \{1, 3, 5\}).$$

Dann sind jeweils zwei dieser Ereignisse unabhängig, z.B.

$$P(A_2 \cap A_3) = \frac{|\{2, 4, 6\} \times \{1, 3, 5\}|}{|\Omega|} = \frac{9}{36} = \frac{1}{2} \cdot \frac{1}{2} = P(A_2) \cdot P(A_3)$$

Aber:  $A_1, A_2, A_3$  sind nicht stochastisch unabhängig, denn

$$A_1 \cap A_2 \cap A_3 = \emptyset \quad \Rightarrow \quad P(A_1 \cap A_2 \cap A_3) = 0 \neq \frac{1}{8} = P(A_1) \cdot P(A_2) \cdot P(A_3).$$

### 3 Zufallsvariablen und ihre Verteilungen

Für Stochastiker sind oft nicht die  $W$ -keiten auf einem bestimmten  $W$ -Raum  $(\Omega, P)$  von Interesse, sondern die  $W$ -keiten, mit der eine *Zufallsgröße*  $X$  Werte in einer bestimmten Menge  $S$  annimmt. In diesem Kapitel werden wir diesen wichtigen Begriff der Zufallsvariablen kennen lernen. Diese sind nichts anderes als die mathematische Modellierung/Beschreibung von zufälligen Größen sind.

**Frage:** Wie ergänzt man  $S$  zu einem  $W$ -Raum?

**Definition 3.1.** Ist  $(\Omega, P)$  ein diskreter Wahrscheinlichkeitsraum und  $S \neq \emptyset$  eine beliebige Menge, so wird die Abbildung  $X: \Omega \rightarrow S$  auch  **$S$ -wertige Zufallsvariable** genannt.

**Satz 3.2.** Ist  $X: \Omega \rightarrow S$  eine Zufallsvariable auf einem diskreten  $W$ -raum  $(\Omega, P)$ , dann wird durch

$$P^X(B) := P(X^{-1}(B)), \quad \forall B \subseteq S,$$

ein Wahrscheinlichkeitsmaß  $P^X$  auf  $S$  definiert, welches **Verteilung von  $X$**  genannt wird.  $(S, P^X)$  ist ein diskreter Wahrscheinlichkeitsraum.

*Notation* für Urbilder:

$$\begin{aligned} \{X \in B\} &:= \{\omega \in \Omega \mid X(\omega) \in B\} = X^{-1}(B) \\ \{X = x\} &:= \{\omega \in \Omega \mid X(\omega) = x\} = X^{-1}(\{x\}) \\ \{X > x\} &:= \{\omega \in \Omega \mid X(\omega) > x\} = X^{-1}((x, \infty)) \quad (\text{im Fall } S = \mathbb{R}) \end{aligned}$$

etc. Zudem schreiben wir bspw.  $P(X \in B) := P(\{X \in B\})$ .

*Beweis.* Wir weisen die Eigenschaften eines diskreten Wahrscheinlichkeitsmaßes nach. Aus der Normierung und der  $\sigma$ -Additivität von  $P$  folgt

$$P^X(S) = P(X^{-1}(S)) = P(X \in S) = P(\Omega) = 1$$

sowie für beliebige disjunkte Mengen  $A_n \subseteq S, n \in \mathbb{N}$ :

$$\begin{aligned} P^X\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= P\left(X \in \bigcup_{n \in \mathbb{N}} A_n\right) = P\left(\bigcup_{n \in \mathbb{N}} \{X \in A_n\}\right) \\ &= \sum_{n \in \mathbb{N}} P(\{X \in A_n\}) = \sum_{n \in \mathbb{N}} P^X(A_n). \end{aligned}$$

Ist schließlich  $\Omega_0 \subseteq \Omega$  eine diskrete Menge mit  $P(\Omega_0) = 1$ , dann definieren wir  $S_0 := \{s \in S \mid \exists \omega \in \Omega_0 : X(\omega) = s\}$ .  $S_0$  ist höchstens abzählbar und es gilt

$$P^X(S_0) = P(X \in S_0) = P(X^{-1}(S_0)) \geq P(\Omega_0) = 1.$$

Insgesamt haben wir damit gezeigt, dass  $(S, P^X)$  ein diskreter Wahrscheinlichkeitsraum ist.  $\square$

*Beispiel 3.3.*

Der zugrundeliegende  $W$ -raum ist  $\Omega = \{1, \dots, 6\}^2$  versehen mit der Gleichverteilung  $P$ . Die Augensummen wird beschrieben durch die Zufallsvariable

$$X: \Omega \rightarrow \{2, \dots, 12\} =: S, \quad X(\omega_1, \omega_2) := \omega_1 + \omega_2.$$

Die Verteilung  $P^X$  von  $X$  hat die **Zähldichte**  $f_X$  für alle  $k \in S$  gegeben durch

$$\begin{aligned}
 f_X(k) &= P^X(\{k\}) = P(X = k) \\
 &= P(\{(\omega_1, \omega_2) \in \Omega \mid X(\omega_1, \omega_2) = \omega_1 + \omega_2 = k\}) \\
 &= P(\{(\omega_1, k - \omega_1) \mid 1 \leq \omega_1 \leq 6, 1 \leq k - \omega_1 \leq 6\}) \\
 &= P(\{(\omega_1, k - \omega_1) \mid \max(1, k - 6) \leq \omega_1 \leq \min(6, k - 1)\}) \\
 &= \begin{cases} \frac{k-1}{36} & \text{falls } 2 \leq k \leq 7 \\ \frac{13-k}{36} & \text{falls } 7 < k \leq 12 \end{cases} \\
 &= \frac{6 - |k - 7|}{36}. \quad \diamond
 \end{aligned}$$

### 3.1 Unabhängige Zufallsvariablen

Im vorangegangenen Beispiel gehen wir ganz intuitiv davon aus, dass die beiden Würfelwürfe *unabhängig* von einander sind. Diese Intuition wollen wir nun formalisieren.

Zufallsvariablen sollen als unabhängig gelten, wenn beliebige Ereignisse, die damit ausgedrückt werden können, stochastisch unabhängig sind.

**Definition 3.4.** Sei  $(\Omega, P)$  ein diskreter Wahrscheinlichkeitsraum und  $S_i, i \in \{1, \dots, n\}$ , nicht-leere Mengen. Zufallsvariablen

$$X_i: \Omega \rightarrow S_i, \quad i \in \{1, \dots, n\},$$

heißen **( $P$ -)stochastisch unabhängig**, wenn für beliebige  $B_i \subseteq S_i, i \in \{1, \dots, n\}$ , die Ereignisse

$$\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$$

stochastisch unabhängig sind.

Die Zufallsvariablen  $X_i: \Omega \rightarrow S_i$  besitzen die Verteilungen  $P^{X_i}$  für alle  $i = 1, \dots, n$ .

Auch der Vektor  $(X_1, \dots, X_n): \Omega \rightarrow S_1 \times \dots \times S_n$  ist eine Zufallsvariable mit einer Verteilung  $P^{(X_1, \dots, X_n)}$  auf  $S_1 \times \dots \times S_n$ .

**Satz 3.5.** In der Situation von Definition 3.4 sind äquivalent:

- (i)  $X_1, \dots, X_n$  sind stochastisch unabhängig.
- (ii)  $\forall B_i \subseteq S_i (1 \leq i \leq n): P\{X_i \in B_i \forall 1 \leq i \leq n\} = \prod_{i=1}^n P\{X_i \in B_i\}$
- (iii) Bezeichnen  $f_{X_i}$  die Zähldichten von  $P^{X_i}$  auf  $S_i$ , dann hat die Zähldichte  $f_{(X_1, \dots, X_n)}$  von  $P^{(X_1, \dots, X_n)}$  die Form

$$f_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \prod_{i=1}^n f_{X_i}(t_i) \quad \forall t_i \in S_i, 1 \leq i \leq n.$$

► Die gemeinsame Verteilung  $P^{(X_1, \dots, X_n)}$  unabhängiger Zufallsvariablen  $X_1, \dots, X_n$  besitzt also *Produktgestalt*.

*Beweis.* Wir verwenden einen Ringschluss  $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i)$ .

$(i) \Rightarrow (ii)$  folgt direkt aus den Definition von stochastischer Unabhängigkeit von Ereignissen und Zufallsvariablen.

(ii)  $\Rightarrow$  (iii) Für  $t_i \in S_i, i = 1, \dots, n$  wählen  $B_i = \{t_i\}$  und erhalten aus (ii):

$$\begin{aligned} f_{(X_1, \dots, X_n)}(t_1, \dots, t_n) &= P^{(X_1, \dots, X_n)}(\{t_1, \dots, t_n\}) \\ &= P(X_i = t_i \forall 1 \leq i \leq n) \\ &= \prod_{i=1}^n P(X_i = t_i) \\ &= \prod_{i=1}^n P^{X_i}(\{t_i\}) = \prod_{i=1}^n f_{X_i}(t_i). \end{aligned}$$

(iii)  $\Rightarrow$  (i) Für  $I \subseteq \{1, \dots, n\}$  und  $B_i \subseteq S_i, i \in I$  ist

$$P(X_i \in B_i \forall i \in I) = \prod_{i \in I} P(X_i \in B_i)$$

zu zeigen. Für  $i \in \{1, \dots, n\} \setminus I$  setzen wir  $B_i = S_i$  und erhalten aus (iii)

$$\begin{aligned} P(X_i \in B_i \forall i \in I) &= P(X_i \in B_i \forall 1 \leq i \leq n) \\ &= \sum_{(t_1, \dots, t_n) \in B_1 \times \dots \times B_n} f_{(X_1, \dots, X_n)}(t_1, \dots, t_n) \\ &= \sum_{t_1 \in B_1} \dots \sum_{t_n \in B_n} \prod_{i=1}^n f_{X_i}(t_i) \\ &= \prod_{i=1}^n \left( \sum_{t_i \in B_i} f_{X_i}(t_i) \right) \\ &= \prod_{i=1}^n P(X_i \in B_i) = \prod_{i \in I} P(X_i \in B_i) \quad \square \end{aligned}$$

Für die letzte Gleichung benutzen wir, dass  $P(X_i \in B_i) = 1$  für  $\{1, \dots, n\} \setminus I$ .

Gemeinsame Verteilung von ZV  $X, Y$  sei gegeben durch die Werte  $P(X = i, Y = j), i \in S = \{0, 1, 2\}, j \in T = \{-1, 0, 1\}$ .

$P(X = i, Y = j)$	$j = -1$	$j = 0$	$j = 1$	$P(X = i)$
$i = 0$	0.1	0.2	0.1	0.4
$i = 1$	0.1	0.1	0.1	0.3
$i = 2$	0.1	0.1	0.1	0.3
$P(Y = j)$	0.3	0.4	0.3	

Insbesondere können wir die Randverteilungen und ihre Zähldichten ablesen. Zum Beispiel gilt für  $A_1 \subseteq S$

$$P(X \in A_1) = \sum_{i \in A_1} P(X = i) = \sum_{i \in A_1} P(X = i, Y \in T) = \sum_{i \in A_1} \sum_{j \in T} P(X = i, Y = j).$$

Die Zufallsvariablen sind nicht unabhängig. Zum Beispiel gilt

$$P(X = 0, Y = -1) = 0.1 \neq 0.4 \cdot 0.3 = P(X = 0) \cdot P(Y = -1).$$

Wie müsste die Tabelle aussehen, damit bei den gleichen Randverteilungen die Zufallsvariablen unabhängig wären?

$P(X = i, Y = j)$	$j = -1$	$j = 0$	$j = 1$	$P(X = i)$
$i = 0$	0.12	0.16	0.12	0.4
$i = 1$	0.09	0.12	0.09	0.3
$i = 2$	0.09	0.12	0.09	0.3
$P(Y = j)$	0.3	0.4	0.3	

**Beispiele:** Sind folgende ZV  $X$  und  $Y$  unabhängig?

1. Seien  $X$  und  $Y$  Zufallsvariablen mit Wertebereich  $\{0, 1, 2\}$  bzw.  $\{1, 2, 3\}$ . und gemeinsamer Zähl-dichte  $(x, y) \mapsto P(X = x, Y = y)$  von  $(X, Y)$  gegeben durch:

$x \setminus y$	1	2	3
0	1/10	0	3/10
1	2/10	0	1/10
2	1/10	2/10	0

2. Seien  $X$  und  $Y$  Zufallsvariablen mit Wertebereich  $\{-2, 0, 1\}$  bzw.  $\{1, 3\}$  und gemeinsamer Zähl-dichte

$x \setminus y$	1	3
-2	2/20	1/20
0	4/20	4/20
1	6/20	3/20

**Bemerkung:** Ist  $f_{(X,Y)}$  die gemeinsame Zähl-dichte auf  $S \times T$ , dann sind die Zähl-dichten  $f_X$  und  $f_Y$  (genannt **Rand-Zähl-dichten**) gegeben durch

$$f_X(x) = \sum_{y \in T} f_{(X,Y)}(x, y) \quad \text{und} \quad f_Y(y) = \sum_{x \in S} f_{(X,Y)}(x, y)$$

*Beispiel 3.6* ( $n$  unabhängige Bernoulliexperimente).

Ein Bernoulli-Zufallsexperiment wird  $n$ -mal unabhängig durchgeführt, wobei die Erfolgswahrscheinlichkeit jeweils  $p \in [0, 1]$  sei.

Wir betrachten den Grundraum  $\Omega = \{0, 1\}^n$ , wobei für  $\omega = (\omega_1, \dots, \omega_n) \in \Omega$   $\omega_i = 1$  gelte, falls das  $i$ -te Experiment erfolgreich war und sonst  $\omega_i = 0$ .

$\Omega$  sei mit einem W'maß  $P$  versehen.

Nun betrachten wir die Zufallsvariablen

$$X_i : \Omega \rightarrow \{0, 1\} \quad \text{mit} \quad X_i(\omega) = \omega_i \quad \text{also}$$

$$X_i = \begin{cases} 0, & \text{falls } i\text{-tes Experiment Misserfolg,} \\ 1, & \text{falls } i\text{-tes Experiment Erfolg.} \end{cases}$$

Da  $X_i$  Bernoulli-verteilt mit Erfolgsw'keit  $p$  sein soll, muss gelten

$$f_{X_i}(1) = P^{X_i}(\{1\}) = p, \quad f_{X_i}(0) = P^{X_i}(\{0\}) = 1 - p$$

$$\iff f_{X_i}(\omega_i) = p^{\omega_i} (1 - p)^{1 - \omega_i}, \quad \omega_i \in \{0, 1\}.$$

Wegen  $X_i(\omega) = \omega_i, i = 1, \dots, n$  gilt für den Zufallsvektor

$$(X_1, \dots, X_n)(\omega) = \omega \quad \text{für alle} \quad \omega \in \Omega.$$

►  $P = P^{(X_1, \dots, X_n)}$ .

Wegen der geforderten Unabhängigkeit von  $X_i$ , ist die Zähldichte von  $(X_1, \dots, X_n)$  als **Produkt-dichte** gegeben:

$$\begin{aligned} f(\omega_1, \dots, \omega_n) &= f_{(X_1, \dots, X_n)}(\omega_1, \dots, \omega_n) = \prod_{i=1}^n f_{X_i}(\omega_i) \\ &= \prod_{i=1}^n p^{\omega_i} (1-p)^{1-\omega_i} \\ &= p^{\sum_{i=1}^n \omega_i} \cdot (1-p)^{\sum_{i=1}^n (1-\omega_i)}. \end{aligned}$$

Wir betrachten nun die Zufallsvariable  $Y := \sum_{i=1}^n X_i$ , also die Anzahl der Erfolge in  $n$  Experimenten.  $P^Y$  hat die Zähldichte

$$\begin{aligned} f_Y(k) &= P(\{Y = k\}) \\ &= P\left(\left\{(\omega_1, \dots, \omega_n) \in \Omega \mid Y(\omega_1, \dots, \omega_n) = \sum_{i=1}^n X_i(\omega) = \sum_{i=1}^n \omega_i = k\right\}\right) \\ &= \sum_{\substack{(\omega_1, \dots, \omega_n) \in \Omega \\ \sum_{i=1}^n \omega_i = k}} f(\omega) \\ &= \binom{n}{k} p^k (1-p)^{n-k}, \end{aligned}$$

denn jeder Summand ist gleich  $p^k (1-p)^{n-k}$  und es gibt  $\binom{n}{k}$  Möglichkeiten, die  $k$  der  $n$  Stellen auszuwählen, an denen  $\omega_i$  den Wert 1 annimmt.

► Die Zähldichte  $f_Y$  der Verteilung von  $Y$  ist gegeben durch

$$f_Y(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{für alle } k \in \{0, \dots, n\}.$$

►  $Y$  ist binomialverteilt:  $P^Y = \mathcal{B}_{(n,p)}$

*Fazit:* Sind  $X_1, \dots, X_n$  stochastisch unabhängig mit  $P^{X_i} = \mathcal{B}_{(1,p)}$ , dann gilt für deren Summe  $Y = \sum_{i=1}^n X_i$  gilt

$$P^Y = \mathcal{B}_{(n,p)}.$$

◇

*Beispiel 3.7* (Capture-Recapture-Verfahren).

*Ziel:* Die Anzahl  $N$  von Fischen in einem See soll geschätzt werden.

*Verfahren:*

1. Fange  $M$  Fische, markiere sie (mit Nummern) und lasse sie wieder frei.
2. Fange wieder  $n$  Fische, darunter seien  $m$  markierte.

*Annahme:* Die Fangwahrscheinlichkeit im 2. Schritt sei für markierte und unmarkierte Fische gleich.

*Modell:*

$n$ -maliges Ziehen ohne Zurücklegen aus  $N$  Fischen und ohne Berücksichtigung der Reihenfolge. Der  $W$ -raum ist damit gegeben durch

$$\Omega = \{(\omega_1, \dots, \omega_n) \in \{1, \dots, N\}^n \mid \omega_1 < \omega_2 < \dots < \omega_n\}$$

versehen mit der Gleichverteilung  $P$ .

Markierte Fische entsprechen den Nummern  $1, \dots, M$ . Die Anzahl der gefangenen markierten Fische entspricht der Zufallsvariable

$$X(\omega) = \sum_{i=1}^n \mathbb{1}_{\{1, \dots, M\}}(\omega_i)$$

mit der *Indikatorfunktion*

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \text{falls } \omega \in A \\ 0, & \text{falls } \omega \notin A \end{cases}, \quad A \subseteq \Omega.$$

Die möglichen Werte von  $X(\omega)$  sind also gegeben durch die Bedingungen

$$0 \leq X(\omega) \leq \min(n, M), \quad n - X(\omega) \leq N - M,$$

womit folgt, dass der Wertebereich von  $X$

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

ist.

Zu berechnen ist nun:

$$P^X(\{m\}) = P(X = m) = \frac{|\{\omega \in \Omega \mid X(\omega) = m\}|}{|\Omega|},$$

wobei  $|\Omega| = \binom{N}{n}$  die Anzahl der Möglichkeiten ist,  $n$  Fische aus  $N$  Fischen auszuwählen.

Es gibt genau  $\binom{M}{m}$  Möglichkeiten,  $m$  Fische aus  $M$  markierten Fischen zu ziehen.

Es gibt genau  $\binom{N-M}{n-m}$  Möglichkeiten,  $n-m$  Fische aus allen  $N-M$  nicht markierten Fischen zu ziehen.

► Insgesamt gibt es  $\binom{M}{m} \cdot \binom{N-M}{n-m}$  Möglichkeiten,  $n$  Fische aus  $N$  zu ziehen, so dass genau  $m$  davon markiert sind.

$$\rightsquigarrow P^X(\{m\}) = P(X = m) = \frac{\binom{M}{m} \cdot \binom{N-M}{n-m}}{\binom{N}{n}} \mathbb{1}_S(m) \quad \forall m \in \mathbb{N}_0 \quad (*)$$

*Definition 3.8.* Das Wahrscheinlichkeitsmaß  $P = \mathcal{H}_{(N, M, n)}$  auf  $\mathbb{N}_0$  mit der Zähldichte Die rechte Seite gibt die Zähldichte (\*) heißt **hypergeometrischen Verteilung**.

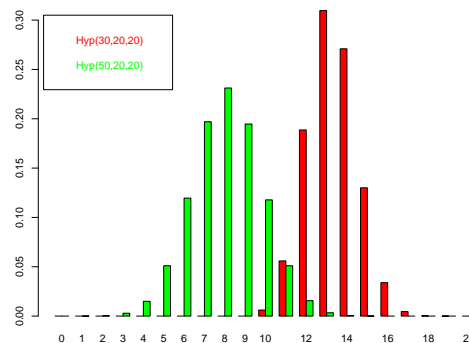


Abb.: Zähldichte der Hypergeometrischen Verteilung mit  $M = n = 20$  und  $N \in \{30, 50\}$ .  $\diamond$

Die **hypergeometrische Verteilung**  $\mathcal{H}_{(N,M,n)}$  beschreibt die Anzahl der markierten Gegenstände beim  $n$ -maligem **Ziehen ohne Zurücklegen** aus  $N$  Gegenständen, von denen  $M$  markiert sind. Zieht man **mit Zurücklegen**, so ist die Wahrscheinlichkeit, einen markierten Gegenstand zu ziehen, bei jeder Ziehung  $M/N$ . Die Anzahl der gezogenen markierten Gegenstände hat daher gerade die **Binominalverteilung**  $\mathcal{B}_{(n,M/N)}$ .

Falls  $n \ll N$ , dann ist Ziehen mit oder ohne Zurücklegen fast identisch und daher

$$\mathcal{H}_{(N,M,n)}\{m\} \approx \mathcal{B}_{(n, \frac{M}{N})}\{m\} \quad \forall 0 \leq m \leq n$$

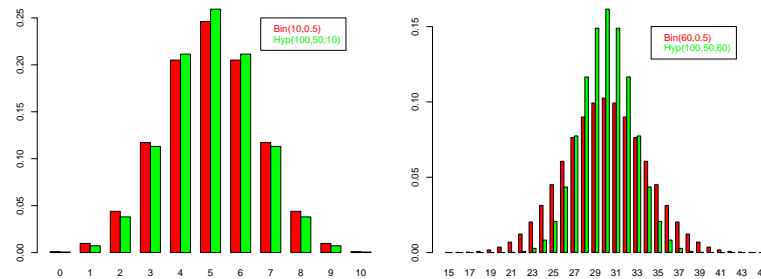


Abb.: Vergleich der Binomal- und hypergeometrischen Verteilung für  $\frac{n}{N} = 0.1$  (links) und  $\frac{n}{N} = 0.6$  (rechts)

*Beispiel 3.9.*

In Beispiel 3.7 ist uns ein erstes **statistisches Modell** begegnet: Auf  $\mathbb{N}_0$  betrachten wir die Familie von Verteilungen  $(\mathcal{H}_{(N,M,n)})_{N \in \mathbb{N}}$ , wobei  $M, n$  fest und bekannt sind, aber  $N \in \mathbb{N}$  ein **unbekannter Parameter** ist.

**Ziel:** Konstruktion eines Schätzers für den wahren Parameter  $N$ , unter dem Beobachtungen erzeugt wurden.

**Lösungsidee:** Wir werden sehen, dass die erwartete Anzahl markierter Fische  $n \cdot \frac{M}{N}$  ist, wobei  $N$  unbekannt ist.

Gehen wir nun davon aus, dass die tatsächliche Anzahl der beobachteten, markierten Fische  $X(\omega)$  in etwa gleich der erwarteten Anzahl ist, erhalten wir den **Schätzer**

$$X(\omega) \approx n \cdot \frac{M}{N} \quad \rightsquigarrow \quad \hat{N} := \left\lceil n \cdot \frac{M}{X(\omega)} \right\rceil,$$

wobei  $\lceil x \rceil$  die Gauß-Klammer von  $x$  bezeichnet, also die größte ganze Zahl kleiner oder gleich  $x$ .

◇

*Beispiel 3.10.*

**Modell:**

$N$  = Populationsgröße, (*unbekannt aber nicht zufällig*),

$M = 1000$ , markierter Fische,

$n = 1000$ , Stichprobengröße, Anzahl der Fische beim zweiten Fang,

$k = 100$ , Anzahl markierter Fische beim zweiten Fang (*Beobachtung*).

Im Beispiel erhalten wir

$$\hat{N} = \frac{1000 \cdot 1000}{100} = 10000.$$

◇

Es sei  $\lambda > 0$ . Eine  $\mathbb{N}_0$ -wertige Zufallsvariable  $X$  ist **Poissonverteilt mit Parameter  $\lambda$** , kurz  $X \sim \text{Pois}(\lambda)$ , wenn

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Nach Hausaufgabe handelt es sich tatsächlich um eine Zähl-dichte

**Anwendungen:**

- Radioaktiver Zerfall/ Anzahl des Ausschläge des Geigerzählers pro Zeiteinheit (viele Atomkerne, kleine W-keit für Zerfall).
- Anzahl der Anfragen bei einem Server pro Zeiteinheit (viele „Kunden“, kleine W-keit für Anfrage).
- Anzahl (seltener) Unfälle in einer Woche in einer Großstadt,
- Anzahl (seltener) Mutationen an einem DNA-Strang.

Die Anwendungen sind durch den Poisson'schen Grenzwertsatz begründet.

Für eine große Anzahl an Experimenten  $n$  und eine kleine Erfolgsw'keit  $p$  kann  $\mathcal{B}_{(n,p)}$  durch eine **strukturell einfachere** Verteilung approximiert werden:

**Satz und Definition 3.11** (Poisson'scher Grenzwertsatz). *Ist  $p_n \in (0, 1)$ ,  $n \in \mathbb{N}$ , so dass  $\lim_{n \rightarrow \infty} np_n = \lambda > 0$ , dann gilt*

$$\lim_{n \rightarrow \infty} \mathcal{B}_{(n,p_n)}(\{k\}) = \lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} = e^{-\lambda} \cdot \frac{\lambda^k}{k!} =: f_\lambda(k) \quad \forall k \in \mathbb{N}_0.$$

Dieses Resultat ist auch unter Namen „Gesetz der kleinen Zahlen“ bekannt.

*Beweis.* Für festes  $k$  gilt

$$\begin{aligned} \binom{n}{k} &= \frac{n^k n(n-1) \cdots (n-k+1)}{k!} \\ &= \frac{n^k}{k!} \underbrace{1 \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)}_{\rightarrow 1 \text{ für } n \rightarrow \infty}. \end{aligned}$$

Daraus folgt für  $n \rightarrow \infty$

$$\begin{aligned} \frac{\mathcal{B}_{(n,p_n)}(\{k\})}{e^{-\lambda} \frac{\lambda^k}{k!}} &= \frac{\binom{n}{k} p_n^k (1 - p_n)^{n-k}}{e^{-\lambda} \frac{\lambda^k}{k!}} \\ &= \underbrace{\frac{k!}{n^k} \binom{n}{k}}_{\rightarrow 1} \cdot \underbrace{\left(\frac{np_n}{\lambda}\right)^k}_{\rightarrow 1} \cdot \underbrace{(1 - p_n)^{-k}}_{\rightarrow 1} \cdot \underbrace{\left(1 - \frac{np_n}{n}\right)^n}_{\rightarrow e^{-\lambda}} e^\lambda \rightarrow 1. \quad \square \end{aligned}$$

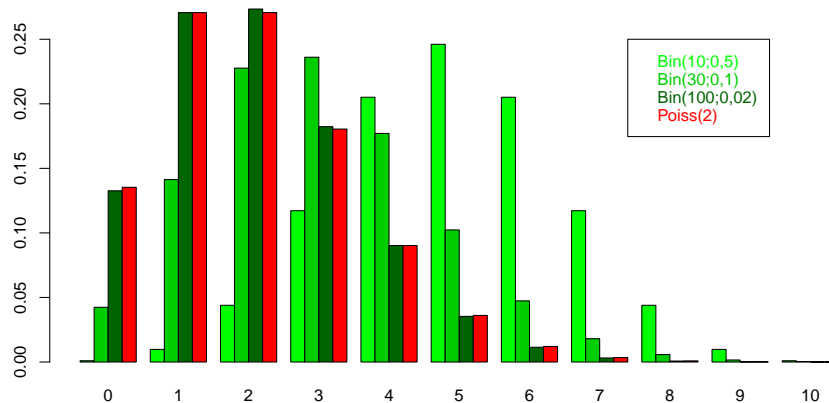


Abb.: Illustration des Poisson'schen Grenzwertsatzes: Zähldichten der Binomialverteilung mit Parametern  $n \cdot p \rightarrow 2$  und Poissonverteilung mit Intensität 2.

*Anwendung:* Die Anzahl der Zerfälle pro Minute in einer radioaktiven Probe ist (näherungsweise) Poisson-verteilt. Dies lässt sich mit dem Poisson'schen Grenzwertsatz erklären, wenn sich die Zerfälle einzelner Atome nicht gegenseitig beeinflussen.

*Beispiel 3.12 (Überbuchung).*

Für einen Flug stehen 200 Plätze zur Verfügung. Da erfahrungsgemäß 3% aller Ticketkäufer nicht zum Flug erscheinen, verkauft die Fluggesellschaft  $n = 203$  Tickets.

*Frage:* Wie hoch ist die W'keit, dass der Flug überbucht ist, wenn man vereinfachend annimmt, dass das Nichterscheinen bei allen Ticketkäufern unabhängig voneinander erfolgt?

Sei  $X$  die Anzahl der Ticketkäufer, die nicht zum Flug erscheint. Dann ist  $X$  gerade  $\mathcal{B}_{(n,p)}$ -verteilt mit  $p = 0,03$ .

Die W'keit einer Überbuchung ist daher gerade

$$P(X \leq 2) = \mathcal{B}_{(n,p)}(\{0, 1, 2\}) \approx 0,0555.$$

Verwendet man die Poisson-Approximation aus Satz 3.7, so erhält man

$$P(X \leq 2) \approx \mathcal{P}_{np}(\{0, 1, 2\}) \approx 0,0581.$$

Der relative Approximationsfehler beträgt knapp 5%. ◇

Oft treten Summen **unabhängiger** Zufallsvariablen auf, z.B. als Gesamtmessfehler, der sich aus unabhängigen Einzelfehlerquellen zusammensetzt.

Die Frage, die sich dann stellt ist wie man mit Verteilungen von Summen rechnen kann.

Seien  $X$  und  $Y$  Augenzahlen zweier Würfel und  $Z = X + Y$ . D.h. wir haben  $\Omega = \{1, \dots, 6\}^2$  mit Gleichverteilung  $P$  für die beiden Würfelwürfe und der Wertebereiche von  $X$ ,  $Y$  ist  $\{1, \dots, 6\}$  und der von  $Z$  ist  $\{2, \dots, 12\}$ .

Wir können die Zufallsvariablen als  $\mathbb{Z}$ -wertig auffassen indem wir ihre Verteilung außerhalb der Wertebereiche auf 0 setzen:

$$P(Y = -1) = P(X = 10) = P(\emptyset) = 0$$

*Frage:* Wie bestimmt man die Verteilung von  $Z$ ?

*Möglichkeit 1:* Abzählen einzelner passender Kombinationen auf

$$P(Z = 8) = P(X + Y = 8) = P(\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}) = \frac{5}{36}$$

Möglichkeit 2: Systematisch berechnen mit Schreibweise  $f_X(i) = P(X = i)$ , etc.

$$\begin{aligned}
 P(Z = 8) &= P(X + Y = 8) = \sum_{i=1}^6 P(X + Y = 8, X = i) \\
 &= \sum_{i=1}^6 P(Y = 8 - i, X = i) \stackrel{\text{Unabh}}{=} \sum_{i=1}^6 P(Y = 8 - i)P(X = i) \\
 &= f_Y(7)f_X(1) + f_Y(6)f_X(2) + \dots + f_Y(2)f_X(6) \\
 &= f_Y(6)f_X(2) + f_Y(3)f_X(5) + f_Y(4)f_X(4) + f_Y(5)f_X(3) + f_Y(2)f_X(6) \\
 &= \frac{5}{36}
 \end{aligned}$$

Wobei wir  $f_Y(7) = P(Y = 7) = 0$  benutzt haben.

**Satz und Definition 3.13.** Sind  $X, Y$   $\mathbb{R}$ -wertige Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum mit Zähldichten  $f_X$  von  $P^X$  und  $f_Y$  von  $P^Y$ , dann heißt

$$(f_X * f_Y)(z) = \sum_{x \in \mathbb{R}: f_X(x) > 0} f_X(x) \cdot f_Y(z - x), \quad \forall z \in \mathbb{R},$$

die **Faltung** von  $f_X$  und  $f_Y$ . Hierbei ist  $f_X * f_Y$  wieder eine Zähldichte mit dem Träger  $\Omega_T := \{z \in \mathbb{R} | \exists x, y \in \mathbb{R} : z = x + y, f_X(x) > 0, f_Y(y) > 0\}$  und die zugehörige diskreten Verteilung  $P^X * P^Y$  nennen wir **Faltung** von  $P^X$  und  $P^Y$ .

*Beweis.* Da die Träger von  $P^X$  und  $P^Y$  abzählbar sind, muss auch  $\Omega_T$  abzählbar sein. Weiterhin ist  $(f_X * f_Y)(z) \geq 0$  für alle  $z \in \mathbb{R}$  und  $(f_X * f_Y)(z) > 0$  kann nur gelten, wenn es mindestens eine Zerlegung  $z = x + y$  mit  $f_X(x) > 0, f_Y(y) > 0$  gibt. Um nachzuweisen, dass  $f_X * f_Y$  eine Zähldichte ist, bleibt zu zeigen

$$\begin{aligned}
 \sum_{z \in \mathbb{R}} (f_X * f_Y)(z) &= \sum_{z \in \mathbb{R}} \sum_{x \in \mathbb{R}} f_X(x) \cdot f_Y(z - x) \\
 &\stackrel{y=z-x}{=} \sum_{x \in \mathbb{R}} \sum_{y \in \mathbb{R}} f_X(x) \cdot f_Y(y) \\
 &= \left( \sum_{x \in \mathbb{R}} f_X(x) \right) \left( \sum_{y \in \mathbb{R}} f_Y(y) \right).
 \end{aligned}$$

Man beachte, dass in allen Summen höchstens abzählbar viele Summanden von null verschieden sind.  $\square$

**Satz 3.14.** Sind  $X, Y$  unabhängige  $\mathbb{R}$ -wertige Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum, so gilt

$$P^X * P^Y = P^{X+Y}.$$

*Beweis.*

Für jedes  $z \in \mathbb{R}$  gilt nach der Definition der Faltung und mit Hilfe der Abzählbarkeit des Trägers von  $P^X$  sowie der  $\sigma$ -Additivität von  $P$ :

$$\begin{aligned}
 (f_X * f_Y)(z) &= P^{X+Y}(\{z\}) = P(X + Y = z) \\
 &= P\left(\{X + Y = z\} \cap \left(\bigcup_{x \in \mathbb{R}: f_X(x) > 0} \{X = x\}\right)\right) \\
 &= P\left(\bigcup_{x \in \mathbb{R}: f_X(x) > 0} (\{X = x\} \cap \{X + Y = z\})\right) \\
 &= \sum_{x \in \mathbb{R}: f_X(x) > 0} P(X = x, Y = z - x).
 \end{aligned}$$

Nun verwenden wir die Unabhängigkeit von  $X$  und  $Y$ , um für alle  $n \in \mathbb{Z}$  zu folgern:

$$(f_X * f_Y)(z) = \sum_{x \in \mathbb{R}} P(X = x)P(Y = z - x) = \sum_{x \in \mathbb{R}} f_X(x) \cdot f_Y(z - x). \quad \square$$

*Beispiel 3.15.* Wir berechnen die Faltung zweier Poisson-Verteilungen.

Seien  $\lambda_1, \lambda_2 > 0$ . Dann hat  $\mathcal{P}_{\lambda_1} * \mathcal{P}_{\lambda_2}$  die Zähldichte

$$\begin{aligned} (f_{\lambda_1} * f_{\lambda_2})(n) &= \sum_{k=0}^{\infty} f_{\lambda_1}(k) \cdot f_{\lambda_2}(n-k) \\ &= \sum_{k=0}^n e^{-\lambda_1} \cdot \frac{\lambda_1^k}{k!} \cdot e^{-\lambda_2} \cdot \frac{\lambda_2^{n-k}}{(n-k)!} \\ &= e^{-(\lambda_1+\lambda_2)} \cdot \frac{1}{n!} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \cdot \lambda_1^k \cdot \lambda_2^{n-k} \\ &= e^{-(\lambda_1+\lambda_2)} \cdot \frac{1}{n!} \cdot (\lambda_1 + \lambda_2)^n \\ &= f_{\lambda_1+\lambda_2}(n) \quad \forall n \in \mathbb{N}_0. \end{aligned}$$

Da diskrete Wahrscheinlichkeitsmaße durch ihre Zähldichten eindeutig bestimmt sind, gilt also

$$\mathcal{P}_{\lambda_1} * \mathcal{P}_{\lambda_2} = \mathcal{P}_{\lambda_1+\lambda_2}. \quad \diamond$$

*Beispiel 3.16 (Quicksort).*

Die Zahlen  $x_1, \dots, x_n$ , die alle als verschieden angenommen werden, sollen sortiert werden. Der Algorithmus Quicksort erledigt die Aufgabe wie folgt:

1. Wähle zufällig gleichverteilt ein  $x_j$  aus.
2. Ordne Zahlen  $x_i < x_j$  links von  $x_j \rightsquigarrow X_l$  Vektor von Zahlen  $< x_j$ .
3. Ordne Zahlen  $x_i > x_j$  rechts von  $x_j \rightsquigarrow X_r$  Vektor von Zahlen  $> x_j$ .
4. Verfahre mit  $X_l$  und  $X_r$  getrennt ebenso, usw. bis nur noch Vektoren der Länge 1 übrig sind.

So etwa bei 3 7 2 6 13 1

1. zufällige Wahl 7:

$$\underbrace{3 \ 2 \ 6 \ 1}_{X_l} \mid 7 \mid \underbrace{13}_{X_r} \rightsquigarrow 5 \text{ Vergleiche}$$

2. zufällige Wahl 3 (bei  $X_l$ ):

$$\underbrace{2 \ 1}_{X'_l} \mid 3 \mid 6 \mid 7 \mid 13 \rightsquigarrow 3 \text{ Vergleiche}$$

3. zufällige Wahl 1 (bei  $X'_l$ ):

$$1 \mid 2 \mid 3 \mid 6 \mid 7 \mid 13 \rightsquigarrow \frac{1 \text{ Vergleich}}{9 \text{ Vergleiche}}$$

Wir wollen nun eine *Laufzeitanalyse* des Algorithmus durchführen.

Zunächst der worst case:

Die gewählte Zahl ist jeweils kleinste oder größte.

► Die Anzahl der Vergleiche ist dann:

$$(n-1) + (n-2) + \dots + 1 = \frac{n \cdot (n-1)}{2}.$$

► Die größtmögliche Anzahl an nötigen Vergleichen ist damit  $n(n-1)/2$ .

Nun zum best case:

Es sei speziell  $n = 2^k - 1$  und die ausgewählte Zahl jeweils die mittlere.

- 1.)  $2^k - 2$  Vergleiche  $\rightsquigarrow$  2 Blöcke mit Länge  $2^{k-1} - 1$
- 2.)  $2 \cdot (2^{k-1} - 2)$  Vergleiche  $\rightsquigarrow$  4 Blöcke mit Länge  $2^{k-2} - 1$
- 3.) usw.

► Gesamtanzahl der Vergleiche:

$$\begin{aligned} & (2^k - 2) + 2 \cdot (2^{k-1} - 2) + \dots + 2^{k-2} \cdot (2^2 - 2) \\ &= (k-1)2^k - 2 \sum_{j=0}^{k-2} 2^j \\ &= (k-1)2^k - 2(2^{k-1} - 1) \\ &= (k-2) \cdot 2^k + 2 \approx n \cdot \log_2 n. \end{aligned}$$

Später: Die erwartete Anzahl der benötigten Vergleiche ist etwa

$$\underbrace{2 \log 2}_{\approx 1,39} \cdot \text{Anzahl Vergleiche im best case.}$$

Wir bestimmen nun die **Zähldichte der zufälligen Anzahl von Vergleichen**, die zum Sortieren benötigt werden.

Da die Auswahl im 1. Schritt des Algorithmus jeweils zufällig gleichverteilt erfolgt, hängt die Verteilung der Anzahl von Vergleichen nur von der Anzahl der Daten ab, nicht von ihren genauen Werten oder ihrer Reihenfolge.

Wir definieren

$$Z(X) := \text{Zahl der Vergleiche, um Vektor } X \text{ zu sortieren,}$$

so dass nach der Konstruktion des Algorithmus

$$Z(X) = n - 1 + Z(X_l) + Z(X_r)$$

gilt und die Zähldichte

$$f_n(m) := P(Z(x_1, \dots, x_n) = m) = P(Z(X_l) + Z(X_r) = m - (n-1)), \quad m \in \mathbb{N},$$

nur von  $n$  abhängt.

Bezeichne nun  $K$  die zufällige Stelle, die die im 1. Schritt des Algorithmus ausgewählte Zahl **nach** Ordnung hat. Das Ereignis  $\{K = k\}$  bedeutet also gerade, dass die  $k$ -te kleinste Zahl ausgewählt worden ist.

Bei gegebenem Wert  $k$  sind  $Z(X_l), Z(X_r)$  stochastisch unabhängig, da die zu vergleichenden Zahlen in den Teilvektoren  $X_l$  und  $X_r$  jeweils unabhängig gewählt werden. Da dann  $Z(X_l)$  und  $Z(X_r)$  die Länge  $k - 1$  bzw.  $n - k$  haben, gilt

$$P(Z(X_l) + Z(X_r) = m - n + 1 \mid K = k) = (f_{k-1} * f_{n-k})(m - n + 1).$$

Es folgt mit dem Satz von der totalen Wahrscheinlichkeit

$$\begin{aligned} f_n(m) &= P(Z(X_l) + Z(X_r) = m - n + 1) \\ &= \sum_{k=1}^n P(K = k) P(Z(X_l) + Z(X_r) = m - n + 1 \mid K = k) \\ &= \sum_{k=1}^n \frac{1}{n} \cdot (f_{k-1} * f_{n-k})(m - n + 1) \\ &= \frac{1}{n} \sum_{k=1}^n \sum_{j=0}^{m-n+1} f_{k-1}(j) \cdot f_{n-k}(m - n + 1 - j). \end{aligned}$$

► Für  $n \geq 3$  kann  $f_n$  rekursiv berechnet werden, da

$$\begin{aligned} f_0(0) &= 1, & f_0(k) &= 0 \quad \forall k \neq 0, & f_1(0) &= 1, & f_1(k) &= 0 \quad \forall k \neq 0, \\ f_2(1) &= 1, & f_2(k) &= 0 \quad \forall k \neq 1. \end{aligned}$$

Es gilt z.B.  $f_3(2) = 1/3, f_3(3) = 2/3, f_3(k) = 0 \quad \forall k \in \mathbb{N}_0 \setminus \{2, 3\}$ .

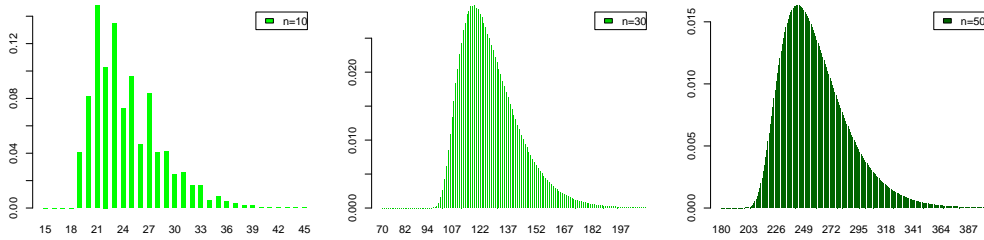


Abb.: Zähldichten der benötigten Vergleiche für Quicksort für  $n \in \{10, 30, 50\}$ .

► Die Verteilung der Vergleichszahl konzentriert sich in einem Bereich, der deutlich kleiner die als worst case Anzahl ist.  $\diamond$

## 3.2 Markovketten

Im vorangegangenen Beispiel sind die Anzahlen der zum Sortieren der Teilvektoren  $X_l$  und  $X_r$  benötigten Vergleiche nicht vollständig stochastisch unabhängig voneinander.

Die Anzahl der noch benötigten Vergleiche hängt jedoch von den bis dahin erfolgten Sortierschritten nur über die Länge der Vektoren in der aktuellen Zerlegung des Datensatzes ab, nicht von dem Weg, wie diese Zerlegung erreicht worden ist.

Folgen von Zufallsvariablen, die eine entsprechende Abhängigkeitsstruktur aufweisen, treten in vielen Anwendungen auf.

Im Folgenden sei  $S$  immer eine höchstens abzählbare Menge.

**Definition 3.17.** Eine Folge  $(X_n)_{n \in \mathbb{N}_0}$  von  $S$ -wertigen Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum  $(\Omega, P)$  heißt **Markovkette**, falls sie die folgende sog. **Markoveigenschaft** erfüllt:

$$\begin{aligned} P(X_{n+1} = s_{n+1} \mid X_n = s_n, X_{n-1} = s_{n-1}, \dots, X_0 = s_0) &= P(X_{n+1} = s_{n+1} \mid X_n = s_n), \\ \forall n \in \mathbb{N}, s_0, \dots, s_{n+1} \in S \text{ mit } P(X_n = s_n, X_{n-1} = s_{n-1}, \dots, X_0 = s_0) &> 0. \end{aligned}$$

Die Markovkette heißt **homogen**, falls die sogenannten **Ein-Schritt-Übergangswahrscheinlichkeiten**

$$p_{ts} := P(X_{n+1} = t \mid X_n = s)$$

für alle  $n \in \mathbb{N}_0$  mit  $P(X_n = s) > 0$  gleich sind.

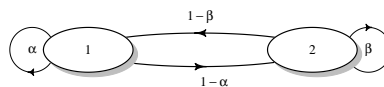
► Bei Markovketten hängt das zukünftige stochastische Verhalten gegeben dem gegenwärtigen Zustand nicht von der echten Vergangenheit ab.

Das stochastische Verhalten einer homogenen Markov-Kette ist eindeutig bestimmt durch die Startverteilung  $P^{X_0}$  (z.B. durch die Angabe der zugehörigen Zähldichte  $f_0(s) = P(X_0 = s)$ ,  $s \in S$ ) und die Übergangswahrscheinlichkeiten  $p_{ts}$  mit  $t, s \in S$ . Dann gilt z.B.

$$\begin{aligned} P(X_0 = s_0, X_1 = s_1, X_2 = s_2) &= P(X_2 = s_2 \mid X_1 = s_1, X_0 = s_0) \cdot P(X_1 = s_1, X_0 = s_0) \\ &= p_{s_2 s_1} \cdot P(X_1 = s_1 \mid X_0 = s_0) \cdot P(X_0 = s_0) \\ &= p_{s_2 s_1} \cdot p_{s_1 s_0} \cdot f_0(s_0). \end{aligned}$$

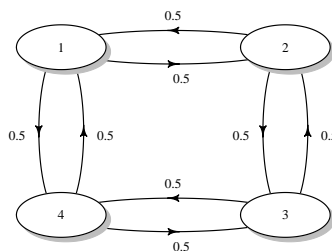
► Allgemeiner gilt für beliebige Ereignisse  $A_i$ ,  $0 \leq i \leq n$ ,

$$P(X_i \in A_i \forall 0 \leq i \leq n) = \sum_{s_0 \in A_0} \cdots \sum_{s_n \in A_n} p_{s_n s_{n-1}} \cdot p_{s_{n-1} s_{n-2}} \cdots \cdots p_{s_1 s_0} \cdot f_0(s_0).$$



Übergangsdiaagramm allgemeiner MK auf  $I = \{1, 2\}$ , mit  $\alpha, \beta \in [0, 1]$ . Die Übergangsmatrix ist

$$P = \begin{pmatrix} \alpha & 1 - \alpha \\ 1 - \beta & \beta \end{pmatrix}.$$



$I = \{1, 2, 3, 4\}$ . Zugehörige MK hat die Übergangsmatrix

$$P = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}.$$

Die MK ist *periodisch*: Ist der Irrfahrer zur Zeit 0 in  $\{1, 3\}$ , dann ist er mit W'keit 1 zu ungeraden Zeitpunkten in einem der Zustände aus  $\{2, 4\}$  und zu geraden Zeitpunkten in einem der Zustände aus  $\{1, 3\}$ .

*Beispiel 3.18* (Glücksspiel).

Spieler  $A$  und  $B$  spielen folgendes Glücksspiel:

Spieler  $A$  bestimmt pro Runde den Einsatz und wirft eine Münze. Fällt Kopf, so zahlt  $B$  den Einsatz an  $A$ , sonst zahlt  $A$  den Einsatz an  $B$ . Spieler  $A$  darf bestimmen, wie lange gespielt wird. Zu Anfang hat Spieler  $A$   $1\text{€}$  und benötigt  $5\text{€}$ . Er spielt daher so lange, bis er die  $5\text{€}$  zusammen hat oder sein gesamtes Kapital verspielt hat. Er entschließt sich für folgende „kühne Strategie“ (*bold play*)

$$\text{Einsatz von } A = \begin{cases} \text{gesamtes aktuelles Kapital,} & \text{falls Kapital} \leq \frac{5}{2}\text{€} \\ 5 - \text{Kapital,} & \text{falls Kapital} > \frac{5}{2}\text{€} \end{cases}$$

Die Wahrscheinlichkeit, dass Kopf fällt, sei jeweils  $p$ , und die Würfe sollen sich gegenseitig nicht beeinflussen.

Die Zufallsvariable  $X_n$  **bezeichne das Kapital von  $A$  nach  $n$  Spielen**. (Der Wert soll sich nicht mehr ändern, sobald das Spiel beendet ist.)

►  $(X_n)_{n \in \mathbb{N}_0}$  **ist eine homogene Markov-Kette**.

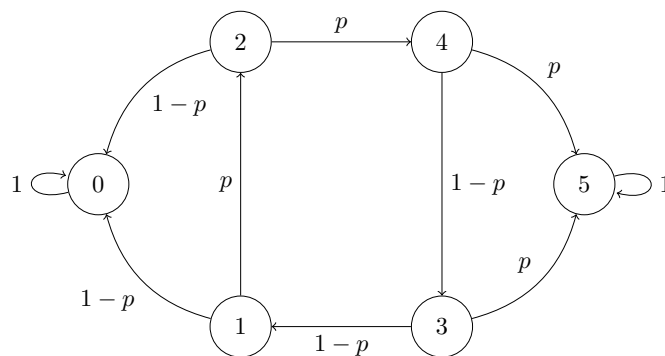


Abb.: **Übergangsgraph**, d.h. der Graph aller möglichen Übergänge von  $X_n$  nach  $X_{n+1}$  mit die zugehörigen Übergangswahrscheinlichkeiten im „bold play“.

*Frage:* Mit welcher Wahrscheinlichkeit erhält Spieler  $A$  schließlich  $5\text{€}$ ?

Bezeichne  $p_k$  die Wahrscheinlichkeit, dass Spieler  $A$  bei einem Kapital von  $k$  z.Z.  $n$  tatsächlich  $5\text{€}$  erhält, bevor er pleite geht:

$$p_k := P(X_l = 5 \text{ für ein } l > n \mid X_n = k),$$

wobei  $p_k$  nicht von  $n$  abhängt.

Unter Verwendung der Markoveigenschaft erhalten wir

$$\begin{aligned}
 p_1 &= P(X_l = 5 \text{ für ein } l > n \mid X_n = 1) \\
 &= P(X_{n+1} = 2, X_l = 5 \text{ für ein } l > n \mid X_n = 1) \\
 &\quad + \underbrace{P(X_{n+1} = 0, X_l = 5 \text{ für ein } l > n \mid X_n = 1)}_{=0} \\
 &= \frac{P(X_n = 1, X_{n+1} = 2, X_l = 5 \text{ für ein } l > n + 1)}{P(X_n = 1, X_{n+1} = 2)} \cdot \frac{P(X_n = 1, X_{n+1} = 2)}{P(X_n = 1)} \\
 &= P(X_l = 5 \text{ für ein } l > n + 1 \mid X_n = 1, X_{n+1} = 2) \cdot P(X_{n+1} = 2 \mid X_n = 1) \\
 &\stackrel{ME}{=} P(X_l = 5 \text{ für ein } l > n + 1 \mid X_{n+1} = 2) p_{21} \\
 &= p_2 \cdot p
 \end{aligned}$$

Auf gleiche Art und Weise erhält man insgesamt die Beziehungen

$$\begin{aligned}
 p_1 &= p \cdot p_2 \\
 p_2 &= p \cdot p_4 \\
 p_3 &= p + (1 - p) \cdot p_1 \\
 p_4 &= p + (1 - p) \cdot p_3
 \end{aligned}$$

Als lineares Gleichungssystem in Matrixschreibweise erhalten wir

$$\underbrace{\begin{pmatrix} 1 & -p & 0 & 0 \\ 0 & 1 & 0 & -p \\ p-1 & 0 & 1 & 0 \\ 0 & 0 & p-1 & 1 \end{pmatrix}}_{=:D} \cdot \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ p \\ p \end{pmatrix}$$

Da  $D$  invertierbar ist, existieren eindeutige Lösungen  $p_1, \dots, p_4$ .

$$\rightsquigarrow p_1 = \frac{(2-p) \cdot p^3}{1 - p^2 + 2p^3 - p^4}$$

Wir betrachten nun den vorsichtigeren Spieler  $\tilde{A}$ , der in der gleichen Situation immer nur  $1 \in$  setzt. (*timid play*)

Es ergeben sich folgende Übergangswahrscheinlichkeiten:

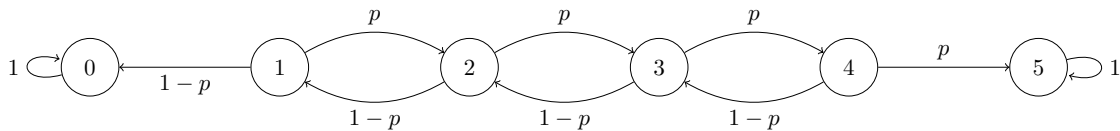


Abb.: Graph aller möglichen Übergänge von  $X_n$  nach  $X_{n+1}$  mit die zugehörigen Übergangswahrscheinlichkeiten im „timid play“.

Das Gleichungssystem der Erfolgswahrscheinlichkeiten lautet nun

$$\begin{pmatrix} 1 & -p & 0 & 0 \\ p-1 & 1 & -p & 0 \\ 0 & p-1 & 1 & -p \\ 0 & 0 & p-1 & 1 \end{pmatrix} \cdot \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ p \end{pmatrix}$$

$$\leadsto p_1 = \frac{p^4}{1 - 3p + 4p^2 - 2p^3 + p^4}$$

Der Vergleich beider Strategien ergibt folgende Graphik:

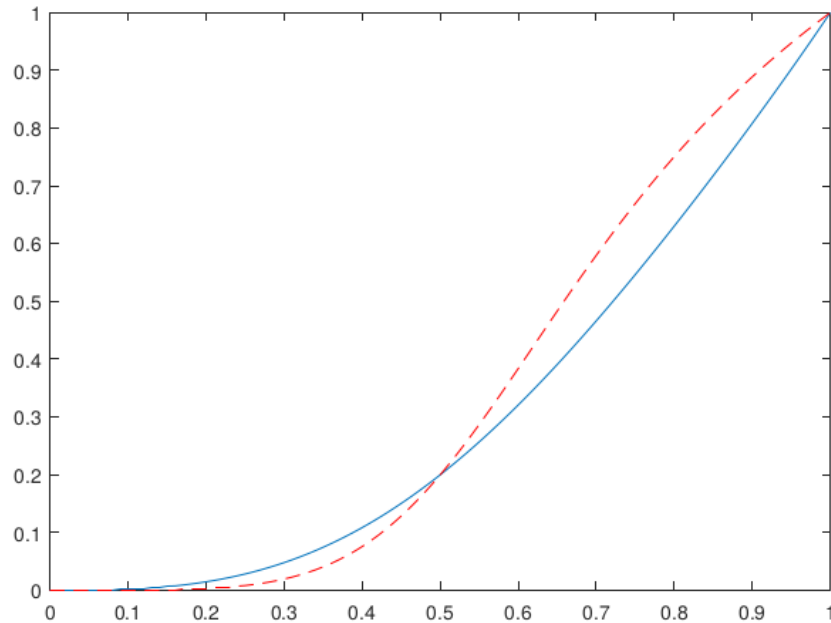


Abb.: Erfolgsw'keit  $p_1$  als Funktion von  $p$  für die Strategien *bold play* (blau durchgezogen) bzw. *timid play* (rot gestrichelt)

Der Vergleich der Wahrscheinlichkeiten  $p_1$ , dass Spieler  $A$  bzw.  $\tilde{A}$  mit der jeweiligen Strategie schließlich 5€ erhält, zeigt:

- Für  $p < 1/2$ : Bei der kühnen Strategie ist  $p_1$  größer als bei der vorsichtigen Strategie. Man kann zeigen: **Die kühne Strategie maximiert die Erfolgswahrscheinlichkeit  $p_1$ !**
- Für  $p > 1/2$ : Die Erfolgswahrscheinlichkeit  $p_1$  bei der vorsichtigen Strategie ist größer als bei der kühnen Strategie.
- Für  $p = 1/2$ : Bei einer fairen Münze beträgt die Erfolgswahrscheinlichkeit bei beiden Strategien  $p_k = k/5$ . Man kann zeigen: Dies gilt für alle möglichen Strategien. Das heißt **bei einem fairen Spiel kann man seine Gewinnwahrscheinlichkeit nicht durch geschicktes Spielen verbessern!**

◇

## 4 Erwartungswerte und Momente von Zufallsvariablen

### 4.1 Erwartungswerte und ihre Eigenschaften

Neben den Eintrittswahrscheinlichkeiten gibt es andere Kenngrößen, die einen Eindruck über Zufallsvariablen vermitteln.

**Definition 4.1.** (i) Der **Erwartungswert** einer  $\mathbb{R}$ -wertigen Zufallsvariable  $X: \Omega \rightarrow \mathbb{R}$  auf einem diskreten Wahrscheinlichkeitsraum  $(\Omega, P)$  ist definiert als

$$E_P[X] := E[X] := \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\}),$$

falls  $\sum_{\omega \in \Omega} |X(\omega)| \cdot P(\{\omega\}) < \infty$ . (Ansonsten besitzt  $X$  keinen (endlichen) Erwartungswert.)

(ii) Sei  $(\mathbb{R}, Q)$  ein diskreter W'raum und  $\text{Id}: \mathbb{R} \rightarrow \mathbb{R}$  die Identität, dann heißt

$$\mu(Q) := E_Q[\text{Id}] = \sum_{x \in \mathbb{R}} x \cdot Q(\{x\})$$

**Mittelwert** von  $Q$ , falls  $\sum_{x \in X} |x| \cdot Q(\{x\}) < \infty$ .

*Beispiele 4.2.*

1. Sei  $X = \mathbb{1}_A$  für ein  $A \subseteq \Omega$ . Dann folgt

$$E[\mathbb{1}_A] = E[X] = \sum_{\omega \in A} \cdot P(\{\omega\}) = P(A).$$

2. Nun sei  $P = \mathcal{P}_\lambda$  auf  $\Omega = \mathbb{N}_0$  und  $X = \text{Id}$ , sodass  $X$  Poisson-verteilt ist:  $P(X = n) = P(\{n\}) = e^{-\lambda} \cdot \frac{\lambda^n}{n!}$  für alle  $n \in \mathbb{N}_0$ . Dann gilt:

$$\begin{aligned} \mu(\mathcal{P}_\lambda) = E[X] &= \sum_{n=0}^{\infty} n P(\{n\}) = \sum_{n=0}^{\infty} n \cdot e^{-\lambda} \cdot \frac{\lambda^n}{n!} \\ &= \lambda \sum_{n=1}^{\infty} e^{-\lambda} \cdot \frac{\lambda^{n-1}}{(n-1)!} \stackrel{k=n-1}{=} \lambda \sum_{k=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\ &= \lambda \sum_{k=0}^{\infty} \mathcal{P}_\lambda(\{k\}) = \lambda. \end{aligned}$$

**Satz 4.3** (Transformationssatz). Sei  $(\Omega, P)$  ein diskreter Wahrscheinlichkeitsraum,  $X: \Omega \rightarrow S$  eine Zufallsvariable und  $g: S \rightarrow \mathbb{R}$ . Dann gilt:

$$E_P[\underbrace{g(X)}_{=g \circ X}] = E_{P^X}[g] = \sum_{s \in S} g(s) \cdot P^X(\{s\}).$$

Insbesondere gilt für reellwertige Zufallsvariablen  $X: \Omega \rightarrow \mathbb{R}$

$$E_P[X] = \sum_{x \in X(\Omega)} x \cdot P^X(\{x\}) = \sum_{x \in X(\Omega)} x \cdot P(X = x).$$

► Zu Berechnung des Erwartungswertes  $E_P[g(X)]$  benötigen wir nur die Verteilung  $P^X$  der ZV  $X$ , aber nicht das zugrundeliegende W'maß  $P$ .

*Beweis.* Mit Hilfe der Zerlegung  $\Omega = \bigcup_{s \in S} X^{-1}(\{s\})$  erhalten wir:

$$\begin{aligned}
 E_P(g(X)) &= \sum_{\omega \in \Omega} g(X(\omega)) \cdot P(\{\omega\}) \\
 &= \sum_{s \in S} \sum_{\omega \in \Omega: X(\omega)=s} g(X(\omega)) \cdot P(\{\omega\}) \\
 &= \sum_{s \in S} g(s) \cdot \sum_{\omega \in X^{-1}(\{s\})} P(\{\omega\}) \\
 &= \sum_{s \in S} g(s) \cdot P(X = s) \\
 &= \sum_{s \in S} g(s) \cdot P^X(\{s\}).
 \end{aligned}$$

□

*Beispiel 4.4.* Sei  $X$  Poissonverteilt mit Erwartungswert  $\lambda > 0$ , d.h.  $P^X = \mathcal{P}_\lambda$ . Dann gilt mit  $g(x) := x^2$

$$\begin{aligned}
 E[X^2] &= E_{\mathcal{P}_\lambda}[g] = \sum_{n=0}^{\infty} g(n) \mathcal{P}_\lambda(\{n\}) \\
 &= \sum_{n=0}^{\infty} n^2 \frac{\lambda^n}{n!} e^{-\lambda} \\
 &= \sum_{n=0}^{\infty} n(n-1) \frac{\lambda^n}{n!} e^{-\lambda} + \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} \\
 &= \sum_{n=2}^{\infty} \frac{\lambda^{n-2}}{(n-2)!} \lambda^2 e^{-\lambda} + \mu(\mathcal{P}_\lambda) \\
 &= \lambda^2 \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} + \lambda \\
 &= \lambda^2 + \lambda.
 \end{aligned}$$

◇

*Beispiel 4.5.* Sei  $X$  eine Zufallsvariable mit  $P(X = 1) = 1 - p$  und  $P(X = 2) = p$ . Dann gilt  $E[X] = 1 \cdot (1 - p) + 2p$  und

$$E[1/X] = 1 \cdot (1 - p) + \frac{1}{2}p.$$

Und im Allgemeinen ist  $E[1/X] \neq 1/E[X]$ .

◇

Zwei elementare Rechenregeln für Erwartungswerte liefert uns folgendes Resultat.

**Satz 4.6.** Seien  $X, Y$  Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum  $(\Omega, P)$ , die Erwartungswerte besitzen. Dann gilt:

- (i)  $E[aX + Y] = a \cdot E[X] + E[Y]$  für alle  $a \in \mathbb{R}$  (**Linearität**).
- (ii) Gilt  $X \leq Y$  (d.h.  $X(\omega) \leq Y(\omega)$  für alle  $\omega \in \Omega$ ), dann folgt  $E(X) \leq E(Y)$  (**Monotonie**).

*Beweis.* (i) Es gilt

$$\begin{aligned}
 E[aX + Y] &= \sum_{\omega \in \Omega} (aX + Y)(\omega)P(\{\omega\}) \\
 &= \sum_{\omega \in \Omega} (aX(\omega) + Y(\omega))P(\{\omega\}) \\
 &= a \sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) + \sum_{\omega \in \Omega} Y(\omega)P(\{\omega\}) \\
 &= aE[X] + E[Y].
 \end{aligned}$$

(ii) Aus der  $\omega$ -weisen Abschätzung folgt

$$E[X] = \sum_{\omega \in \Omega} \underbrace{X(\omega)}_{\leq Y(\omega)} \underbrace{P(\{\omega\})}_{\geq 0} \leq \sum_{\omega \in \Omega} Y(\omega)P(\{\omega\}) = E[Y]. \quad \square$$

*Beispiel 4.7* (Mittelwert der Binomialverteilung).

Sei  $X$  eine Zufallsvariable auf  $(\Omega, P)$  mit  $P^X = \mathcal{B}_{(n,p)}$ . Gesucht ist  $E[X]$ .

1. *Lösung:* Berechnen nach Definition:

$$\begin{aligned}
 E[X] &= \sum_{k=0}^n k \cdot P(X = k) \\
 &= \sum_{k=0}^n k \cdot \binom{n}{k} \cdot p^k (1-p)^{n-k} \\
 &= n \cdot p \cdot \sum_{k=1}^n \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} \cdot p^{k-1} \cdot (1-p)^{(n-1)-(k-1)} \\
 &\stackrel{j=k-1}{=} n \cdot p \cdot \sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} \cdot p^j \cdot (1-p)^{n-1-j} \\
 &= n \cdot p \cdot (p + 1 - p)^{n-1} \\
 &= n \cdot p.
 \end{aligned}$$

2. *Lösung:* Seien  $X_i$  unabhängige  $\mathcal{B}_{(1,p)}$ -verteilte Zufallsvariablen, d.h.  $P(X_i = 1) = p$  und  $P(X_i = 0) = 1 - p$ .

Gemäß Beispiel 3.6 ist dann  $\sum_{i=1}^n X_i$  gerade  $\mathcal{B}_{(n,p)}$ -verteilt, hat also dieselbe Verteilung wie  $X$ . Daher folgt

$$\begin{aligned}
 E[X] &= E\left[\sum_{i=1}^n X_i\right] \\
 &= \sum_{i=1}^n E[X_i] \\
 &= \sum_{i=1}^n (0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1)) = n \cdot p.
 \end{aligned}$$

Wir erinnern uns an den **Erdős–Rényi-Graph** aus Beispiel 1.6 mit  $n \in \mathbb{N}$  Knoten: Ob eine Kante zwischen zwei Knoten gezogen wird, sei durch unabhängige Bernoulli-verteilte ZVn  $X_i \sim \mathcal{B}_{(1,p)}$ ,  $i = 1, \dots, \binom{n}{2}$  mit Verbindungsweite  $p \in (0, 1)$  beschrieben.

► Wir erwarten im Mittel  $\binom{n}{2}p$  Kanten im Erdős–Rényi-Graph. ◇

*Beispiel 4.8* (Erwartete Laufzeit von Quicksort).

Wir setzen Beispiel 3.14 fort.

Sei  $Z(X) = Z(x_1, \dots, x_n)$  die zufällige Zahl der Vergleiche, die benötigt werden, um  $x_1, \dots, x_n$  zu sortieren. Gesucht ist

$$\mu_n := E[Z(X)] = \sum_{j=0}^{\infty} j \cdot P(Z(X) = j) = \sum_{j=0}^{\infty} j f_n(j).$$

Dabei bezeichnet  $f_n$  die Zähldichte von  $Z(X)$ , die nur von  $n$  abhängt (vgl. Bsp. 3.14), so dass auch  $\mu_n$  nur von  $n$  abhängt.

Es wird zufällig eine Zahl  $x_j$  gleichverteilt aus  $x_1, \dots, x_n$  ausgewählt; alle Zahlen kleiner als  $x_j$  werden in einem Vektor  $X_l$  zusammen gefasst, die größeren Zahlen in einem Vektor  $X_r$ . Es folgt

$$\begin{aligned} \mu_n &= E[Z(X)] \\ &= E[n - 1 + Z(X_l) + Z(X_r)] \\ &= n - 1 + E[Z(X_l)] + E[Z(X_r)]. \end{aligned}$$

Sei  $x_j$  die  $K$ -te kleinste Zahl. Dann hat der Vektor  $X_l$  der Zahlen kleiner als  $x_j$  die Länge  $K - 1$  und der Vektor  $X_r$  der Zahlen größer als  $x_j$  die Länge  $n - K$ .

Daher gilt

$$\begin{aligned} E[Z(X_l)] &= \sum_{j=0}^{\infty} j P(Z(X_l) = j) \\ &= \sum_{j=0}^{\infty} j \sum_{k=1}^n P(Z(X_l) = j \mid K = k) P(K = k) \\ &= \sum_{j=0}^{\infty} j \sum_{k=1}^n f_{k-1}(j) \cdot \frac{1}{n} \\ &= \frac{1}{n} \cdot \sum_{k=1}^n \sum_{j=0}^{\infty} j f_{k-1}(j) \\ &= \frac{1}{n} \cdot \sum_{k=1}^n \mu_{k-1} = \frac{1}{n} \cdot \sum_{j=0}^{n-1} \mu_j. \end{aligned}$$

Ebenso erhält man

$$E[Z(X_r)] = \frac{1}{n} \cdot \sum_{k=1}^n \mu_{n-k} = \frac{1}{n} \cdot \sum_{j=0}^{n-1} \mu_j,$$

zusammen ergibt sich die Rekursionsformel

$$\mu_n = n - 1 + \frac{2}{n} \sum_{j=0}^{n-1} \mu_j.$$

Es folgt für alle  $n \geq 2$

$$n\mu_n = n(n-1) + 2 \sum_{j=0}^{n-1} \mu_j \quad \text{und} \quad (n-1)\mu_{n-1} = (n-1)(n-2) + 2 \sum_{j=0}^{n-2} \mu_j$$

und somit

$$n\mu_n - (n-1)\mu_{n-1} = 2(n-1) + 2\mu_{n-1} \quad \iff \quad \mu_n = \frac{n+1}{n} \mu_{n-1} + 2 \frac{n-1}{n}.$$

Mit vollständiger Induktion kann man zeigen, dass

$$\mu_n = 2(n+1) \sum_{j=1}^n \frac{1}{j} - 4n \begin{cases} \leq & 2n \log n \\ \geq & 2n \log n - 4n \end{cases} \sim 2n \log n$$

Aus Beispiel 3.14 ist bekannt, dass die Laufzeit im best case  $\sim n \log_2 n = n \log n / \log 2$  ist.

► Die erwartete Laufzeit für große Datensätze verhält sich, wie das 2 log 2-fache der kürzesten Laufzeit!

◇

Eine hilfreiche Anwendung von Erwartungswerten ist die *Siebformel von Sylvester-Poincaré* bzw. das *Einschluss-Ausschluss-Prinzip*. Diese kann man zwar auch elementar mit den Mitteln aus Kapitel 1 beweisen, jedoch ermöglichen Erwartungswerte dank der oben bewiesenen Rechenregeln eine deutlich einfachere Beweisführung (statt über vollständige Induktion).

Mit den Mitteln aus Kapitel 1 lässt sich leicht zeigen:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C) \end{aligned}$$

Allgemeiner gilt:

**Satz 4.9.** Seien  $A_1, \dots, A_n \subseteq \Omega$  Ereignisse in einem diskreten Wahrscheinlichkeitsraum  $(\Omega, P)$ . Dann gilt:

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_j \cap A_i) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \\ &\quad \dots + (-1)^{n-1} P(A_1 \cap \dots \cap A_n) \\ &= \sum_{I \subseteq \{1, \dots, n\}, I \neq \emptyset} (-1)^{|I|+1} P\left(\bigcap_{i \in I} A_i\right). \end{aligned}$$

*Beweis.* Wegen  $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$  und  $\mathbb{1}_{A \cap B} = \mathbb{1}_A \cdot \mathbb{1}_B$  für beliebige Ereignisse  $A, B \subseteq \Omega$  gilt

$$\mathbb{1}_{\bigcup_{i=1}^n A_i} = 1 - \mathbb{1}_{(\bigcup_{i=1}^n A_i)^c} = 1 - \mathbb{1}_{\bigcap_{i=1}^n A_i^c} = 1 - \prod_{i=1}^n \mathbb{1}_{A_i^c} = 1 - \prod_{i=1}^n (1 - \mathbb{1}_{A_i}).$$

Ausmultiplizieren liefert

$$\mathbb{1}_{\bigcup_{i=1}^n A_i} = \sum_{i=1}^n \mathbb{1}_{A_i} - \sum_{1 \leq i < j \leq n} \mathbb{1}_{A_i} \mathbb{1}_{A_j} + \dots + (-1)^{n+1} \prod_{i=1}^n \mathbb{1}_{A_i}.$$

Die Linearität des Erwartungswertes ergibt damit

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= E[\mathbb{1}_{\bigcup_{i=1}^n A_i}] = \sum_{i=1}^n E[\mathbb{1}_{A_i}] - \sum_{1 \leq i < j \leq n} \underbrace{E[\mathbb{1}_{A_i} \mathbb{1}_{A_j}]}_{=P(A_i \cap A_j)} \\ &\quad + \dots + (-1)^{n+1} \underbrace{E\left[\prod_{i=1}^n \mathbb{1}_{A_i}\right]}_{=P(\bigcap_{i=1}^n A_i)}. \end{aligned}$$

Wir erhalten also genau die behauptete Formel. □

*Beispiel 4.10* (Rencontre-Problem).

Beim Schrottwichteln bringt jeder der  $n$  teilnehmenden Studis ein „Geschenk“ mit. Diese werden rein zufällig auf die  $n$  Studis verteilt.

*Frage:* Mit welcher W'keit erhält wenigstens ein Studi sein eigenes Geschenk?

*Modell:*  $\Omega = S_n := \{\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\} \mid \pi \text{ bijektiv}\}$  (Menge aller Permutationen von  $1, \dots, n$ ) versehen mit der Gleichverteilung  $P$ .

*Interpretation:*  $\pi(i)$  entspricht dem Studi, welcher das Geschenk von Studi  $i$  erhält.

► Gesucht ist  $P(A)$  für

$$A := \{\pi \in \Omega \mid \exists i \in \{1, \dots, n\} : \pi(i) = i\} = \bigcup_{i=1}^n A_i$$

mit  $A_i := \{\pi \in \Omega \mid \pi(i) = i\}$ .

Die Siebformel liefert

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

wobei

$$\begin{aligned} P\left(\bigcap_{i \in I} A_i\right) &= \frac{|\bigcap_{i \in I} A_i|}{n!} \\ &= \frac{1}{n!} |\{\pi \in \Omega \mid \pi(i) = i \forall i \in I\}| \\ &= \frac{(n - |I|)!}{n!}. \end{aligned}$$

$$\blacktriangleright \quad P(A) = \sum_{I \subseteq \{1, \dots, n\}, I \neq \emptyset} (-1)^{|I|+1} \frac{(n - |I|)!}{n!}.$$

Weiter gilt:

$$\begin{aligned} P(A) &= \sum_{I \subseteq \{1, \dots, n\}, I \neq \emptyset} (-1)^{|I|+1} \frac{(n - |I|)!}{n!} \\ &= \sum_{k=1}^n \sum_{I \subseteq \{1, \dots, n\}, |I|=k} (-1)^{k+1} \frac{(n - k)!}{n!} \\ &= \sum_{k=1}^n \binom{n}{k} (-1)^{k+1} \frac{(n - k)!}{n!} \\ &= \sum_{k=1}^n (-1)^{k+1} \frac{1}{k!} \\ &= 1 - \sum_{k=0}^n \frac{(-1)^k}{k!} \rightarrow 1 - e^{-1} \quad \text{für } n \rightarrow \infty. \end{aligned}$$

Die W'keit konvergiert sehr schnell:

$n$	1	2	5	10	$\infty$
$P(A)$	1	0,5	0,63	$\approx 0,63212054$	$\approx 0,63212056$

◇

## 4.2 Momente von Zufallsvariablen

**Definition 4.11.** Sei  $X$  eine  $\mathbb{R}$ -wertige Zufallsvariable auf einem diskreten Wahrscheinlichkeitsraum  $(\Omega, P)$ .

(i) Existieren  $E[X]$  und  $E[X^2]$ , so ist die **Varianz** von  $X$  durch

$$\text{Var}(X) := E[(X - E[X])^2] = \sum_{x \in X(\Omega)} (x - E[X])^2 \cdot P(X = x)$$

definiert.  $\sqrt{\text{Var}(X)}$  heißt **Standardabweichung** von  $X$ .

(ii) Für  $k \in \mathbb{N}$  heißt  $E[X^k]$  (im Falle der Existenz) das  **$k$ -te Moment von  $X$** . Dabei ist  $X^k : \Omega \rightarrow \mathbb{R}$  definiert durch  $X^k(\omega) = (X(\omega))^k$ .

Tatsächlich folgt aus der Existenz von  $E[X^2]$  auch immer die Existenz von  $E[X]$  (*Cauchy-Schwarz-Ungleichung*).

*Bemerkung 4.12.* Die Varianz ist die mittlere quadratische Abweichung von  $X$  von seinem Erwartungswert, also ein Maß dafür, wie stark die zufälligen Werte um den Erwartungswert streuen. Es gilt

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

Insbesondere gilt:

$$0 \leq \text{Var}(X) = E[X^2] - E[X]^2 \implies E[X^2] \geq E[X]^2.$$

Allgemeiner gilt für alle  $a \in \mathbb{R}$ :

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

► Die Minimalstelle der Funktion  $a \mapsto E[(X - a)^2]$  ist  $a = E[X]$ .

*Beispiel 4.13.*

1. Sei  $X$   $\mathcal{P}_\lambda$ -verteilt, d.h.  $P(X = n) = e^{-\lambda} \cdot \frac{\lambda^n}{n!}$  für alle  $n \in \mathbb{N}_0$ . Nach Beispiel 4.2.2 ist  $E[X] = \lambda$  und nach Beispiel 4.4 ist  $E[X^2] = \lambda^2 + \lambda$ . Damit folgt

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

2. Sei  $X \sim \mathcal{B}_{(1,p)}$  Bernoulli-verteilt mit Erfolgsw'keit  $p \in [0, 1]$ . Dann ist  $E[X] = p$  und

$$\text{Var}(X) = E[X^2] - E[X]^2 \stackrel{X \in \{0,1\}}{=} E[X] - E[X]^2 = p(1 - p). \quad \diamond$$

Wir kommen zu zwei wichtigen Abschätzung für die sogenannten Tails einer Verteilung, d.h. Abschätzungen für die Wahrscheinlichkeit, dass eine Zufallsvariable große Werte annimmt.

**Satz 4.14.**

(i)

**Markov-Ungleichung**  $X$  sei eine Zufallsvariable mit existierendem Erwartungswert  $E[X]$ . Dann gilt

$$P(|X| \geq c) \leq \frac{E[|X|]}{c} \quad \forall c > 0.$$

(ii)

**Chebyshev-Ungleichung**  $X$  sei eine Zufallsvariable mit existierender Varianz  $\text{Var}(X)$ . Dann gilt

$$P(|X - E[X]| \geq c) \leq \frac{\text{Var}(X)}{c^2} \quad \forall c > 0.$$

*Beweis.* (i) Aufgrund der Monotonie des Erwartungswertes und wegen  $\mathbb{1}_{\{|x| \geq c\}} \leq \frac{|x|}{c}$  für alle  $x \in \mathbb{R}$  und  $c > 0$  erhalten wir

$$P(|X| \geq c) = E[\mathbb{1}_{\{|X| \geq c\}}] \leq E\left[\frac{|X|}{c}\right] = \frac{E[|X|]}{c}.$$

(ii) Da für positive Zahlen  $a, b > 0$  die Äquivalenz  $a \leq b \iff a^2 \leq b^2$  gilt (Monotonie von  $\mathbb{R}_+ \ni a \mapsto a^2$ ) erhalten wir analog zu (i):

$$P(|X - E[X]| \geq c) = P((X - E[X])^2 \geq c^2) \leq \frac{E[(X - E[X])^2]}{c^2} = \frac{\text{Var}(X)}{c^2}. \quad \square$$

*Beispiel 4.15.*

Sei  $X \mathcal{P}_\lambda$ -verteilt und  $\lambda > 0$ . Somit wissen wir  $E[X] = \lambda = \text{Var}(X)$ .

1. Für die Markov-Ungleichung gilt mit  $c > 0$ :

$$P(X \geq c) \leq \frac{E[X]}{c} = \frac{\lambda}{c}$$

2. Die Chebyshev-Ungleichung liefert für  $c > \lambda$ :

$$\begin{aligned} P(X \geq c) &= P(X - \lambda \geq c - \lambda) \\ &\leq P(|X - \lambda| \geq c - \lambda) \leq \frac{\text{Var}(X)}{(c - \lambda)^2} = \frac{\lambda}{(c - \lambda)^2} \end{aligned}$$

Es gilt:  $\begin{cases} \text{Markov ist schärfer, falls} & c < \lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}} \\ \text{Chebyshev ist schärfer, falls} & c > \lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}} \end{cases}$

Oft sind beide Schranken viel größer als  $P(X \geq c)$  (möglicherweise sogar  $\geq 1$ ).

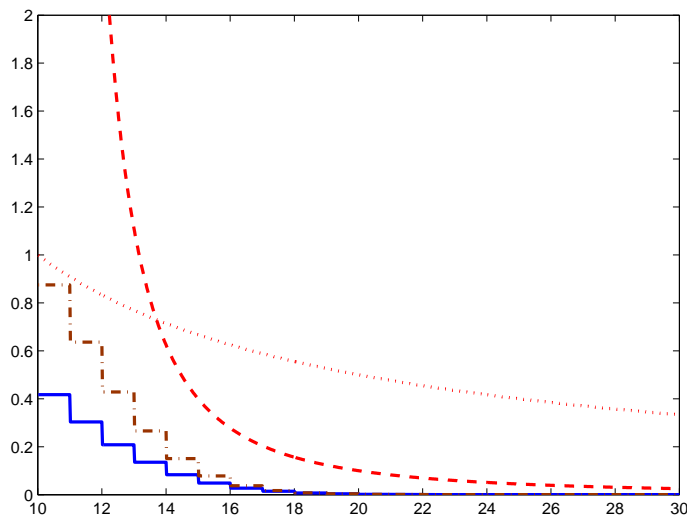


Abb.: Für eine  $\mathcal{P}_\lambda$ -verteilte ZV  $X$  mit  $\lambda = 10$  die W'keiten  $P(X \geq c)$  (blau durchgezogen) und  $P(|X - \lambda| \geq c - \lambda)$  (braun Strich-Punkt), sowie die Markov-Schranke  $\lambda/c$  (rot gepunktet) und die Chebyshev-Schranke  $\lambda/(c - \lambda)^2$  (rot gestrichelt) jeweils als Funktion von  $c > \lambda$

◇

*Beispiel 4.16.*

Wir betrachten noch einmal das Beispiel Quicksort.

Sei  $Z(x_1, \dots, x_n)$  die benötigte Anzahl der Vergleiche, um die Zahlen  $x_1, \dots, x_n$  zu sortieren. Nach Beispiel 4.7 wissen wir

$$2n \log n - 4n \leq E[Z(x_1, \dots, x_n)] \leq 2n \log n.$$

Auf ähnliche Weise kann man zeigen, dass  $\text{Var}(Z(x_1, \dots, x_n)) \leq 3n(n-1)$ .

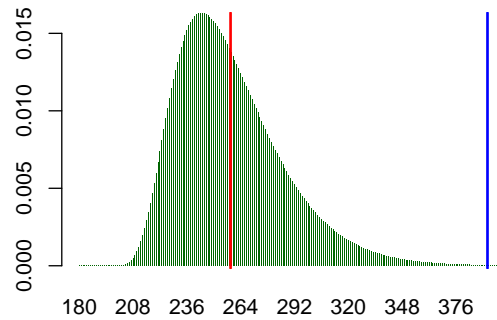
$$\begin{aligned} & \blacktriangleright P(Z(x_1, \dots, x_n) \geq 2n \log n + a \cdot n) \\ & \leq P(|Z(x_1, \dots, x_n) - E(Z(x_1, \dots, x_n))| \geq \underbrace{2n \log n - E(Z(x_1, \dots, x_n))}_{\geq 0} + a \cdot n) \\ & \leq P(|Z(x_1, \dots, x_n) - E(Z(x_1, \dots, x_n))| \geq a \cdot n) \\ & \leq \frac{\text{Var}(Z(x_1, \dots, x_n))}{(a \cdot n)^2} \quad (\text{Chebyshev-Ungleichung}) \\ & \leq \frac{3n(n-1)}{(a \cdot n)^2} < \frac{3}{a^2}. \end{aligned}$$

Für  $a = \varepsilon \cdot \log n$  mit  $\varepsilon > 0$  ergibt sich:

$$P(Z(x_1, \dots, x_n) \geq (2 + \varepsilon)n \log n) \leq \frac{3}{\varepsilon^2 (\log n)^2} \xrightarrow{n \rightarrow \infty} 0 \quad \forall \varepsilon > 0$$

► Mit großer W'keit verhält sich die Anzahl der Vergleiche im Wesentlichen wie  $2n \log n$ , wenn der Datenumfang  $n$  groß ist.

Abb.: Zähldichte der Vergleichszahl von Quicksort für  $n = 50$  mit Mittelwert  $\mu_{50} \approx 259$  (rot) und Schranke  $2n \log n \approx 391$  (blau).



◇

*Bemerkung 4.17.* Die Varianz ist im Gegensatz zum Erwartungswert nicht linear im Argument. Vielmehr gilt für Zufallsvariablen  $X, Y$  mit  $E[X^2], E[Y^2] < \infty$ :

$$\begin{aligned} \text{Var}(aX + b) &= E[(aX + b - E[aX + b])^2] = E[(aX + b - (aE[X] + b))^2] \\ &= a^2 \cdot \text{Var}(X), \quad \forall a, b \in \mathbb{R}, \end{aligned}$$

sowie

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

Die vorangegangene Formel für die Varianz der Summe von Zufallsvariablen bringt uns direkt auf einen weiteren zentralen Begriff.

**Definition 4.18.** Es seien  $X, Y$  Zufallsvariablen mit Erwartungswerten  $E[X]$  und  $E[Y]$ .

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

heißt (im Falle der Existenz) die **Kovarianz von  $X$  und  $Y$** .

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

heißt dann **Korrelation** von  $X$  und  $Y$ , falls  $\text{Var}(X) > 0$  und  $\text{Var}(Y) > 0$ .  
 $X$  und  $Y$  heißen **unkorreliert**, falls  $\text{Cov}(X, Y) = 0$ .

*Beispiel 4.19.* Seien  $X: \Omega \rightarrow \{0, 1\}$  und  $Y: \Omega \rightarrow \{-1, 0, 1\}$  ZVn auf einem diskreten W'raum  $(\Omega, P)$ . In der nachfolgenden Tabelle ist die Zähl-dichte  $P\{X = x, Y = y\}$  der gemeinsamen Verteilung von  $(X, Y)$  angegeben:

$x \backslash y$	-1	0	1
0	1/10	1/20	1/10
1	3/10	3/10	3/20

Zur Berechnung der Kovarianz und Korrelation:

- Berechne die Zähl-dichten von  $X, Y$  über die Zeilen- bzw. Spaltensummen.
- Berechne  $E[X] = 3/4$ ,  $E[Y] = -3/20$  und  $E[Y^2] = 13/20$ .
- Berechne  $E[XY] = \sum_{x=0}^1 \sum_{y=-1}^1 xyP(X = x, Y = y) = -3/20$ .
- Es folgt  $\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = -3/80$ .
- Berechne  $\text{Var}(X) = E[X^2] - E[X]^2 = 3/16$  und  $\text{Var}(Y) = 251/400$ .
- ▶ Es folgt  $\text{Corr}(X, Y) = -\sqrt{3/251}$ . ◇

*Bemerkung 4.20.* Existiert die Kovarianz zweier Zufallsvariablen  $X, Y$ , so gilt für alle  $a, b, c, d \in \mathbb{R}$

$$\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y).$$

Ist  $Z$  eine weitere Zufallsvariable, so dass  $\text{Cov}(X, Z)$  existiert, so gilt außerdem

$$\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z).$$

Diese sogenannte *Bilinearität* der Kovarianz, d.h. die Kovarianz ist in beiden Argumenten jeweils linear, kann man leicht anhand der Definition nachrechnen.

**Satz 4.21.** *Seien  $X$  und  $Y$  stochastisch unabhängige reellwertige Zufallsvariablen auf einem diskreten W'raum  $(\Omega, P)$  mit Erwartungswerten  $E[X]$  und  $E[Y]$ . Dann gilt*

$$E[X \cdot Y] = E[X] \cdot E[Y].$$

*Insbesondere sind unabhängige Zufallsvariablen unkorreliert.*

Man beachte, dass aus Unkorreliertheit nicht Unabhängigkeit folgt:

*Beispiel 4.22.* Sei  $X$  gleichverteilt auf  $\{-1, 0, 1\}$  und  $Y = X^2$ . Dann gilt

$$E[XY] = E[X^3] = -1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0 = 0 \cdot \frac{2}{3} = E[X]E[Y], \text{ aber}$$

$$P(X = 1, Y = 1) = P(X = 1) = \frac{1}{3} \neq \frac{1}{3} \cdot \frac{2}{3} = P(X = 1)P(Y = 1). \quad \diamond$$

*Beweis.* Für die abzählbaren Träger  $\Omega_X, \Omega_Y \subseteq \Omega$  von  $X$  bzw.  $Y$  gilt aufgrund der Unabhängigkeit

$$\begin{aligned} E[XY] &= \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} xyP(X = x, Y = y) \\ &= \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} xyP(X = x)P(Y = y) \\ &= \left( \sum_{x \in \Omega_X} xP(X = x) \right) \left( \sum_{y \in \Omega_Y} yP(Y = y) \right) \\ &= E[X]E[Y]. \end{aligned} \quad \square$$

**Korollar 4.23.** Für Zufallsvariablen  $X_1, \dots, X_n$  gilt

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

falls die Varianzen alle existieren. Sind die Zufallsvariablen  $X_1, \dots, X_n$  unabhängig (oder schwächer: unkorreliert), so gilt insbesondere:

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

*Beweis.* Es gilt aufgrund  $\text{Var}(Z) = \text{Cov}(Z, Z)$  für Zufallsvariablen  $Z$  mit existierendem zweiten Moment, der Bilinearität der Kovarianz und der Symmetrie  $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$ :

$$\begin{aligned} \text{Var} \left( \sum_{i=1}^n X_i \right) &= \text{Cov} \left( \sum_{i=1}^n X_i, \sum_{j=1}^n X_j \right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n \text{Cov}(X_i, X_i) + \sum_{i,j \in \{1, \dots, n\}, i \neq j} \text{Cov}(X_i, X_j) \\ &= \text{Var} \left( \sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j). \end{aligned}$$

Der Zusatz folgt direkt aus  $\text{Cov}(X_i, X_j) = 0$  für unabhängige bzw. unkorrelierte Zufallsvariablen  $X_i, X_j$ .  $\square$

*Beispiel 4.24.*  $X$  sei eine  $\mathcal{B}_{(n,p)}$ -verteilte Zufallsvariable.

►  $X$  hat dieselbe Verteilung wie  $\sum_{i=1}^n X_i$ , wobei  $X_i \sim \mathcal{B}_{(1,p)}$  stochastisch unabhängig sind (vgl. Beispiel 3.6).

► Es gilt

$$\begin{aligned} \text{Var}(X) &= \text{Var} \left( \sum_{i=1}^n X_i \right) \\ &\stackrel{\text{Korr. 4.23}}{=} \sum_{i=1}^n \text{Var}(X_i) \\ &= n \cdot \text{Var}(X_1) \\ &\stackrel{\text{Bsp. 4.13}}{=} np(1-p). \end{aligned}$$

◇

*Bemerkung 4.25.*

*Ziel:* Der Wert einer ZV  $X$  wurde beobachtet. Daraus soll der Wert der ZV  $Y$  vorhergesagt werden. Zur Approximation von  $Y$  soll eine lineare Funktionen von  $X$  verwendet werden.

Als Maß für die Approximationsgüte / Vorhersagegüte verwenden wir den mittleren quadratischen Vorhersagefehler

$$E[(Y - (aX + b))^2].$$

► Wähle  $a$  und  $b$  so, dass  $E[(Y - (aX + b))^2]$  minimal wird.

Es gilt

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

► Der erste Summand ist  $\geq 0$  und hängt nicht von  $b$  ab, sodass der mittlere quadratische Vorhersagefehler für  $b = E[Y] - aE[X]$  minimiert wird, für den der 2. Summand (auch  $\geq 0$ ) verschwindet.

Der erste Summand

$$\begin{aligned} \text{Var}(Y - aX) &= \text{Var}(Y) + 2 \text{Cov}(Y, -aX) + \text{Var}(-aX) \\ &= \text{Var}(Y) - 2a \text{Cov}(X, Y) + a^2 \text{Var}(X) \\ &= \left( a\sqrt{\text{Var}(X)} - \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}} \right)^2 + \text{Var}(Y) - \frac{\text{Cov}(X, Y)^2}{\text{Var}(X)} \end{aligned}$$

wird durch die Wahl

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

minimiert, falls  $\text{Var}(X) > 0$  gilt.

► Die beste lineare Approximation für  $Y$  durch  $X$  ist also

$$\frac{\text{Cov}(X, Y)}{\text{Var}(X)} \cdot (X - E[X]) + E[Y].$$

Die beste lineare Approximation für  $Y$  durch  $X$  ist

$$\frac{\text{Cov}(X, Y)}{\text{Var}(X)} \cdot (X - E[X]) + E[Y].$$

Wir standardisieren  $X$  und  $Y$  nun affin linear so, dass sie Erwartungswert 0 und Varianz 1 besitzen:

$$\tilde{X} := \frac{X - E[X]}{\sqrt{\text{Var}(X)}}, \quad \tilde{Y} = \frac{Y - E[Y]}{\sqrt{\text{Var}(Y)}}.$$

Die beste lineare Approximation für  $\tilde{Y}$  durch  $\tilde{X}$  ist dann folglich

$$\frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \cdot \text{Var}(Y)} \cdot \frac{X - E[X]}{\sqrt{\text{Var}(X)}} = \text{Corr}(X, Y) \cdot \tilde{X}.$$

Der zugehörige mittlere quadratische Approximationsfehler ist

$$E[(\tilde{Y} - \text{Corr}(X, Y) \cdot \tilde{X})^2] = 1 - (\text{Corr}(X, Y))^2.$$

► Insbesondere folgt  $\text{Corr}(X, Y) \in [-1, 1]$ , da der Erwartungswert einer nicht-negativen Zufallsvariable nicht negativ sein kann.

Im Fall  $|\text{Corr}(X, Y)| = 1$  ist der Vorhersagefehler gleich 0. Im Fall  $|\text{Corr}(X, Y)| = 0$  ist der Vorhersagefehler maximal.

**Achtung:** Die Korrelation ist nur ein Maß für die Stärke des linearen Zusammenhangs zwischen  $X$  und  $Y$ .

Es ist möglich, dass  $X, Y$  unkorreliert sind, aber  $Y$  vollständig durch  $X$  bestimmt ist, solange der Zusammenhang nicht linear ist (siehe Beispiel 4.22).

Neben dem Erwartungswert gibt es weitere „Lageparameter“, die „mittlere Werte“ einer Zufallsvariable beschreiben.

**Definition 4.26.** Ist  $X$  eine  $\mathbb{R}$ -wertige Zufallsvariable, so heißt jede Zahl  $m(X)$  **Median** von  $X$  (bzw. vom  $P^X$ ), falls für die reelle (deterministische) Zahl  $m = m(X)$

$$P(X \leq m) \geq \frac{1}{2} \quad \text{und} \quad P(X \geq m) \geq \frac{1}{2}$$

gilt.

Mediane müssen nicht eindeutig bestimmt sein.

*Beispiel 4.27.* Ist  $X$  auf  $\{0, \dots, n\}$  gleichverteilt, so ist  $\frac{n}{2}$  der eindeutig bestimmte Median, wenn  $n$  gerade ist. Es gilt nämlich mit der Gaußklammer  $[x] := \max\{k \in \mathbb{Z} | k \leq x\}$ :

$$\begin{aligned} P(X \leq m) = \frac{[m+1]}{n+1} &\implies \left( P(X \leq m) \geq \frac{1}{2} \iff m \geq \frac{n}{2} \right), \\ P(X \geq m) = \frac{[n-m+1]}{n+1} &\implies \left( P(X \geq m) \geq \frac{1}{2} \iff m \leq \frac{n}{2} \right). \end{aligned}$$

Ist  $n$  hingegen ungerade, so ist jede Zahl  $m \in \left[ \frac{n-1}{2}, \frac{n+1}{2} \right]$  ein Median von  $X$ , da

$$\begin{aligned} P\left(X \leq \frac{n-1}{2}\right) &= \frac{\left[\frac{n-1}{2} + 1\right]}{n+1} = \frac{1}{2}, \\ P\left(X \geq \frac{n+1}{2}\right) &= \frac{\left[n - \frac{n-1}{2} + 1\right]}{n+1} = \frac{1}{2}. \end{aligned}$$

In diesem Fall wird mitunter der Mittelpunkt des Intervalls, welches aus den Medianen gebildet werden kann, „der“ Median genannt.  $\diamond$

*Bemerkung 4.28.* Der Median verhält sich unter linearen Transformationen wie der Erwartungswert. Mit  $a, b \in \mathbb{R}$  gilt also

$$m \text{ ist Median von } X \iff am + b \text{ ist Median von } aX + b$$

Größen, die sich so verhalten, werden auch als **Lageparameter** bezeichnet.

Der Erwartungswert ist auf natürliche Art und Weise mit der Varianz bzw. der Standardabweichung als Streuungsmaß verknüpft (vgl. Bem. 4.12).

Auf analoge Weise ist der erwartete absolute Abstand vom Median  $m(X)$ , also  $E[|X - m(X)|]$ , als Streuungsmaß mit dem Median verknüpft: Man kann zeigen, dass ein Wert  $m$  genau dann Median ist, wenn er  $E(|X - m|)$  minimiert.

Wie für die Standardabweichung gilt unter linearer Transformation mit  $a, b \in \mathbb{R}$

$$E(|aX + b - m(aX + b)|) = |a|E(|X - m(X)|).$$

Nicht immer wird in praktischen Anwendungen sauber zwischen Erwartungswert und Median unterscheiden, obwohl diese weit auseinander liegen können. So ist die “erwartete Lebenszeit“ eines Neugeborenen in Wirklichkeit der (empirische) Median.

Insbesondere Ausreißer haben einen unterschiedlich starken Einfluss auf  $E[X]$  bzw.  $m(X)$ .

*Beispiel 4.29.* Sei  $X$  eine  $\{0, 1, 2, n\}$ -wertige ZV mit  $P(X = i) = \frac{1-\varepsilon}{3}$ ,  $i \in \{0, 1, 2\}$ , und  $P(X = n) = \varepsilon$  für ein  $\varepsilon \in (0, \frac{1}{4})$ ,  $n > 2$ . Dann gilt

$$E[X] = 1 - \varepsilon + \varepsilon \cdot n \quad \text{und} \quad m(X) = 1,$$

denn  $P(X \leq 1) = \frac{2}{3}(1 - \varepsilon) \geq \frac{1}{2}$  und  $P(X \geq 1) = \frac{2}{3}(1 - \varepsilon) + \varepsilon \geq \frac{1}{2}$ .

► Während  $E[X]$  linear in  $n$  wächst, spielt die Größe des Ausreißers  $n$  für  $m(X)$  keine Rolle.  $m(X)$  ist damit **robust** gegenüber Ausreißern. ◊

### 4.3 Anwendung auf Schätzprobleme

In Anwendungen kennt man typischerweise die zugrundeliegende Wahrscheinlichkeitsverteilung eines Zufallsexperiments nicht, kann aber ggf. eine Familie von W’maßen beschreiben, welche in Frage kommen, vgl. Beispiel 3.7.

**Definition 4.30.** Ein **statistisches Experiment** oder **statistisches Modell** ist gegeben durch  $(\Omega, (P_\vartheta)_{\vartheta \in \Theta})$ , wobei  $P_\vartheta$  diskrete W’maße auf  $\Omega$  sind.  $\vartheta$  ist der unbekannte **Parameter**.  $\Theta$  heißt **Parameterraum**.

► Ziel der Statistik sind Rückschlüsse von den Beobachtungen auf den zugrundeliegenden Parameter, insbesondere wollen wir den Parameter **schätzen** oder wir wollen **testen**, ob der Parameter einen bestimmten Wert hat oder in einem bestimmten Bereich liegt.

Wir betrachten ein statistisches Experiment  $(\Omega, (P_\vartheta)_{\vartheta \in \Theta})$  und wollen einen **Schätzer**  $\hat{\vartheta}: \Omega \rightarrow \Theta$  für  $\vartheta$  aufgrund einer Beobachtung  $\omega$  konstruieren. Folgende zwei Lösungsansätze führen häufig zu guten Schätzverfahren:

- **Maximum-Likelihood (maximale Plausibilität):** Wähle den Parameter  $\hat{\vartheta} = \hat{\vartheta}(\omega)$  unter dem das Ergebnis/die Beobachtung  $\omega$  die größte W’keit hat (am plausibelsten ist), d.h.

$$P_{\hat{\vartheta}(\omega)}(\{\omega\}) = \max_{\vartheta \in \Theta} P_\vartheta(\{\omega\}).$$

Betrachten wir Zähldichte  $f_\vartheta(\omega) = P_\vartheta(\{\omega\})$  als Funktion in  $\vartheta$  für einen festes  $\omega$ , nennen wir sie **Likelihood-Funktion**.

- **Momentenmethode:** Wir betrachten eine Zufallsvariable  $X: \Omega \rightarrow \mathbb{R}$  mit existierendem Erwartungswert  $E_\vartheta[X]$  und  $P_\vartheta$  für jedes  $\vartheta \in \Theta$ . Falls  $X$  nahe an  $E_\vartheta[X]$  liegt (wir also eine relativ geringe Streuung um den Erwartungswert haben), wählen wir das  $\hat{\vartheta} = \hat{\vartheta}(\omega)$ , für das gilt:

$$E_{\hat{\vartheta}(\omega)}[X] = X(\omega).$$

*Beispiel 4.31.* Gegeben eine einzige Beobachtung einer ZV mit Werten in  $\{0, 1, 2\}$  mit Verteilung  $P_\vartheta$ ,  $\vartheta \in \{\vartheta_0, \vartheta_1\}$  und folgenden Zähldichten:

	$x = 0$	$x = 1$	$x = 2$
$\vartheta = \vartheta_0$	0.7	0.2	0.1
$\vartheta = \vartheta_1$	0.2	0.3	0.5

Wird  $X = 0$  beobachtet, dann ist  $P_{\vartheta_0}$  plausibler. Man würde in diesem Fall schätzen, dass  $\vartheta$  durch  $\vartheta_0$  gegeben ist.

Bei einer Beobachtung  $X = 1$  oder  $X = 2$  ist umgekehrt  $\vartheta = \vartheta_1$  plausibler.

Folgende Schätzfunktion bietet sich also an

$$\hat{\vartheta}(X) = \begin{cases} \vartheta_0 & : X = 0, \\ \vartheta_1 & : X \neq 0. \end{cases}$$

◇

*Beispiel 4.32.* Ein Meinungsumfrageinstitut möchte herausfinden, wie hoch der Anteil  $\vartheta \in (0, 1)$  der Befürworter einer  $CO_2$ -Steuer ist. Hierfür werden unabhängig voneinander und zufällig gleichverteilt ausgewählt  $n$  Leute befragt (ggf. Mehrfachzählung). Die Anzahl der Befürworter  $X$  in der Umfrage führt auf das *statistische Modell*

$$(\Omega, (P_\vartheta^X)_{\vartheta \in \Theta}) \quad \text{mit} \quad \Omega = \{0, \dots, n\}, P_\vartheta^X = \mathcal{B}_{n, \vartheta}, \vartheta \in \Theta = [0, 1].$$

► *Maximum-Likelihood:* Wir maximieren  $\vartheta \mapsto f_\vartheta(k) = \binom{n}{k} \vartheta^k (1 - \vartheta)^{n-k}$ : Für  $0 < k < n$  gilt

$$0 \stackrel{!}{=} \frac{\partial}{\partial \vartheta} f_\vartheta(k) = \binom{n}{k} (k(1 - \vartheta) - (n - k)\vartheta) \vartheta^{k-1} (1 - \vartheta)^{n-k-1}.$$

Lösen dieser Gleichung ergibt  $\hat{\vartheta}(k) = \frac{k}{n}$  (Prüfe hinreichende Bedingung und die Fälle  $k \in \{0, n\}$ ).

► *Momentenmethode:* Wir betrachten die Zufallsvariable  $X: \Omega \rightarrow \mathbb{R}, X(k) = k$  und erhalten

$$k \stackrel{!}{=} E_\vartheta[X] = n\vartheta.$$

Lösen dieser Gleichung führt ebenfalls auf  $\hat{\vartheta}(k) = \frac{k}{n}$ .

► **Erwarteter quadratischer Fehler** von  $\hat{\vartheta} = \frac{X}{n}$  unter  $P_\vartheta$ , d.h.  $X \sim \mathcal{B}_{n, \vartheta}$ :

$$E_\vartheta[(\hat{\vartheta} - \vartheta)^2] = \frac{1}{n^2} E_\vartheta[(X - n\vartheta)^2] = \frac{1}{n^2} \text{Var}_\vartheta(X) = \frac{\vartheta(1 - \vartheta)}{n}.$$

► Die Genauigkeit der Schätzung steigt für wachsende  $n$ .

◇

## 5 Grenzwertsätze

In diesem Kapitel werden wir zwei fundamentale Resultate der Wahrscheinlichkeitstheorie kennen lernen, welche das asymptotische Verhalten von gewichteten Mitteln unabhängiger Zufallsvariablen beschreiben.

Das erste Resultat dieser Art ist das sogenannte „Gesetz der großen Zahlen“.

Wir betrachten Summen von  $n$  Zufallsvariablen  $X_i$  für  $n \rightarrow \infty$ . Für große  $n$  ist die Verteilung von  $\sum_{i=1}^n X_i$  in der Regel nicht exakt oder nur schwer berechenbar.

*Ziel:* Eine gute Approximation der Verteilung der Summe für große  $n$ .

**Satz 5.1** (Schwaches Gesetz der großen Zahlen). *Seien  $X_i$ ,  $i \in \mathbb{N}$ , unkorrelierte Zufallsvariablen mit Erwartungswerten  $E[X_i]$  und existiere ein  $M < \infty$ , sodass  $\text{Var}(X_i) \leq M$  für alle  $i \in \mathbb{N}$ . Dann gilt für alle  $\varepsilon > 0$*

$$P\left(\frac{1}{n} \left| \sum_{i=1}^n (X_i - E[X_i]) \right| \geq \varepsilon\right) \leq \frac{M}{n\varepsilon^2} \rightarrow 0 \quad \text{für } n \rightarrow \infty$$

*Beweis.* Die Chebyshev-Ungleichung und die Unkorreliertheit der  $X_i$  zusammen mit Korollar 4.23 liefern:

$$\begin{aligned} P\left(\frac{1}{n} \left| \sum_{i=1}^n (X_i - E[X_i]) \right| \geq \varepsilon\right) &= P\left(\left| \frac{1}{n} \sum_{i=1}^n X_i - E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] \right| \geq \varepsilon\right) \\ &\leq \frac{1}{\varepsilon^2} \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2\varepsilon^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2\varepsilon^2} \sum_{i=1}^n \text{Var}(X_i) \\ &\leq \frac{1}{n^2\varepsilon^2} \sum_{i=1}^n M = \frac{M}{n\varepsilon^2} \rightarrow 0 \quad \text{für } n \rightarrow \infty \quad \square \end{aligned}$$

Dieser Satz führt uns auf die sogenannte *stochastische Konvergenz*.

**Definition 5.2.** Seien  $Y, Y_n$   $\mathbb{R}$ -wertige Zufallsvariablen.  $Y_n$  **konvergiert ( $P$ -)stochastisch gegen  $Y$** , falls

$$\forall \varepsilon > 0: \quad P(|Y_n - Y| \geq \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

Wir schreiben  $Y_n \xrightarrow{P} Y$  oder  $Y_n \rightarrow Y$   $P$ -stochastisch.

*Bemerkung 5.3.* Unter der Bedingung von Satz 5.1 gilt also

$$\frac{1}{n} \sum_{i=1}^n (X_i - E[X_i]) \xrightarrow{P} 0.$$

Ist  $E(X_i) = \mu$  für alle  $i \in \mathbb{N}$ , so gilt zudem  $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mu$ .

Beispiel 5.4.

Möchte man testen, ob eine Münze fair ist, dann werfe man sie  $n$  Mal und definiere für  $1 \leq i \leq n$

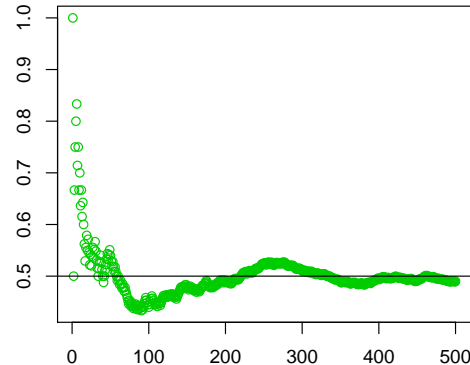
$$X_i := \begin{cases} 1, & \text{falls im } i\text{-ten Wurf Kopf fällt,} \\ 0, & \text{falls im } i\text{-ten Wurf Zahl fällt.} \end{cases}$$

$X_i$  sind unabhängig und  $\mathcal{B}_{1,p}$ -verteilt. Für eine faire Münze gilt  $p = 1/2$ .

► Wegen  $E[X_i] = p$  und des G.d.g.Z. konvergiert die relative Häufigkeit, mit der Kopf fällt:

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{P} p \quad P\text{-stochastisch.}$$

Abb.: Realisierung von  $\frac{1}{n} \sum_{i=1}^n X_i$  für  $X_i \sim \mathcal{B}_{1,1/2}$  und  $n \in \{1, \dots, 500\}$ .



Ist die Münze fair, so ist die relative Häufigkeit mit hoher W'keit nahe  $1/2$ .

Frage: Bei welchen Abweichungen von  $1/2$  kann man diese als deutlichen Hinweis auffassen, dass die Münze unfair ist (also  $p \neq 1/2$  gilt)?

Ist tatsächlich  $p = 1/2$ , so zeigt der Beweis von Satz 5.1

$$P\left(\left|\frac{Y_n}{n} - \frac{1}{2}\right| \geq \varepsilon\right) \leq \frac{\text{Var}(X_1)}{n\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} = \frac{1}{4n\varepsilon^2}.$$

► Eine Abweichung größer als  $\varepsilon = 1/\sqrt{4\tau n}$  tritt höchstens mit W'keit  $\tau$  auf.

Ist z.B.  $n = 500$  ( $1/\sqrt{n} \approx 0,045$ ), so liefert die Wahl  $\varepsilon = 0,1$

$$P(Y_n \notin (200, 300)) = P\left\{\left|\frac{Y_n}{n} - \frac{1}{2}\right| \geq 0,1\right\} \leq 0,05.$$

Liegt die tatsächliche Anzahl von Köpfen nicht im Intervall  $(200, 300)$ , so ist die Münze also vermutlich unfair. Allerdings ist das Intervall unnötig lang. Später wird der sog. zentrale Grenzwertsatz genauere Abschätzungen liefern.

◇

Beispiel 5.5 (Monte-Carlo-Simulationen).

Ziel: Berechne für eine Zufallsvariable  $Z$  den Erwartungswert  $E[Z]$ .

Problem: Verteilung  $P^Z$  ist oft zwar prinzipiell bekannt, aber nicht analytisch bestimmbar. Insbesondere ist dies oft der Fall, wenn  $Z$  von der Form  $f(Y_1, \dots, Y_k)$  für "einfache" Zufallsvariablen  $Y_i$  ist.

Idee: Simuliere unabhängige Zufallsvariablen  $Z_i$ , die verteilt sind wie  $Z$  (tats. „Pseudozufallszahlen“  $z_i$ , die sich im Wesentlichen wie  $Z_i$  verhalten).

Wegen des Gesetzes der großen Zahlen

$$\frac{1}{n} \sum_{i=1}^n Z_i \rightarrow E[Z] \quad P\text{-stochastisch}$$

ist zu hoffen, dass  $\frac{1}{n} \sum_{i=1}^n z_i \approx E[Z]$ , wenn  $n$  hinreichend groß ist.

Konkretes Beispiel:

Seien  $Y_1, Y_2$  unabhängig und gleichverteilt auf  $\{0, \frac{1}{L}, \frac{2}{L}, \dots, 1\} =: S_L$  für ein  $L \in \mathbb{N}$ . Somit sind  $(Y_1, Y_2)$  gleichverteilt auf  $S_L^2$ .

Falls  $L$  sehr groß ist, ist  $(Y_1, Y_2)$  näherungsweise gleichverteilt auf  $[0, 1]^2$ . Betrachte die Indikatorfunktion, dass  $(Y_1, Y_2)$  im Viertelkreis liegt:

$$Z = 1_{\{Y_1^2 + Y_2^2 \leq 1\}}.$$

Dann gilt für großes  $L$

$$\begin{aligned} E[Z] &= P(Y_1^2 + Y_2^2 \leq 1) \\ &\approx \frac{\text{Fläche Viertelkreis}}{\text{Fläche Einheitsquadrat}} = \frac{\pi}{4}. \end{aligned}$$

► Mittels  $z_i$  erhalten wir für große  $L$  und  $n$  eine Approximation von  $\pi$ :

$$4 \cdot \frac{1}{n} \sum_{i=1}^n z_i \approx 4 \cdot E(Z) \approx \pi.$$

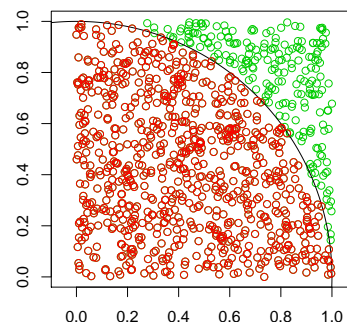


Abb.:  $n = 1000$  Realisierung von  $(Y_1, Y_2)$  mit  $L = 1000$  und resultierender Approximation  $\pi \approx 3.1480$ .

◇

Wir kommen nun zum zweiten fundamentalen Grenzwertresultat.

**Satz und Definition 5.6** (Zentraler Grenzwertsatz). Seien  $X_i, i \in \mathbb{N}$ , unabhängig und identisch verteilte Zufallsvariablen, d.h.  $P^{X_i} = P^{X_1} \forall i$ , (kurz: „i.i.d.“ für „independent and identically distributed“) mit Erwartungswert  $\mu = E[X_i]$  und Varianz  $\sigma^2 = \text{Var}(X_i) \in (0, \infty)$ .

Dann gilt für alle  $x \in \mathbb{R}$ :

$$P\left(\frac{1}{\sqrt{n \cdot \sigma^2}} \sum_{i=1}^n (X_i - \mu) \leq x\right) \xrightarrow{n \rightarrow \infty} \Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

$\Phi$  heißt **Standardnormalverteilungsfunktion** und

$$\varphi(t) := \Phi'(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-t^2/2}, \quad t \in \mathbb{R},$$

heißt **Standardnormalverteilungsdichte**

Der Beweis ist z.B. im Lehrbuch von Dümbgen zu finden.

Seien  $X_1, X_2, \dots$  unabhängige und identisch verteilte Zufallsvariablen mit Erwartungswert  $\mu \in \mathbb{R}$  und Varianz  $\sigma^2 \in (0, \infty)$ .

Dann gilt nach dem Gesetz der großen Zahlen:

$$\frac{1}{n} \sum_{i=1}^n X_i - \mu \xrightarrow{P} 0,$$

d.h. mit für große  $n$  ist mit großer Wahrscheinlichkeit ist  $\sum_{i=1}^n X_i$  nah an  $n\mu$ .

Zentraler Grenzwertsatz quantifiziert dieses „nah“. Er besagt, dass für große  $n$

$$\frac{\sqrt{n}}{\sigma} \left( \frac{1}{n} \sum_{i=1}^n X_i - \mu \right) = \frac{1}{\sigma \sqrt{n}} \left( \sum_{i=1}^n X_i - n\mu \right) = \frac{1}{\sigma \sqrt{n}} \sum_{i=1}^n (X_i - \mu)$$

annähernd standardnormalverteilt ist.

Die Standardnormalverteilungsfunktion bzw. -dichte besitzen einige wichtige Eigenschaften.

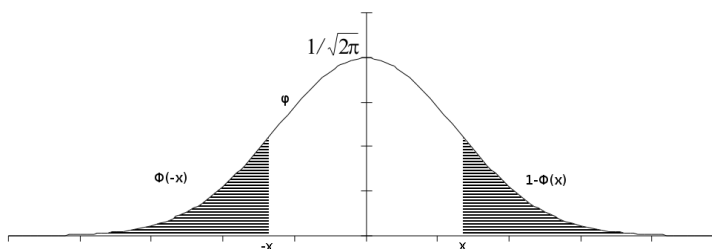
Es gelten:

$$\begin{aligned} \Phi(-\infty) &:= \lim_{x \rightarrow -\infty} \Phi(x) = 0, & \Phi(\infty) &:= \lim_{x \rightarrow \infty} \Phi(x) = 1, \\ \varphi(t) &= \varphi(-t), & \forall t &\in \mathbb{R}. \end{aligned}$$

Daraus folgt für alle  $x \in \mathbb{R}$

$$\begin{aligned} \Phi(-x) &= \int_{-\infty}^{-x} \varphi(t) dt \stackrel{u=-t}{=} \int_x^{\infty} \underbrace{\varphi(-u)}_{=\varphi(u)} du = \underbrace{\int_{-\infty}^{\infty} \varphi(u) du}_{=\Phi(\infty)=1} - \underbrace{\int_{-\infty}^x \varphi(u) du}_{=\Phi(x)} \\ &= 1 - \Phi(x) \end{aligned}$$

Insbesondere gilt damit  $\Phi(0) = \frac{1}{2}$ .



*Beispiel 5.7.*

Es seien  $X_1, 1 \leq i \leq n$ , i.i.d. Poisson-verteilt mit Parameter  $\lambda > 0$ . Somit gilt  $E[X_i] = \text{Var}(X_i) = \lambda$ , vgl. Beispiele 4.2 und 4.13.

► Der Zentrale Grenzwertsatz liefert (für großes  $n$ )

$$P\left(\frac{1}{\sqrt{n\lambda}} \sum_{i=1}^n (X_i - \lambda) \leq x\right) \approx \Phi(x)$$

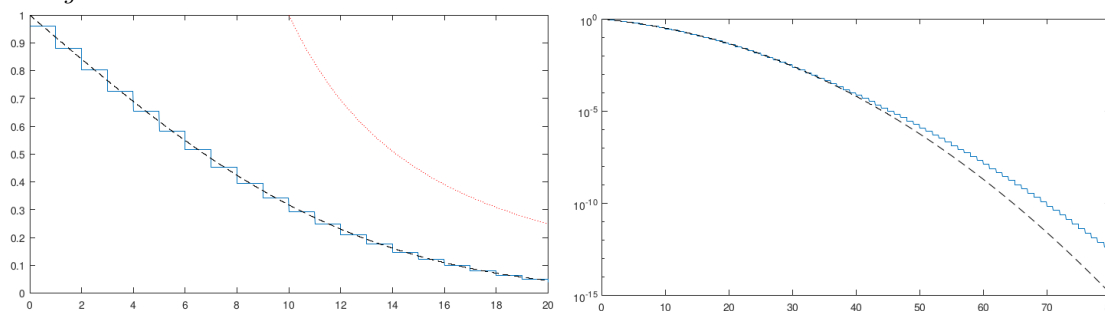
und daher

$$P\left(\left|\sum_{i=1}^n (X_i - \lambda)\right| \geq c\right) \approx \Phi(-c/\sqrt{n\lambda}) + 1 - \Phi(c/\sqrt{n\lambda}) = 2\Phi(-c/\sqrt{n\lambda}).$$

Andererseits ist  $\sum_{i=1}^n X_i$  Poisson-verteilt mit Parameter  $n\lambda$  (s. Bsp. 3.13), sodass

$$P\left(\left|\sum_{i=1}^n (X_i - \lambda)\right| \geq c\right) = 1 - \mathcal{P}_{n\lambda}(\lceil n\lambda - c \rceil, \dots, \lfloor n\lambda + c \rfloor).$$

*Vergleich dieser Wahrscheinlichkeiten:*



*Abb.:* W'keiten für  $n\lambda = 100$ , jeweils abgebildet als Funktion von  $c$ . Blau durchgezogen: tatsächliche W'keit, schwarz gestrichelt: Normalapproximation, rot gepunktet: Chebyshev'sche Schranke. Links:  $c \in [0, 20]$ , rechts  $c \in [0, 80]$  mit logarithmische Skala der  $y$ -Achse.

► Während für moderate  $c$  die Approximation sehr gut ist, ergeben sich größere Abweichungen für extremere Werte von  $c$ .  $\diamond$

Wenden wir den Zentralen Grenzwertsatz auf eine  $\mathcal{B}_{1,p}$ -verteilte i.i.d. Folge  $(X_i)_{i \geq 1}$  an, ergibt sich wegen  $E[X_i] = p$  und  $\text{Var}(X_i) = p(1-p)$ :

**Korollar 5.8** (Zentraler Grenzwertsatz von Moivre-Laplace). Ist  $Y_n$  eine  $\mathcal{B}_{(n,p)}$ -verteilte Zufallsvariable mit  $p \in (0, 1)$  für alle  $n \in \mathbb{N}$ , so gilt

$$P\left(a < \frac{Y_n - np}{\sqrt{np \cdot (1-p)}} \leq b\right) \xrightarrow{n \rightarrow \infty} \Phi(b) - \Phi(a) \quad \forall -\infty \leq a \leq b \leq \infty$$

*Beweis.* Für i.i.d.  $\mathcal{B}_{1,p}$ -verteilte (Bernoulli-)Zufallsvariablen  $(X_i)_{i \in \mathbb{N}}$  gilt, dass  $Y_n$  die selbe Verteilung hat wie  $\sum_{i=1}^n X_i$ . Zudem gilt (siehe Beispiele oben)  $\mu = E[X_i] = p$  sowie  $\sigma^2 := \text{Var}(X_i) = p(1-p)$ . Daher folgt aus dem Zentralen Grenzwertsatz

$$\begin{aligned} P\left(a < \frac{Y_n - np}{\sqrt{np \cdot (1-p)}} \leq b\right) &= P\left(\frac{Y_n - np}{\sqrt{np \cdot (1-p)}} \leq b\right) - P\left(\frac{Y_n - np}{\sqrt{np \cdot (1-p)}} \leq a\right) \\ &= P\left(\frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n\sigma^2}} \leq b\right) - P\left(\frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n\sigma^2}} \leq a\right) \\ &\xrightarrow{n \rightarrow \infty} \Phi(b) - \Phi(a) \quad \forall -\infty \leq a \leq b \leq \infty. \quad \square \end{aligned}$$

*Beispiel 5.9.*

Bei einem Flug kann das Flugzeug  $n_0 = 200$  Personen befördern. Mit Wahrscheinlichkeit  $p = 0,96$  erscheinen die Ticketinhaber jeweils unabhängig voneinander zum Flug.

*Frage:* Wie viele Tickets darf das Unternehmen verkaufen, damit die Wahrscheinlichkeit einer Überbuchung (d.h., dass mehr als  $n_0$  Personen zum Abflug erscheinen) maximal 0,05 beträgt? Bezeichne

$$\begin{aligned} X &= \text{Anzahl der Personen, die zum Flug erscheinen,} \\ n &= \text{Anzahl der verkauften Tickets,} \end{aligned}$$

so ist  $X$  gerade  $\mathcal{B}_{(n,p)}$ -verteilt.

► Wähle  $n$  nun maximal so, dass  $P(X > n_0) \leq 0,05$ .

Der Zentrale Grenzwertsatz (bzw. Korollar 5.8) liefert die Approximation

$$\begin{aligned} P(X > n_0) &= P\left(\frac{X - np}{\sqrt{np(1-p)}} > \frac{n_0 - np}{\sqrt{np(1-p)}}\right) \\ &\approx 1 - \Phi\left(\frac{n_0 - np}{\sqrt{np(1-p)}}\right) \stackrel{!}{\leq} 0,05. \end{aligned}$$

Soll nun die rechte Seite kleiner oder gleich 0,05 sein, so ist dies äquivalent zu

$$\begin{aligned} n &\leq \frac{n_0 + \frac{1-p}{2}(\Phi^{-1}(0,95))^2}{p} - \sqrt{\left(\frac{n_0 + \frac{1-p}{2}(\Phi^{-1}(0,95))^2}{p}\right)^2 - \left(\frac{n_0}{p}\right)^2} \\ &\approx 203,55. \end{aligned}$$

► Man kann also maximal 203 Tickets verkaufen.

Die tats. Überbuchungsw. beträgt dann nur  $\mathcal{B}_{(203,p)}\{201, 202, 203\} \approx 0,0113$ . ◇

*Beispiel 5.10* (Stetigkeitskorrektur).

Die obige Approximation lässt sich merklich verbessern!

Da  $X$  nur natürliche Zahlen als Werte annehmen kann, gilt

$$P(X > n_0) = P(X > t) \quad \forall t \in [n_0, n_0 + 1).$$

Es ergibt sich wie oben

$$P(X > t) \approx 1 - \Phi\left(\frac{t - np}{\sqrt{np(1-p)}}\right).$$

Oft wird der mittlere Wert  $t = n_0 + \frac{1}{2}$  des Konstanzintervalls verwendet. Diese Wahl nennt man **Stetigkeitskorrektur**, da die Sprungfunktion  $t \mapsto P\{X \leq t\}$  durch eine stetige Funktion approximiert wird.

Damit ergibt sich  $n \leq 204,06$ , d.h. max. 204 verkaufte Tickets. In der Tat gilt

$$P(X > n_0) = \mathcal{B}_{(n,p)}(n_0 + 1, \dots, n) \approx \begin{cases} 0,048 & \text{für } n = 204, \\ 0,094 & \text{für } n = 205. \end{cases}$$

◇

*Beispiel 5.11* (Konstruktion eines Tests, Fortsetzung von Beispiel 5.4).

Eine Münze wird  $n = 500$  mal unabhängig voneinander geworfen. Die W'keit für Kopf sei  $p$ .

*Gesucht:* Ein **Test**, ob  $p = 1/2$ , d.h. ob die Münze fair ist.

Wir setzen wieder

$$X_i = \begin{cases} 1, & \text{falls im } i\text{-ten Wurf Kopf fällt,} \\ 0, & \text{falls im } i\text{-ten Wurf Zahl fällt,} \end{cases}$$

und erhalten das *statistische Experiment*:

$$(\{0,1\}^n, (P_p^{(X_1, \dots, X_n)})_{p \in [0,1]}) \quad \text{mit} \quad P_p^{(X_1, \dots, X_n)}(\{\omega\}) = \prod_{i=1}^n p^{\omega_i} (1-p)^{1-\omega_i}, \omega \in \{0,1\}^n.$$

Die relative Häufigkeit, mit der Kopf fällt, ist gegeben durch

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i.$$

Die Münze soll als (vermutlich) unfair angesehen werden, falls  $|\bar{X}_n - 1/2|$  einen sog. **kritischen Wert**  $c$  überschreitet. ► Wie sollte  $c$  gewählt werden?

► Der Test lehnt Annahme, dass Münze fair ist, ab, wenn  $|\bar{X}_n - 1/2| > c$ .

Bei der Wahl des kritischen Werts  $c$  sind zwei gegenläufige Effekte zu berücksichtigen:

- Wählt man  $c$  groß, so ist es unwahrscheinlich, dass man eine *tatsächlich faire Münze irrtümlich als unfair bezeichnet* (**Fehler 1. Art**); andererseits wird man aber oft eine *tatsächlich unfaire Münze nicht als solche erkennen* (**Fehler 2. Art**).
- Wählt man  $c$  klein, so ist umgekehrt die W'keit, einen Fehler 1. Art zu begehen groß, aber die W'keit, einen Fehler 2. Art zu begehen, klein.

► Üblicherweise gibt man sich eine (kleine) obere Schranke  $\alpha \in (0,1)$ , das sog. **Signifikanzniveau**, für die W'keit vor, einen Fehler 1. Art zu begehen (der oft als problematischer angesehen wird), und versucht dann, die W'keit für einen Fehler 2. Art zu minimieren.

► Wähle  $\alpha$  (z.B.  $\alpha = 0,05$ ) und dann  $c = c_\alpha$  so, dass für eine faire Münze ( $p = 1/2$ ) die Wahrscheinlichkeit durch  $\alpha$  beschränkt ist:

$$P_{1/2} \left( \left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{2} \right| > c \right) \leq \alpha.$$

Mittels des Zentralen Grenzwertsatzes erhalten wir die Approximation

$$\begin{aligned} P_{1/2} \left( \left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{2} \right| > c \right) &\approx 1 - \Phi(\sqrt{n} \cdot 2c) + \Phi(-\sqrt{n} \cdot 2c) \\ &= 2\Phi(-2c\sqrt{n}) \\ &\stackrel{!}{=} \alpha. \end{aligned}$$

► Es folgt im Fall  $n = 500$ ,  $\alpha = 0,05$

$$c = -\frac{\Phi^{-1}(\alpha/2)}{2\sqrt{n}} \approx 0,0438.$$

Die Münze kann also als unfair angesehen werden, wenn die relative Häufigkeit, mit der Kopf fällt, **nicht** in dem Intervall  $(0,4562; 0,5438)$  liegt.

Vergleicht man das Ergebnis mit dem, dass wir aus der Chebyshev-Ungleichung gewonnen haben, also dem Intervall  $(0,4; 0,6)$ , so zeigt sich, dass es die Approximation mit dem Zentralen Grenzwertsatz ermöglicht, eine unfaire Münze viel eher zu erkennen.  $\diamond$

## 6 Wahrscheinlichkeitsmaße auf $\mathbb{R}$

### 6.1 Grundbegriffe reloaded

Kontinuierliche Ergebnisse eines Zufallsexperiments (z.B. Längen, Gewichte, Zeiten) lassen sich nicht auf natürliche Weise durch diskrete Wahrscheinlichkeitsräume beschreiben.

Im Allgemeinen kann man dann nicht mehr jeder Untermenge des Grundraums eine Wahrscheinlichkeit zuweisen, vgl. Bsp. 6.7 unten.

► Wahrscheinlichkeiten kann man nur noch „gutartigen“ Mengen zuordnen, u.a.:

- Intervalle sind gutartig.
- Komplemente gutartiger Mengen sind gutartig.
- Abzählbare Vereinigungen gutartiger Mengen sind gutartig.

Es bezeichne im folgenden  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  das System aller „gutartigen“ Mengen.

Wie müssen Systeme „gutartiger“ Mengen konstruiert werden?

**Definition 6.1.** Sei  $\Omega \neq \emptyset$  ein beliebiger Grundraum. Eine Menge  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  von Untermengen von  $\Omega$  heißt  **$\sigma$ -Algebra auf  $\Omega$** , falls folgende Bedingungen erfüllt sind:

- (a)  $\Omega \in \mathcal{A}$ ,
- (b)  $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$ ,
- (c)  $A_n \in \mathcal{A} \quad \forall n \in \mathbb{N} \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$ .

$(\Omega, \mathcal{A})$  heißt dann **Messraum**. Die Mengen  $A \in \mathcal{A}$  heißen **Ereignisse**.

►  $\sigma$ -Algebren modellieren die Menge aller Informationen, die wir über ein Zufallsexperiment bekommen können, entsprechend lassen sich (a)-(c) interpretieren.

*Beispiel 6.2.* Betrachte auf  $\Omega = \mathbb{R}$  die  $\sigma$ -Algebren:

1.  $\mathcal{A} = \{\emptyset, (-\infty, 0), [0, \infty), \mathbb{R}\} \rightsquigarrow$  Wir wissen nur, ob das Ergebnis  $<$  oder  $\geq 0$  ist,
2.  $\mathcal{A} = \mathcal{P}(\mathbb{R}) \rightsquigarrow$  Wir wissen alles, insb. können wir alle  $\omega \in \Omega$  unterscheiden.
3.  $\mathcal{A} = \mathcal{B}_{\mathbb{R}}$  sei die kleinste  $\sigma$ -Algebra, welche alle Intervalle  $(a, b]$  für beliebige  $-\infty < a < b < \infty$  enthält  $\rightsquigarrow$  Für jedes Intervall wissen wir, ob ein Ergebnis darin liegt oder nicht.

◇

►  $\mathcal{B}_{\mathbb{R}}$  heißt **Borel- $\sigma$ -Algebra** auf  $\mathbb{R}$ . Man kann zeigen, dass  $\mathcal{B}_{\mathbb{R}}$  auch alle offenen und alle abgeschlossenen Teilmengen von  $\mathbb{R}$  enthält. Obwohl alle „üblichen“ Untermengen von  $\mathbb{R}$  in  $\mathcal{B}_{\mathbb{R}}$  liegen, gilt  $\mathcal{B}_{\mathbb{R}} \neq \mathcal{P}(\mathbb{R})$ .

Auf  $\Omega = \mathbb{R}$  werden wir typischerweise  $\mathcal{B}_{\mathbb{R}}$  betrachten.

Auf eine  $\sigma$ -Algebra  $\mathcal{A}$  eingeschränkt können wir die Definition eines diskreten W'maßes auf allgemeine Grundräume  $\Omega$  übertragen:

**Definition 6.3.** Sei  $(\Omega, \mathcal{A})$  ein Messraum mit Grundraum  $\Omega \neq \emptyset$  und  $\sigma$ -Algebra  $\mathcal{A}$ . Eine Abbildung

$$P: \mathcal{A} \rightarrow [0, 1]$$

heißt **Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$** , falls

- (a)  $P(\Omega) = 1$ ,

(b)  $A_n \in \mathcal{A}, n \in \mathbb{N}$ , disjunkt  $\Rightarrow P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n)$  ( $\sigma$ -Additivität).

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

► Diskrete Wahrscheinlichkeitsräume ergeben sich als Spezialfälle mit  $\mathcal{A} = \mathcal{P}(\Omega)$  und W'maßen mit abzählbarem Träger.

Die Sätze und Definitionen

1.4 elementare Rechenregeln für endliche W'räume,

1.11 Rechenregeln für diskrete W'maße,

2.1 Bedingte Wahrscheinlichkeiten

2.3 Satz von der Totalen Wahrscheinlichkeit und von Bayes,

2.6 stochastische Unabhängigkeit von Ereignissen,

übertragen sich sinngemäß auf allgemeine Wahrscheinlichkeitsräume, wobei als Ereignisse jeweils nur Mengen aus  $\mathcal{A}$  betrachtet werden.

► Wesentlicher Unterschied zu diskreten Wahrscheinlichkeitsräumen:

Während diskrete W'maße  $P$  vollständig durch die Zähldichte  $f(\omega) := P(\{\omega\})$ ,  $\omega \in \Omega$  bestimmt sind, ist dies für allgemeine W'maße falsch!

Für eine *einfache* Beschreibung von W'maßen auf  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  kann man zeigen, dass es reicht,  $P((-\infty, x])$  für alle  $x \in \mathbb{R}$  festzulegen.

**Satz und Definition 6.4.** *Ist  $P$  ein Wahrscheinlichkeitsmaß auf  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ , so gilt für die durch*

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

definierte **Verteilungsfunktion** von  $P$ :

(i)  $F$  ist monoton steigend.

(ii)  $F$  ist rechtsseitig stetig, d.h.  $\lim_{t \downarrow x} F(t) = F(x)$  für alle  $x \in \mathbb{R}$ .

(iii)  $F(\infty) := \lim_{x \rightarrow \infty} F(x) = 1$ ,  $F(-\infty) := \lim_{x \rightarrow -\infty} F(x) = 0$ .

Ist umgekehrt  $F: \mathbb{R} \rightarrow [0, 1]$  eine Funktion, die (i) - (iii) erfüllt, so existiert genau ein Wahrscheinlichkeitsmaß auf  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ , das  $F$  als Verteilungsfunktion besitzt.

Der Beweis der Rückrichtung beruht auf Resultaten der Maßtheorie, die deutlich über diese Vorlesung hinausgehen. Die Eigenschaften von Verteilungsfunktionen kann man jedoch relativ leicht aus den Eigenschaften von W'maßen herleiten. Zunächst wollen wir jedoch einige Beispiele studieren. Hierzu kann man viele Verteilungen auf  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  besonders einfach durch ihre Dichten beschreiben:

**Satz und Definition 6.5.** *Sei  $P$  ein Wahrscheinlichkeitsmaß auf  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ . Existiert eine integrierbare Funktion  $f: \mathbb{R} \rightarrow [0, \infty)$ , sodass*

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

so heißt  $f$  **Dichte** von  $P$  bzw. von der zugehörigen Verteilungsfunktion  $F$ . Ist  $A \in \mathcal{B}_{\mathbb{R}}$ , so dass  $f \cdot \mathbb{1}_A$  integrierbar ist, so gilt

$$P(A) = \int_{\mathbb{R}} f(x) \cdot \mathbb{1}_A(x) dx =: \int_A f(x) dx$$

Umgekehrt ist jede integrierbare Funktion  $f: \mathbb{R} \rightarrow [0, \infty)$  mit  $\int_{-\infty}^{\infty} f(t) dt = 1$  Dichte eines Wahrscheinlichkeitsmaßes auf  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ , das durch (\*) eindeutig festgelegt ist.

Ein erstes Beispiel für  $W$ -Verteilungen auf  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  ist uns bereits im vorangegangenen Kapitel begegnet:

*Beispiel 6.6* (Normalverteilung). Die aus dem Zentralen Grenzwertsatz bekannte Funktion

$$\varphi(t) := \frac{1}{\sqrt{2\pi}} \cdot e^{-t^2/2}, \quad t \in \mathbb{R},$$

ist die Dichte der sog. **Standardnormalverteilung**  $\mathcal{N}_{(0,1)}$ , denn die zugehörige Verteilungsfunktion  $\Phi(x) = \int_{-\infty}^x \varphi(t) dt$  erfüllt  $\Phi(\infty) = 1$ .

Durch Substitution  $s = (t - \mu)/\sigma$  kann man leicht zeigen, dass auch

$$\varphi_{\mu,\sigma}(t) := \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-(t-\mu)^2/(2\sigma^2)}, \quad t \in \mathbb{R},$$

für Parameter  $\mu \in \mathbb{R}$  und  $\sigma > 0$  eine Dichte ist. Die zugehörige Verteilung ist die **Normalverteilung**  $\mathcal{N}_{(\mu,\sigma^2)}$ .  $\diamond$

*Beispiel 6.7* (Gleichverteilung).

Für  $-\infty < a < b < \infty$  ist  $f := \frac{1}{b-a} \mathbb{1}_{(a,b)}$  Dichte eines  $W$ -maßes auf  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ , denn

$$\int_{-\infty}^{\infty} f(t) dt = \int_{-\infty}^{\infty} \frac{\mathbb{1}_{(a,b)}(t)}{b-a} dt = \frac{1}{b-a} \int_a^b 1 dt = \frac{b-a}{b-a} = 1,$$

der sog. **Gleichverteilung**  $\mathcal{U}_{(a,b)}$  auf  $(a,b]$ . Es gilt für alle  $a \leq c < d \leq b$

$$\begin{aligned} \mathcal{U}_{(a,b]}((c,d]) &= \int_{-\infty}^{\infty} f(x) \cdot \mathbb{1}_{(c,d]}(x) dx \\ &= \int_{-\infty}^{\infty} \frac{\mathbb{1}_{(a,b]}(x)}{b-a} \cdot \mathbb{1}_{(c,d]}(x) dx = \frac{d-c}{b-a}. \end{aligned}$$

► Die  $W$ -keit von  $(c,d]$  hängt nur von der Länge des Intervalls ab, nicht von seiner Lage innerhalb von  $(a,b]$ , daher der Name „Gleichverteilung“.

► Man kann zeigen, dass kein Wahrscheinlichkeitsmaß  $P$  auf  $([0,1], \mathcal{P}([0,1]))$  existiert, so dass

$$P((c,d]) = d - c \quad \forall 0 \leq c < d \leq 1.$$

Es gibt keine Gleichverteilung auf  $[0,1]$ , die jeder Untermenge von  $[0,1]$  eine  $W$ -keit zuweist  $\rightsquigarrow$  Einschränkung auf  $\mathcal{B}_{\mathbb{R}}$ .  $\diamond$

*Beispiel 6.8* (Exponentialverteilung).

Für alle  $\lambda > 0$  ist  $f_{\lambda}(x) := \frac{1}{\lambda} e^{-x/\lambda} \cdot \mathbb{1}_{[0,\infty)}(x)$  Dichte eines  $W$ -maßes auf  $\mathbb{R}$ , denn

$$\int_{-\infty}^{\infty} f_{\lambda}(x) dx = \int_0^{\infty} \frac{1}{\lambda} e^{-x/\lambda} dx = -e^{-x/\lambda} \Big|_0^{\infty} = 1.$$

Das zugehörige  $W$ -maß  $\mathcal{E}_{\lambda}$  heißt **Exponentialverteilung** mit Parameter  $\lambda > 0$ . Es besitzt die Verteilungsfunktion

$$F_{\lambda}(x) = \int_{-\infty}^x f_{\lambda}(t) dt = -e^{-t/\lambda} \Big|_0^x = 1 - e^{-x/\lambda}, \quad \forall x \geq 0,$$

und  $F_{\lambda}(x) = 0$  für alle  $x < 0$ .  $\diamond$

Für  $x, y \geq 0$  ist die bedingte Wahrscheinlichkeit, dass ein Wert größer als  $x + y$  beobachtet wird, wenn bekannt ist, dass der Wert  $x$  überschreitet, gerade gleich

$$\begin{aligned} \mathcal{E}_\lambda((x + y, \infty) \mid (x, \infty)) &= \frac{\mathcal{E}_\lambda((x + y, \infty) \cap (x, \infty))}{\mathcal{E}_\lambda((x, \infty))} = \frac{\mathcal{E}_\lambda((x + y, \infty))}{\mathcal{E}_\lambda((x, \infty))} \\ &= \frac{1 - \mathcal{E}_\lambda((-\infty, x + y])}{1 - \mathcal{E}_\lambda((-\infty, x])} \\ &= \frac{1 - F_\lambda(x + y)}{1 - F_\lambda(x)} = \frac{e^{-(x+y)/\lambda}}{e^{-x/\lambda}} \\ &= e^{-y/\lambda} = \mathcal{E}_\lambda((y, \infty)). \end{aligned}$$

► Exponentialverteilungen beschreiben Lebensdauern von Dingen, die nicht altern, d.h. die W'keit noch  $y$  Jahre zu überleben, gegeben dass bereits  $x$  Jahre überlebt wurden, hängt nicht von  $x$  ab. Da das bis  $x$  erlebte keinen Einfluss hat, sprechen wir auch von *Gedächtnislosigkeit*.

*Beispiel 6.9.*

Ein Seil der Länge 1 wird so lange an beiden Enden gezogen, bis es reißt. Die Wahrscheinlichkeit, dass das Seil in der Umgebung der Stelle  $x \in (0, 1)$  reißt, sei proportional zum Abstand zum näher gelegenen Ende, also proportional zu  $\min(x, 1 - x)$ .

*Modell:*  $P$  sei das Wahrscheinlichkeitsmaß mit der Dichte

$$f(x) = \begin{cases} c \cdot \min(x, 1 - x), & x \in (0, 1), \\ 0, & \text{sonst.} \end{cases}$$

Wähle  $c > 0$  so, dass  $\int_{-\infty}^{\infty} f(x) dx \stackrel{!}{=} 1$ , das heißt

$$c \int_0^1 \min(x, 1 - x) dx = 2c \int_0^{1/2} x dx = cx^2 \Big|_0^{1/2} = \frac{c}{4} \stackrel{!}{=} 1.$$

Es folgt  $c = 4$ , also  $f(x) = 4 \min(x, 1 - x) \mathbb{1}_{(0,1)}(x)$ .

Die Wahrscheinlichkeit, dass das längere Teilstück mindestens die Länge  $3/4$  besitzt, beträgt

$$\begin{aligned} P\left(\left(0, \frac{1}{4}\right] \cup \left[\frac{3}{4}, 1\right)\right) &= \int_{-\infty}^{\infty} \mathbb{1}_{(0, \frac{1}{4}] \cup [\frac{3}{4}, 1)}(x) \cdot f(x) dx \\ &= \int_0^{1/4} 4x dx + \int_{3/4}^1 4(1 - x) dx \\ &= 4x^2 \Big|_0^{1/4} \\ &= \frac{1}{4}. \end{aligned} \quad \diamond$$

Wir kommen nun zurück zum Beweis von Satz 6.4. Für den Nachweis der Eigenschaften (ii) und (iii) einer Verteilungsfunktion benötigen wir folgende Eigenschaften von W'maßen:

**Lemma 6.10.** *Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum.*

(i) *Sei  $A_n \in \mathcal{A}$ ,  $n \in \mathbb{N}$ , eine monoton steigende Folge (d.h.  $A_n \subseteq A_{n+1} \quad \forall n \in \mathbb{N}$ ). Dann gilt für  $A := \bigcup_{n \in \mathbb{N}} A_n$  (kurz:  $A_n \uparrow A$ )*

$$P(A) = \lim_{n \rightarrow \infty} P(A_n).$$

(ii) Ist  $B_n \in \mathcal{A}$ ,  $n \in \mathbb{N}$ , eine monoton fallende Folge (d.h.  $B_n \supset B_{n+1} \quad \forall n \in \mathbb{N}$ ) und  $B := \bigcap_{n \in \mathbb{N}} B_n$  (kurz:  $B_n \downarrow B$ ), so gilt

$$P(B) = \lim_{n \rightarrow \infty} P(B_n).$$

Den Beweis dieses Lemmas und der Eigenschaften (i) bis (iii) einer Verteilungsfunktion laut Satz 6.4 führen wir, falls es die Zeit erlaubt.

Eine weitere Anwendung des Lemmas liefert:

**Korollar 6.11.** Für jedes Wahrscheinlichkeitsmaß  $Q$  auf  $(\mathbb{R}, \mathcal{B})$  mit Verteilungsfunktion  $F$  und für alle  $x \in \mathbb{R}$  gilt für die Sprunghöhe von  $F$  in  $x$

$$\Delta F(x) := F(x) - F(x-) := F(x) - \lim_{t \uparrow x} F(t) = Q(\{x\}).$$

► Für jede stetige Verteilungsfunktion gilt  $P(\{x\}) = 0$  für alle  $x \in \mathbb{R}$ . Dies ist insbesondere der Fall, wenn  $P$  eine Dichte besitzt.

Dies zeigt, dass Zähldichten für solche Verteilungen völlig bedeutungslos sind.

## 6.2 Zufallsvariablen und ihre Erwartungswerte

Bei diskreten Wahrscheinlichkeitsräumen  $(\Omega, P)$  ist eine Abbildung  $X: \Omega \rightarrow S$  für  $S \neq \emptyset$  stets eine Zufallsvariable und besitzt somit eine Verteilung:

$$P^X(B) = P(X^{-1}(B)) = P(X \in B) \quad \forall B \subseteq S.$$

Dies ist dann ein diskretes Wahrscheinlichkeitsmaß auf  $S$ .

Im allgemeinen Fall ist ein Wahrscheinlichkeitsmaß  $P$  nur noch auf einem Mengensystem  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  definiert, d.h.  $P: \mathcal{A} \rightarrow [0, 1]$ .

►  $P(X^{-1}(B))$  ist nur noch definiert, wenn  $X^{-1}(B) \in \mathcal{A}$ , d.h. wenn das Urbild eine „gutartige“ Menge ist.

► Wir schränken daher wieder die betrachteten Teilmengen  $B \subseteq S$  auf eine Untermenge von  $\mathcal{P}(S)$  ein.

**Satz und Definition 6.12.** Es seien  $(\Omega, \mathcal{A})$ ,  $(S, \mathcal{C})$  Messräume und  $X: \Omega \rightarrow S$  eine Abbildung.  $X$  heißt  $S$ -wertige **Zufallsvariable**, falls

$$X^{-1}(C) \in \mathcal{A} \quad \forall C \in \mathcal{C}$$

Man schreibt  $X: (\Omega, \mathcal{A}) \rightarrow (S, \mathcal{C})$  und sagt, dass  $X$   $(\mathcal{A}, \mathcal{C})$ -**messbar** ist.

Sei  $P$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$ . Dann wird durch

$$P^X(C) := P(X^{-1}(C)) = P(X \in C), \quad C \in \mathcal{C},$$

ein Wahrscheinlichkeitsmaß  $P^X$  auf  $(S, \mathcal{C})$  definiert, die sogenannte **Verteilung von  $X$  unter  $P$** .

**Definition 6.13.** Sei  $X: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  eine  $\mathbb{R}$ -wertige Zufallsvariable auf dem W'raum  $(\Omega, \mathcal{A}, P)$ .

(i) Die Verteilungsfunktion der Verteilung  $P^X$  von  $X$

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

wird auch **Verteilungsfunktion von  $X$**  genannt.

(ii) Besitzt  $F$  eine Dichte  $f$ , d.h. gilt

$$P(X \leq x) = \int_{-\infty}^x f(t) dt \quad \forall x \in \mathbb{R},$$

so heißt diese auch **Dichte von  $X$** .

Im diskreten Wahrscheinlichkeitsraum wird der Erwartungswert definiert als

$$E[X] = \sum_{x \in X(\Omega)} x \cdot P(X = x)$$

Diese Definition ist für allgemeine  $\mathbb{R}$ -wertige Zufallsvariablen unsinnig!

Nach Korollar 6.11 gilt

$$P(X = x) = P^X(\{x\}) = 0,$$

falls  $X$  eine stetige Verteilungsfunktion besitzt und insbesondere, falls  $X$  eine Dichte besitzt.

► Bei obiger Definition wäre der Erwartungswert auch dann 0, wenn die Zufallsvariable nur strikt positive Werte annimmt.

► Man ersetzt daher bei Zufallsvariablen mit Dichten in der obigen Definition die Summe durch ein geeignetes Integral.

**Definition 6.14.** Ist  $X$  eine  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ -wertige Zufallsvariable, die eine Dichte  $f$  besitzt, so wird der **Erwartungswert von  $X$**  definiert als

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx,$$

falls  $\int_{-\infty}^{\infty} |x|f(x) dx < \infty$ .

Analog zu Definition 4.1 nennt man den Erwartungswert von  $X$  auch den Mittelwert der Verteilung  $P^X$  von  $X$ .

*Beispiel 6.15.* 1. Sei  $X$  gleichverteilt auf  $(a, b]$ , d.h. mit Dichte  $f = \mathbb{1}_{(a,b]}/(b-a)$ . Dann hat  $X$  den Erwartungswert

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

2. Sei  $X$  exponentialverteilt mit Parameter  $\lambda$ , d.h.  $X$  besitze die Dichte  $f_{\lambda}(x) := e^{-x/\lambda}/\lambda \cdot \mathbb{1}_{[0,\infty)}(x)$ .

Dann berechnet sich der Erwartungswert mittels partieller Integration zu

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f_{\lambda}(x) dx \\ &= \frac{1}{\lambda} \int_0^{\infty} x e^{-x/\lambda} dx \\ &= -x e^{-x/\lambda} \Big|_0^{\infty} + \int_0^{\infty} e^{-x/\lambda} dx \\ &= -\lambda e^{-x/\lambda} \Big|_0^{\infty} \\ &= \lambda. \end{aligned}$$

◇

**Satz 6.16.** *Ist in der Situation von Definition 6.14  $g: \mathbb{R} \rightarrow \mathbb{R}$  eine Abbildung, so dass  $g(X)$  eine Zufallsvariable ist, deren Erwartungswert existiert, so gilt*

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x) dx.$$

Mit diesem Ergebnis lassen sich Varianz, Kovarianz und Korrelation (und Median) analog zu Kapitel 4 (s. 4.11 und 4.18) definieren und berechnen.

Ferner gelten die gleichen Rechenregeln, wie dort hergeleitet.

*Beispiel 6.17.*

Sei  $X$  gleichverteilt auf  $[a, b]$ , d.h. mit Dichte  $f = \mathbb{1}_{[a,b]}/(b-a)$ . Dann hat  $X^2$  den Erwartungswert

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}.$$

Daher folgt für die Varianz

$$\begin{aligned} \text{Var}(X) &= E[X^2] - E[X]^2 \\ &= \frac{b^2 + ab + a^2}{3} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{4(b^2 + ab + a^2) - 3(a^2 + 2ab + b^2)}{12} \\ &= \frac{b^2 - 2ab + a^2}{12} \\ &= \frac{(b-a)^2}{12}. \end{aligned}$$

◇

### 6.3 Unabhängigkeit

Man kann nun die Unabhängigkeit von Zufallsvariablen genauso wie in Definition 3.4 einführen. Insbesondere definieren wir für reellwertige ZVN:

**Definition 6.18.** Zufallsvariablen  $X_1, \dots, X_n: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  heißen **stochastisch unabhängig**, wenn die Ereignisse  $\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$  für alle  $B_1, \dots, B_n \in \mathcal{B}_{\mathbb{R}}$  stochastisch unabhängig sind.

Diskrete Zufallsvariablen sind genau dann unabhängig, wenn ihre gemeinsame Zähldichte  $(x_1, \dots, x_n) \mapsto f(x_1, \dots, x_n)$  gleich dem Produkt ihrer einzelnen Zähldichten  $f_{X_i}(x_i)$  ist (s. Satz 3.5).

Die Verteilung einer allgemeinen Zufallsvariablen lässt sich nicht durch eine Zähldichte beschreiben. Es lässt sich jedoch für reellwertige Zufallsvariablen eine analoge Charakterisierung mit Hilfe ihrer Verteilungsfunktionen angeben.

**Satz 6.19.** *Für Zufallsvariablen  $X_i: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ ,  $1 \leq i \leq n$ , mit Verteilungsfunktionen  $F_{X_i}$ ,  $1 \leq i \leq n$ , sind äquivalent:*

(i)  $X_1, \dots, X_n$  sind stochastisch unabhängig.

(ii) 
$$P(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n P(X_i \in B_i) \quad \forall B_1, \dots, B_n \in \mathcal{B}_{\mathbb{R}}.$$

(iii) 
$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i) = \prod_{i=1}^n F_{X_i}(x_i) \quad \forall x_1, \dots, x_n \in \mathbb{R}.$$

► Zufallsvariablen sind genau dann unabhängig, wenn die Verteilungsfunktion ihrer gemeinsamen Verteilung das Produkt ihrer Verteilungsfunktionen ist.

*Beispiel 6.20.*

- (i) Seien  $X_1, X_2$  stochastisch unabhängige exponential  $\mathcal{E}_1$ -verteilte Zufallsvariablen mit Vf.  $F_{X_1}(x) = F_{X_2}(x) = (1 - e^{-x})1_{(0, \infty)}(x)$  für alle  $x \in \mathbb{R}$ .

Dann gilt für alle  $x_1, x_2 \in \mathbb{R}$

$$\begin{aligned} P(X_1 \leq x_1, X_2 \leq x_2) &= P(X_1 \leq x_1) \cdot P(X_2 \leq x_2) \\ &= (1 - e^{-x_1})(1 - e^{-x_2})1_{(0, \infty)^2}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2). \end{aligned}$$

- (ii) Seien nun  $Y_1, Y_2$   $\mathbb{R}$ -wertige Zufallsvariablen, so dass für alle  $y_1, y_2 \in \mathbb{R}$ :

$$P(Y_1 \leq y_1, Y_2 \leq y_2) = (1 - e^{\min(y_1, y_2)})1_{(0, \infty)^2}(y_1, y_2). \quad (*)$$

Dann folgt, dass  $Y_1$  (und analog  $Y_2$ )  $\mathcal{E}_1$ -verteilt ist, denn  $\forall y_1 \in \mathbb{R}$ :

$$F_{Y_1}(y_1) = P\{Y_1 \leq y_1\} = \lim_{y_2 \rightarrow \infty} P\{Y_1 \leq y_1, Y_2 \leq y_2\} = (1 - e^{-y_1})1_{(0, \infty)}(y_1).$$

Da die rechte Seite von (\*) nicht als Produkt einer nur von  $y_1$  und einer nur von  $y_2$  abhängende Funktion geschrieben werden kann, sind  $Y_1$  und  $Y_2$  jedoch nicht stochastisch unabhängig. Tatsächlich ist  $(Y_1, Y_2)$  verteilt wie  $(X_1, X_1)$  (Warum?).  $\diamond$

Borel- $\sigma$ -Algebren, Verteilungsfunktionen und Dichten kann man analog auch für  $\Omega = \mathbb{R}^n$  definieren.

**Definition 6.21.** (i) Die **Borel- $\sigma$ -Algebra  $\mathcal{B}_{\mathbb{R}^n}$  auf  $\mathbb{R}^n$**  wird definiert als die kleinste  $\sigma$ -Algebra, die alle Rechteckmengen

$$\bigtimes_{i=1}^n (a_i, b_i] \quad \text{mit} \quad \infty < a_i < b_i < \infty \quad \text{für alle} \quad 1 \leq i \leq n$$

enthält.

- (ii) Ist  $P$  ein Wahrscheinlichkeitsmaß auf  $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ , so wird die zugehörige **multivariate Verteilungsfunktion  $F$**  definiert durch

$$F(x_1, \dots, x_n) := P\left(\bigtimes_{i=1}^n (-\infty, x_i]\right), \quad (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Übertragen wir multivariate Verteilungsfunktionen auf Zufallsvariablen, erhalten wir:

**Satz und Definition 6.22.** Ist  $(\Omega, \mathcal{A})$  ein Messraum und  $X_i : \Omega \rightarrow \mathbb{R}$ ,  $1 \leq i \leq n$ , so gilt die Äquivalenz

$$\begin{aligned} X = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n \quad \mathcal{A}, \mathcal{B}^n\text{-messbar} \\ \iff X_i : \Omega \rightarrow \mathbb{R} \quad \mathcal{A}, \mathcal{B}_{\mathbb{R}}\text{-messbar} \quad \forall 1 \leq i \leq n. \end{aligned}$$

In dem Fall wird  $X$  auch  **$n$ -dimensionaler Zufallsvektor** genannt. Die multivariate Verteilungsfunktion von  $P^X$

$$F(x_1, \dots, x_n) := P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n), \quad (x_1, \dots, x_n) \in \mathbb{R}^n,$$

heißt **gemeinsame Verteilungsfunktion von  $X_1, \dots, X_n$** .

Analog zum eindimensionalen Fall führen wir Dichten auf  $\mathbb{R}^n$  ein.

**Satz und Definition 6.23.** Es sei  $P$  ein  $W$ -maß auf  $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ , mit multivariater Verteilungsfunktion  $F$ . Existiert eine Abbildung  $f: \mathbb{R}^n \rightarrow [0, \infty)$ , so dass für alle  $(x_1, \dots, x_n) \in \mathbb{R}^n$

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_n \dots dt_2 dt_1$$

gilt, so heißt  $f$  (multivariate) **Dichte** von  $P$  bzw. von  $F$ . Es gilt dann  $\forall B \in \mathcal{B}_{\mathbb{R}^n}$

$$P(B) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbb{1}_B(t_1, \dots, t_n) \cdot f(t_1, \dots, t_n) dt_n \dots dt_1.$$

► Insbesondere gilt für  $B = (a_1, b_1] \times \cdots \times (a_n, b_n]$ ,  $a_i < b_i, i = 1, \dots, n$ :

$$P((a_1, b_1] \times \cdots \times (a_n, b_n]) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(t_1, \dots, t_n) dt_n \dots dt_1.$$

Ist  $X = (X_1, \dots, X_n)$  ein Zufallsvektor auf  $(\Omega, \mathcal{A}, P)$ , so heißen die Verteilungen  $P^{X_i}, 1 \leq i \leq n$ , (eindimensionale) **Randverteilungen**.

► Die Verteilungsfunktion  $F_i$  von  $X_i$  bzw.  $P^{X_i}$  berechnet sich wie folgt:

$$\begin{aligned} F_i(x) &= P(X_1 < \infty, \dots, X_{i-1} < \infty, X_i \leq x, X_{i+1} < \infty, \dots, X_n < \infty) \\ &= F(\infty, \dots, \infty, \underbrace{x}_i, \infty, \dots, \infty). \end{aligned}$$

**Satz 6.24.** Besitzt die  $\mathbb{R}^n$ -wertige Zufallsvariable  $X = (X_1, \dots, X_n)$  eine gemeinsame Dichte  $f: \mathbb{R}^n \rightarrow [0, \infty)$ , so hat  $X_i$  für jedes  $1 \leq i \leq n$  die Dichte  $f_i: \mathbb{R} \rightarrow [0, \infty)$  mit

$$f_i(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, \dots, t_{i-1}, x, t_{i+1}, \dots, t_n) dt_n \cdots dt_{i+1} dt_{i+1} \cdots dt_1$$

für alle  $x \in \mathbb{R}$ .

Falls die Verteilungen der Zufallsvariablen Dichten besitzen, so kann man mit diesen auch die Unabhängigkeit charakterisieren, analog zum diskreten Fall.

**Satz 6.25.** Seien  $X_i: \Omega \rightarrow \mathbb{R}$  für  $i = 1, \dots, n$  ZVn auf einem  $W$ -raum  $(\Omega, \mathcal{A}, P)$ . Falls alle Randverteilungen  $P^{X_i}, 1 \leq i \leq n$  jeweils eine Dichte  $f_i$  besitzen, dann sind äquivalent.

- (i)  $X_1, \dots, X_n$  sind unabhängig unabhängig.
- (ii)  $P^{(X_1, \dots, X_n)}$  besitzt eine Dichte  $f$  gegeben durch

$$f(x_1, \dots, x_n) := \prod_{i=1}^n f_i(x_i), \quad (x_1, \dots, x_n) \in \mathbb{R}^n.$$

*Bemerkung 6.26.*

- Durch die Verteilung von  $X = (X_1, \dots, X_n)$  sind auch alle Randverteilungen eindeutig bestimmt; Verteilungsfunktionen und (im Fall der Existenz) Dichten der Randverteilungen lassen sich (prinzipiell) leicht berechnen.
- Umgekehrt ist die Verteilung von  $X$  i.A. *nicht* durch die Randverteilungen eindeutig festgelegt.
- Sind  $X_1, \dots, X_n$  stochastisch unabhängig, so legen bereits die eindimensionalen Randverteilungen  $P^{X_i}$  die Verteilung von  $X$  eindeutig fest.

Es gilt für die Verteilungsfunktionen

$$F_X(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$$

und im Fall der Existenz für die Dichten

$$f_X(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$