

**Stochastik  
für  
Studierende der Informatik**

Vorlesung

an der

Albert–Ludwigs–Universität Freiburg i. Br.

Sommersemester 2008

Prof. Dr. H. R. Lerche

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitende und historische Bemerkungen</b>	<b>1</b>
<b>2</b>	<b>Ereignisse und Wahrscheinlichkeiten</b>	<b>2</b>
2.1	Der Würfel . . . . .	2
2.2	Mengentheoretische Beschreibung von Ereignissen . . . . .	3
2.3	Wahrscheinlichkeiten . . . . .	4
2.4	Chancen-/Gewinnverhältnisse und Auszahlungsquoten . . . . .	6
2.5	Gleichverteilungen . . . . .	7
2.6	Bedingte Wahrscheinlichkeiten und Unabhängigkeit . . . . .	12
2.7	Kombination von Ereignissen . . . . .	21
<b>3</b>	<b>Zufallsvariable und Erwartungswerte</b>	<b>23</b>
3.1	Grundlagen . . . . .	23
3.2	Bernoulli-Verteilung und Binomialverteilung . . . . .	26
3.3	Hypergeometrische Verteilung . . . . .	29
3.4	Poisson-Verteilung und das Gesetz der kleinen Wahrscheinlichkeiten . . . . .	31
3.5	Gesetz der Großen Zahlen . . . . .	34
3.6	Der zentrale Grenzwertsatz . . . . .	39
<b>4</b>	<b>Zufallsvariable und Wahrscheinlichkeitsdichten</b>	<b>43</b>
4.1	Dichten und Verteilungsfunktionen . . . . .	43
4.2	Die Normalverteilung . . . . .	45
4.3	Exponential- und Gamma-Verteilung . . . . .	47
4.4	Lebenszeiten . . . . .	50
4.5	Gemeinsame Verteilung mehrerer Zufallsvariablen . . . . .	51
4.6	Bedingte Verteilungen und Bedingte Erwartungen . . . . .	56
4.7	Empirische Verteilungsfunktion und Quantile . . . . .	61

<b>5</b>	<b>Statistik</b>	<b>65</b>
5.1	Maximum-Likelihood Methode . . . . .	65
5.2	Methode der Momente . . . . .	68
5.3	Konfidenzintervalle . . . . .	69
5.4	Hypothesentests . . . . .	74
5.5	Approximative Konfidenzintervalle und Hypothesentests . . . . .	77
5.6	$\chi^2$ -Anpassungstest . . . . .	79
<b>6</b>	<b>Markov-Ketten</b>	<b>82</b>
6.1	Die Kain und Abel-Aufgabe . . . . .	82
6.2	Markov-Ketten . . . . .	83
6.3	Absorbierende Zustände . . . . .	86
6.4	Rekurrente und transiente Zustände . . . . .	89
6.5	Stationäre Verteilungen . . . . .	91
6.6	Konvergenz gegen die stationäre Verteilung . . . . .	94
6.7	Der Quellencodierungssatz von Shannon . . . . .	96
	<b>Literaturverzeichnis</b>	<b>98</b>

# Kapitel 1

## Einleitende und historische Bemerkungen

Die Wahrscheinlichkeitsrechnung geht in ihren Anfängen auf das Bestimmen von Chancen und Auszahlungen bei Spielen zurück; etwa um 1480 gibt es dazu erste Zeugnisse. Das Spielen aber ist so alt wie die Menschheit und der Zufall war wohl schon immer beim Spielen mit dabei. Im Altertum hat man sehr oft mit würfelähnlichen Gebilden gespielt, meist hergestellt aus Knochen von Tieren. Bei den Griechen und Römern hieß ein solcher "Würfel" Astragalus; er wurde aus Ziegenknochen gefertigt. Im Mittelalter kannte man schon die uns heute geläufigen Würfel und natürlich die dazugehörigen Spiele. Eines der frühesten Werke, das sich mit Chancen und Quoten beim Würfelspielen beschäftigt, geht zurück auf Cardano, ca. 1550. Es heißt "Liber de Ludo Alea". Darin finden sich Überlegungen von der Art, daß, wenn ein Würfel nicht "gezinkt" ist, die Wette auf 1, 3, 5 als genauso günstig anzusehen ist, wie die auf 2, 4, 6. Cardano war auch einer der ersten, der Additions- und Multiplikationsgesetze für Wahrscheinlichkeiten formulierte. Später um 1650 berechneten Pascal und auch Huygens die Wahrscheinlichkeiten von Spielergebnissen. Um 1800 gab es bereits statistische Überlegungen in der Astronomie, z. B. bei der Bestimmung von Planetenorten durch Gauss. Im 19. Jahrhundert wurde die Wahrscheinlichkeitsrechnung noch zur Physik gezählt. Aber Hilberts Bemühungen die Gebiete der Mathematik solide zu begründen, führten dazu, daß Kolmogorov 1933 einen axiomatischen Zugang zur Wahrscheinlichkeitstheorie fand, der diese zu einem Teilgebiet der Mathematik machte. Wahrscheinlichkeiten und Statistiken begegnen uns heute an vielen Stellen des Alltags, sei es im Sport, in der Technik und Wissenschaft, in der Medizin, im Banken- und Versicherungswesen.

Wir wissen seit 80 Jahren, daß die Materie im Kleinen Wahrscheinlichkeitsgesetzen gehorcht. In der sogenannten Quantentheorie sind sie entwickelt und vielfach in der Technik und Wissenschaft überprüft worden.

Wozu Stochastik für Informatiker und Mikrosystemtechniker? Für die Ingenieure ist der Umgang mit Meßwerten unumgänglich und führt notgedrungen auf Statistik. Bei den Informatikern ist die Anwendung etwas versteckter. Aber die mittleren Laufzeiten von Algorithmen oder das Lernen aus Datenmengen sowie die Orientierung von Robotern sind Konzepte, die Stochastik brauchen.

Ich will diese einleitenden Bemerkungen schließen mit einem Beispiel, das die Brücke schlägt von Cardano zu den Ereignissen unserer Tage. Am 08.04.2005 sollte ursprünglich die Hochzeit zwischen Prinz Charles und Camilla Parker-Bowles stattfinden. Ende März 2005 stellten die Buchmacher in London die Wette auf eine Verschiebung der Hochzeit mit einer Quote 19:1, d.h. 19 Pfund Gewinn bei 1 Pfund Einsatz. Tatsächlich trat durch den Tod des Papstes das Ereignis ein und die Londoner Wettbüros durften kräftig zahlen.

# Kapitel 2

## Ereignisse und Wahrscheinlichkeiten

### 2.1 Der Würfel

**Ein Würfel werde einmal geworfen:**

Die Menge der möglichen Ergebnisse ist  $\Omega_1 = \{1, 2, \dots, 6\}$ . Ist der Würfel fair, so hat man  $P(\{i\}) = \frac{1}{6}$ ,  $1 \leq i \leq 6$ . Folglich ergibt sich weiter

$$P(\{2, 4, 6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = 3 \cdot \frac{1}{6} = \frac{1}{2}$$

und ebenso  $P(\{1, 3, 5\}) = \frac{1}{2}$  sowie  $P(\Omega_1) = 1$ . Ist  $A \subset \Omega_1$ , so definiert man

$$P(A) := \frac{|A|}{|\Omega_1|}.$$

Dabei ist  $|A|$  die Anzahl der Elemente der Menge  $A$ .

Nun einige Beispiele zu möglichen Ereignissen.

**Beispiele:**

1. Ergebnis ist ungerade  $\Rightarrow \{1, 3, 5\}$ .
2. Ergebnis ist gerade und kleiner als 4  $\Rightarrow \{2, 4, 6\} \cap \{1, 2, 3\} = \{2\}$ .

**Grundlegende Bemerkung**

Ereignisse, die durch eine logische Aussage bestimmt werden, beschreiben wir stets durch die Menge aller Ergebnisse, die diese logische Aussage erfüllen.

**Ein Würfel werde zweimal geworfen:**

Ein Ergebnis ist z.B.  $(1, 3)$ . Beim 1. Wurf kommt eine 1, beim 2. Wurf kommt eine 3. Hier ist die Menge der möglichen Ereignisse

$$\Omega_2 = \{(1, 1), (1, 2), \dots, (6, 6)\} = \{(i, j) | 1 \leq i \leq 6, 1 \leq j \leq 6\}, \quad |\Omega_2| = 36.$$

Für  $A \subset \Omega_2$  definiert man  $P(A) := \frac{|A|}{|\Omega_2|} = \frac{|A|}{36}$ .

**Beispiel:**

Was ist die Wahrscheinlichkeit, daß die Summe zweier Würfe  $\leq 10$  ist?

$$\begin{aligned} P(\text{Summe} \leq 10) &= 1 - P(\text{Summe} > 10) \\ &= 1 - P(\text{Summe} \geq 11) \\ &= 1 - P(B_{10}^c) = 1 - \frac{|B_{10}^c|}{|\Omega_2|} = 1 - \frac{3}{36} = \frac{11}{12}. \end{aligned}$$

Dabei sind  $B_{10} = \{(1, 1), \dots, (4, 6)\}$  und  $B_{10}^c = \{(5, 6), (6, 5), (6, 6)\}$ .

**Beispiel:**

Was ist die Wahrscheinlichkeit, daß das Ergebnis des 2. Wurfs größer ist als das des 1.? Sei  $C = \{(i, j) | i < j, 1 \leq i \leq 6, 1 \leq j \leq 6\}$ . Dann ist

$$P(\text{2. Wurf} > \text{1. Wurf}) = P(C) = \frac{6(6-1)}{6^2} = \frac{5}{12}.$$

## 2.2 Mengentheoretische Beschreibung von Ereignissen

Sei  $\Omega$  die Menge aller möglichen Ausgänge  $\omega$  eines Zufallsexperiments. Ein Ereignis wird durch eine logische Aussage festgelegt. Dazu gehört genau eine Menge  $A \subset \Omega$ . Wir identifizieren von nun an Ereignisse mit Teilmengen von  $\Omega$ , nämlich genau mit den Mengen, deren Elemente die logische Aussage erfüllen. Das heißt, was eintritt bzw. nicht eintritt, beschreiben wir durch Mengen.

### 2.2.1 Bezeichnungen

- Grundraum  $\Omega$ : sicheres Ereignis
- Leere Menge  $\emptyset$ : unmögliches Ereignis
- $A \subset B$ :  $A$  liegt in  $B$ ; aus  $A$  folgt  $B$
- $B \setminus A$ :  $B$  ohne  $A$ ;  $\omega \in B \setminus A$  genau dann wenn  $\omega \in B$  und  $\omega \notin A$
- $A^c$ : Komplement von  $A$ , Gegenereignis;  $\omega \in A^c$  genau dann wenn  $\omega \notin A$
- $A \cap B$ :  $A$  und  $B$ ;  $\omega \in A \cap B$  genau dann wenn  $\omega \in A$  und  $\omega \in B$
- $A \cup B$ :  $A$  oder  $B$ ;  $\omega \in A \cup B$  genau dann wenn  $\omega \in A$  oder  $\omega \in B$   
(läßt  $\omega \in A \cap B$  zu!)
- $\mathcal{P}(\Omega)$ : Potenzmenge von  $\Omega$ ; das System aller Teilmengen von  $\Omega$ .

**Wichtige Rechenregeln:**

- Kommutativgesetz:  $A \cap B = B \cap A, A \cup B = B \cup A$
- Assoziativgesetz:  $A \cap (B \cap C) = (A \cap B) \cap C,$   
 $A \cup (B \cup C) = (A \cup B) \cup C$
- Distributivgesetz:  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$   
 $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- Gesetze von de Morgan:  $(A \cup B)^c = A^c \cap B^c, (A \cap B)^c = A^c \cup B^c$

Endliche Vereinigungen und Schnitte. Seien  $A_1, \dots, A_n$  Ereignisse

$$\bigcup_{i=1}^n A_i = A_1 \cup \dots \cup A_n \quad \text{und} \quad \bigcap_{i=1}^n A_i = A_1 \cap \dots \cap A_n$$

Abzählbare Vereinigungen und Schnitte. Seien  $A_1, A_2, \dots$  Ereignisse

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \dots \quad \text{und} \quad \bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \dots$$

Gesetze von de Morgan gelten auch hier:

$$\left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c \quad \text{und} \quad \left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c$$

**Sprechweise:** Zwei Ereignisse  $A$  und  $B$  mit  $A \cap B = \emptyset$  heißen disjunkt. Ist  $A_1, A_2, \dots$  eine endliche oder abzählbare Folge von Ereignissen mit  $A_i \cap A_j = \emptyset$  für  $i \neq j$ , so heißen diese Ereignisse paarweise disjunkt.

## 2.3 Wahrscheinlichkeiten

**Definition 2.3.1** Sei  $A \in \mathcal{P}(\Omega)$  und seien  $A_1, A_2, \dots, A_n \in \mathcal{P}(\Omega)$ .  $\{A_1, A_2, \dots, A_n\}$  heißt Partition von  $A$ , falls  $A = \bigcup_{i=1}^n A_i$  und  $A_i \cap A_j = \emptyset$  für  $i \neq j$  gilt.

**Definition 2.3.2** Eine **Wahrscheinlichkeit**  $P$ , ist eine Abbildung  $P : A \mapsto P(A)$  für  $A \subset \Omega$  mit

1.  $P(A) \geq 0$
2.  $P(\Omega) = 1$
3.  $P(A) = \sum_{i=1}^n P(A_i)$ , falls  $\{A_1, \dots, A_n\}$  Partition von  $A$  ist.

**Definition 2.3.3** Sei  $\Omega$  eine endliche oder abzählbar unendliche Menge. Eine Abbildung  $P$  von  $\mathcal{P}(\Omega)$  nach  $\mathbb{R}$  heißt **Wahrscheinlichkeitsmaß** auf  $\mathcal{P}(\Omega)$ , falls gilt

1.  $P(A) \geq 0$  für  $A \in \mathcal{P}(\Omega)$
2.  $P(\Omega) = 1$
3. für paarweise disjunkte Ereignisse  $A_1, A_2, \dots$  aus  $\mathcal{P}(\Omega)$  ist

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

## Rechenregeln:

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

Begründung:  $A_i = \emptyset$  für  $i \geq 1$  ist paarweise disjunkte Folge. Damit  $P(\emptyset) = P(\emptyset) + P(\emptyset) + \dots$

ii)  $P(A \cup B) = P(A) + P(B)$  falls  $A$  und  $B$  disjunkt sind.

Begründung:  $A \cup B = A \cup B \cup \emptyset \cup \emptyset \cup \dots$ . Damit ist  $P(A \cup B) = P(A) + P(B) + 0 + 0 + \dots = P(A) + P(B)$ .

iii)  $P(A^c) = 1 - P(A)$ .

Begründung: folgt aus ii) mit  $\Omega = A \cup A^c$ .

iv) Ist  $A \subset B$ , so gilt  $P(B \setminus A) = P(B) - P(A)$ .

Begründung: folgt aus ii) mit  $B = A \cup (B \setminus A)$  und  $A \cap (B \setminus A) = \emptyset$ .

v) Ist  $A \subset B$ , so gilt  $P(A) \leq P(B)$ .

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

Begründung:  $A \cup B = A \cup (B \setminus (A \cap B))$ .

Für die weiteren Überlegungen ist die folgende Gleichung nützlich:

$$P(A) = \sum_{\omega \in A} P(\{\omega\})$$

Sie ist eine direkte Folge der Definition einer Wahrscheinlichkeit.

Weitergehend hat man

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Dies sieht man durch Fallunterscheidung, wenn man alle Terme als Summen gemäß der vorangegangenen Gleichung liest und zählt wie oft die einzelnen Terme jeweils auftauchen. Entsprechend argumentierend erhält man die Ungleichungen

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &\leq \sum_{i=1}^n P(A_i) \\ P\left(\bigcup_{i=1}^n A_i\right) &\geq \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) \end{aligned}$$

Schließlich gilt allgemein die Ein- und Ausschlußformel:

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

## 2.4 Chancen-/Gewinnverhältnisse und Auszahlungsquoten

### Das Chancenverhältnis

(englisch: odds)

Sei  $A$  ein Ereignis und  $A^c$  das Gegenereignis.

$$R(A) := \frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

heißt das **Chancenverhältnis** von  $A$ .

**Beispiel:** “Sechs” beim Würfeln

Für einen fairen Würfel gilt:

$$P(\{6\}) = \frac{1}{6} \text{ und } P(\{6\}^c) = \frac{5}{6}$$
$$R(A) = \frac{1/6}{5/6} = \frac{1}{5}.$$

In Worten:  $R(A)$  ist 1 zu 5.

### Das Gewinnverhältnis

Bleiben wir bei obigem Beispiel und spielen folgendes Spiel. Der Spieler gewinnt, wenn die “6” kommt, ansonsten verliert er. Der Einsatz sei 1€. Die Auszahlung ist 6€, wenn “6” kommt und 0€, wenn keine “6” kommt. Die Auszahlung setzt sich aus 5€ Gewinn + 1€ Einsatz zusammen. Der mittlere Gewinn (MG) ist

$$\text{MG} = P(\{6\}) \cdot \text{Gewinn} - P(\{6\}^c) \cdot \text{Einsatz} = \frac{1}{6} \cdot 5 - \frac{5}{6} \cdot 1 = 0.$$

Da der mittlere Gewinn gleich 0 ist, spricht man von einem fairen Spiel. Das Gewinnverhältnis beträgt hier 5 zu 1 (5:1).

### Grundprinzip des fairen Wettens

Bei einer fairen Wette, verhalten sich die Gewinnverhältnisse umgekehrt proportional wie die Chancenverhältnisse.

Man kann dieses Prinzip aber auch über den Einsatz ausdrücken: Die Einsätze sind proportional zu den Wahrscheinlichkeiten.

Man denke nur an das Würfelspiel, bei dem der Spieler 1€ Einsatz und der Wettanbieter 5€ zahlen.

## Auszahlungsverhältnisse

Bei Sportwetten wie z.B. b-win werden nicht die Gewinnverhältnisse angegeben, sondern die Auszahlungsverhältnisse, auch einfach Quoten genannt. Das Auszahlungsverhältnis gibt die Auszahlung im Verhältnis zum Einsatz an. Die Auszahlung setzt sich aus Einsatz und Gewinn zusammen. So gab es für das Bundesligaspiel "Karlsruher SC - TSG Hoffenheim" am 18.04.2009 die Quoten:

- 1) 2,85 : 1 bei Karlsruher SC-Sieg
- 2) 3,20 : 1 bei Unentschieden
- 3) 2,30 : 1 bei Karlsruher SC-Niederlage.

Das Angebot ist nicht ganz fair, was natürlich daran liegt, daß b-win etwas Gewinn machen will.

Betrachten wir nochmals die Situation beim Würfel, wenn wir die drei Ereignisse  $\{1, 2, 3\}$ ,  $\{4, 5\}$  und  $\{6\}$  wählen. Faire Quoten sind dann:

- 1) 2 : 1 bei  $\{1, 2, 3\}$
- 2) 3 : 1 bei  $\{4, 5\}$
- 3) 6 : 1 bei  $\{6\}$ .

Die jeweiligen Wahrscheinlichkeiten ergeben sich gerade als die Kehrwerte der Quoten. Dies gilt allgemein.

Hat die faire Wette auf ein Ereignis  $E$  die Wahrscheinlichkeit  $p$  und den Gewinn  $G$ , so gilt bei einem Einsatz von 1

$$p(G + 1) = 1.$$

Denn sei  $E^c$  das Gegenereignis von  $E$ , so ist der mittlere Gewinn

$$MG = G \cdot p - 1 \cdot (1 - p) = 0.$$

Diese Gleichung ist aber äquivalent zur vorangegangenen. Also gilt bei einer fairen Wette der Zusammenhang

$$p = \frac{1}{G + 1}, \text{ wobei } G + 1 \text{ die Wettquote ist.}$$

Bezogen auf das Spiel "Karlsruher SC" gegen "TSG Hoffenheim" findet man

Karlsruher SC-Sieg	$p_1 = 0,35$
Unentschieden	$p_2 = 0,36$
Karlsruher SC-Niederlage	$p_3 = 0,43$

Da  $p_1 + p_2 + p_3 = 1,14$  ist, ist die Wette nicht ganz fair. Bekanntlich ist das Spiel 2:2 ausgegangen.

## 2.5 Gleichverteilungen

Man nimmt an, es gibt endlich viele Ausgänge, die alle gleich wahrscheinlich sind.

Sei  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$  mit  $n = |\Omega|$

$$P(\{\omega_1\}) = P(\{\omega_2\}) = \dots = P(\{\omega_n\}) = 1/n$$

Für beliebiges  $A \subset \Omega$  gilt dann:

$$P(A) = \frac{\text{Anzahl der für A günstigen Ergebnisse}}{\text{Anzahl aller möglichen Ergebnisse}} = \frac{|A|}{|\Omega|}.$$

Dies ist die sogenannte *Abzählregel*.

**Beispiel:**

Dreimaliges Werfen einer fairen Münze. Wappen kodieren wir mit 0 und Zahl mit 1. Dann ist

$$\begin{aligned} \Omega &= \{0, 1\} \times \{0, 1\} \times \{0, 1\} \\ &= \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0), (1, 0, 1), (0, 1, 1), (1, 1, 1)\} \end{aligned}$$

Offensichtlich ist  $|\Omega| = 2^3 = 8$ . Sei  $A$  das Ereignis *mindestens einmal Wappen und mindestens einmal Zahl*. Dann ist

$$A = \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0), (1, 0, 1), (0, 1, 1)\}.$$

Wir haben  $|A| = 6$  und somit  $P(A) = 6/8 = 3/4$ . Man kann natürlich auch über das Gegenereignis *kein mal Wappen oder kein mal Zahl* argumentieren und erhält

$$A^c = \{(0, 0, 0), (1, 1, 1)\}$$

mit  $P(A^c) = 2/8 = 1/4$ .

Jetzt läßt sich auch erläutern, warum es sinnvoll ist, Eigenschaft 3. von Definition 2.3.3 zu fordern (die sogenannte  $\sigma$ -Additivität). Man werfe eine faire Münze beliebig lange. Was ist die Wahrscheinlichkeit irgendwann eine "1" zu werfen? Das entsprechende Ereignis lautet:

$$A = \{(1), (0, 1), (0, 0, 1), \dots\}$$

$$\begin{aligned} P(A) &= P((1)) + P((0, 1)) + P((0, 0, 1)) + \dots \\ &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots \\ &= \sum_{i=1}^{\infty} 2^{-i} = 1. \end{aligned}$$

## Kombinatorische Hilfsmittel

Zunächst führen wir die Binomialkoeffizienten ein. Wie betrachten ein quadratisches Gitter und zählen die Wege von  $(0,0)$  zu dem Punkt  $(k, n-k)$ . Dies geht rekursiv. Wir bezeichnen diese Anzahl mit  $\binom{n}{k}$ .

1							
1	7						
1	6	21					
1	5	15	35				
1	4	10	20	35			
1	3	6	10	15	21		
1	2	3	4	5	6	7	
1	1	1	1	1	1	1	1

Drehen wir nun dieses Schema, so erhalten wir das Pascalsche Dreieck.

				1					
				1	1				
			1	2	1				
		1	3	3	1				
	1	4	6	4	1				
	1	5	10	10	5	1			
1	6	15	20	15	6	1			
1	7	21	35	35	21	7	1		

Nach Konstruktion gehorcht es dem folgenden Bildungsgesetz

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \quad \text{für } 1 \leq k \leq n-1 \text{ und } n \geq 1.$$

Zusätzlich gilt

$$\binom{n}{n} = \binom{n}{0} = 1 \quad \text{für } n \geq 0 \quad \text{und} \quad \binom{n}{k} = 0 \quad \text{für } k > n.$$

Aus dem Bildungsgesetz folgt mit vollständiger Induktion

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Dabei ist  $n! = n \cdot (n-1) \cdots 2 \cdot 1$ , gesprochen “ $n$  Fakultät”.

Wir geben nun zwei Mengen an, deren *Mächtigkeiten* jeweils  $\binom{n}{k}$  ist.

1.) Sei  $A = \{(\omega_1, \dots, \omega_n) \mid \omega_i \in \{0, 1\}, \sum_{i=1}^n \omega_i = k\}$  Dann ist  $|A| = \binom{n}{k}$ .

Diese Menge entspricht genau den Wegen im Gitter mit Länge  $n$  und  $k$  *Anstiegen* in Richtung  $y$ -Achse.

2.) Sei  $\Omega = \{1, \dots, n\}$  und  $\mathcal{P}_k = \{A \subset \Omega \mid |A| = k\}$ . Dann ist  $|\mathcal{P}_k| = \binom{n}{k}$ .

Dies folgt so: Jede Teilmenge mit  $k$  Elementen einer  $n$ -elementigen Menge entspricht eine 0-1 Folge der Länge  $n$ . An der  $i$ -ten Stelle der 0-1 Folge steht eine 1, wenn das  $i$ -te Element in der Menge  $A$  liegt.

### Ein Beispiel:

In einer Urne sind 5 weiße und 4 schwarze nicht nummerierte Kugeln. Es werden 3 Kugeln gezogen. Was ist die Wahrscheinlichkeit 2 weiße und eine schwarze Kugel zu ziehen? Die Anzahl der möglichen Fälle ist  $\binom{9}{3}$ . Die Anzahl der günstigen Fälle ist  $\binom{5}{2} \binom{4}{1}$ . Die gesuchte Wahrscheinlichkeit beträgt also  $\frac{\binom{5}{2} \binom{4}{1}}{\binom{9}{3}} = \frac{10}{21}$ .

Hierbei wurde das folgende Argumentationsprinzip benutzt.

**Das Kombinationsprinzip:** Sei  $\Omega$  eine Menge von  $k$ -Tupeln  $\omega = (\omega_1, \dots, \omega_k)$ , die man als Ergebnisse eines aus  $k$  Telexperimenten bestehenden Zufallsexperimentes auffassen kann, wobei  $\omega_i$  das Ergebnis des  $i$ -ten Telexperimentes ist. Für das erste Telexperiment gebe es  $n_1$  mögliche Ausgänge. Für jedes  $i$  sei  $n_i$  die Zahl der möglichen Ausgänge des  $i$ -ten Telexperimentes, unabhängig davon wie die früheren Telexperimente ausgegangen sind. Dann ist:  $\#\Omega = n_1 n_2 \dots n_k$

**Was heißt Ziehen?** Sei  $N$  eine endliche Menge mit  $n$  verschiedenen Elementen. Aus  $N$  wird eine Stichprobe vom Umfang  $k$  gezogen. Wir interessieren uns für die Anzahl der verschiedenen Stichproben. Dies ist für die Berechnung von Wahrscheinlichkeiten nach der Abzählregel von Nutzen. Dabei gilt es zu beachten, dass unter *Ziehen* unterschiedliches gemeint sein kann.

- Berücksichtigung der Reihenfolge:
  - Reihenfolge der Ziehung wird berücksichtigt: geordnete Probe.
  - Reihenfolge der Ziehung wird nicht berücksichtigt: ungeordnete Probe.
- Zurücklegen:
  - mit Zurücklegen: Mehrfachziehung möglich.
  - ohne Zurücklegen: Mehrfachziehung nicht möglich.

Damit ergeben sich vier Kombinationsmöglichkeiten. Wir werden allerdings die Situation der *ungeordneten Probe mit Zurücklegen* nicht behandeln, da sie im Folgenden keine Rolle spielen wird.

**Geordnete Probe mit Zurücklegen:** Die Anzahl der Proben ist gleich der Anzahl der  $k$ -Tupel  $(x_1, \dots, x_k)$  mit  $x_i \in N$  für  $i = 1, \dots, k$ . Diese Anzahl ist  $n^k$ .

**Geordnete Probe ohne Zurücklegen:** Die Anzahl der Proben ist gleich der Anzahl der  $k$ -Tupel  $(x_1, \dots, x_k)$  mit  $x_i \in N$  für  $i = 1, \dots, k$ , bei denen  $x_i \neq x_j$  für  $i \neq j$  gilt. Mit

dem kombinatorischen Grundprinzip ergibt sich diese Anzahl (die Zahl der Möglichkeiten wird bei jeder Ziehung um eins kleiner!) als:

$$n(n-1)\dots(n-(k-1)) = \prod_{i=0}^{k-1} (n-i).$$

Offensichtlich muß hier  $k \leq n$  gelten. Für den Spezialfall  $k = n$  entspricht jede Stichprobe einer Anordnung der Elemente der Menge  $N$ . Man erhält also die *Permutationen* von  $N$ . Für die Anzahl  $n(n-1)\dots 1$  schreibt man  $n!$  (gesprochen *n Fakultät*). Damit ist

$$n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}.$$

**Ungeordnete Probe ohne Zurücklegen:** Die Anzahl solcher Proben ist gleich der Anzahl der verschiedenen Teilmengen von  $N$  mit genau  $k$  Elementen. Diese Anzahl ist

$$\frac{n(n-1)\dots(n-(k-1))}{k(k-1)\dots 1} = \frac{n!}{k!(n-k)!} = \binom{n}{k}.$$

Dies ist der Binomialkoeffizient  $n$  über  $k$ .

**Anwendung: (Lotto 6 aus 49)**

Es werden 6 Kugeln aus 49 Kugeln gezogen, die von 1 bis 49 nummeriert sind. Die Reihenfolge der Ziehung spielt dabei keine Rolle. Der Ziehungsmechanismus stellt sicher, daß jede Kombination (= ungeordnete Stichprobe ohne Zurücklegen!) gleichwahrscheinlich ist. Damit liegt Gleichverteilung vor.

Der Grundraum läßt sich folgendermassen beschreiben:

$$\Omega = \{(\omega_1, \dots, \omega_6) | 1 \leq \omega_1 < \omega_2 < \omega_3 < \omega_4 < \omega_5 < \omega_6 \leq 49\}.$$

Die Wahrscheinlichkeit für *Sechs Richtige* ist

$$\frac{1}{\binom{49}{6}} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{49 \times 48 \times 47 \times 46 \times 45 \times 44} \approx 7.510^{-8}.$$

Um die Wahrscheinlichkeit für *Drei Richtige* zu bestimmen, benötigen wir die Anzahl der Kombinationen, die genau drei der von uns getippten Zahlen enthalten. Unser Tipp zerlegt die 49 Zahlen in zwei Mengen: die 6 von uns getippten und die 43 von uns nicht getippten. Eine Kombination mit drei richtigen enthält drei Zahlen aus den 6 getippten und drei Zahlen aus den 43 nicht getippten. Die Wahrscheinlichkeit für *Drei Richtige* ist damit

$$\frac{\binom{6}{3} \binom{43}{3}}{\binom{49}{6}} \approx 0.0176..$$

Nun wollen wir noch eine Aufgabe behandeln, die sowohl die kombinatorischen Hilfsmittel als auch die Ein- und Ausschlußformel benutzt.

$n$  Personen kommen mit je einem Geschenk zu einer Party. Die Geschenke werden zufällig verteilt. Jede Person erhält genau ein Geschenk. Wie groß ist die Wahrscheinlichkeit, daß wenigstens eine Person ihr Geschenk zurückerhält?

Wir denken uns die  $n$  Personen durchnummeriert von  $1, \dots, n$ . Sei  $\omega_i$  die Nummer der Person, die ihr Geschenk von  $i$  erhält. Sei  $\Omega = \{\omega = (\omega_1, \dots, \omega_n) \mid 1 \leq \omega_i \leq n, \omega_i \neq \omega_j \text{ für } i \neq j\}$ . Diese stellt die Menge der Permutationen von  $1, \dots, n$  dar.  $\Omega$  hat nach dem Kombinationsprinzip  $n!$  Elemente und man nimmt Gleichverteilung an:

$$P(\{\omega\}) = \frac{1}{n!}.$$

Sei  $A_i$  das Ereignis, daß Person  $i$  ihr Geschenk zurückerhält. Daß genau  $k$  Personen ihr Geschenk zurückerhalten, hat Wahrscheinlichkeit

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_1 \cap \dots \cap A_k) = \frac{(n-k)!}{n!}.$$

Dann ist die Wahrscheinlichkeit, daß mindestens eine Person ihr Geschenk zurückerhält, nach der Ein- und Ausschlußformel gleich

$$\begin{aligned} P(A_1 \cup \dots \cup A_n) &= \sum_{k=1}^n \sum_{\{i_1, \dots, i_k\}} (-1)^{k-1} P(A_{i_1} \cap \dots \cap A_{i_k}) \\ &= \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} \frac{(n-k)!}{n!} \\ &= \sum_{k=1}^n \frac{(-1)^{k-1}}{k!} \\ &= 1 - \sum_{k=0}^n \frac{(-1)^k}{k!} \\ &\rightarrow 1 - e^{-1} \approx 0,63 \quad \text{für } n \rightarrow \infty. \end{aligned}$$

## 2.6 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

### Einführendes Beispiel:

Eine faire Münze wird dreimal hintereinander geworfen. Dabei entspreche 1 dem Ausgang Zahl und 0 dem Ausgang Wappen in einem einzelnen Wurf. Also  $\Omega = \{0, 1\} \times \{0, 1\} \times \{0, 1\}$ . Sei  $A$  das Ereignis *mindestens zweimal Zahl*:

$$A = \{(0, 1, 1), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}.$$

Für die faire Münze sind alle Ergebnisse gleichwahrscheinlich. Damit ist

$$P(A) = \frac{|A|}{|\Omega|} = \frac{4}{8} = \frac{1}{2}.$$

Angenommen wir wissen bereits, daß der erste Wurf Zahl ergeben hat. Wie ändert sich unsere Einschätzung der Wahrscheinlichkeit für das Eintreten von  $A$ ? Wir wissen also, daß das Ereignis

$$B = \{(1, 0, 0), (1, 1, 0), (1, 0, 1), (1, 1, 1)\}$$

auf alle Fälle eintritt. Was ist die *bedingte* Wahrscheinlichkeit von  $A$  gegeben dieses Wissen? Die Intuition legt es nahe, alle Elemente aus  $B$  als gleichwahrscheinlich anzusehen und diese Menge als neuen Grundraum heranzuziehen. Damit erhalten wir als bedingte Wahrscheinlichkeit

$$\frac{|A \cap B|}{|B|} = \frac{3}{4} = \frac{\frac{|A \cap B|}{|\Omega|}}{\frac{|B|}{|\Omega|}} = \frac{P(A \cap B)}{P(B)}.$$

**Definition 2.6.1** Seien  $A$  und  $B$  Ereignisse und sei  $P(B) > 0$ . Dann ist

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

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

Aus der Definition folgt sofort (für  $P(B) > 0$ ) die Multiplikationsregel

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

Aus der Definition folgt ebenfalls

$$P(A^c|B) = 1 - P(A|B).$$

Es gilt nämlich  $B = (A \cap B) \cup (A^c \cap B)$  und  $(A \cap B) \cap (A^c \cap B) = \emptyset$ . Damit ist  $P(A^c \cap B) = P(B) - P(A \cap B)$  und weiter

$$\frac{P(A^c \cap B)}{P(B)} = \frac{P(B) - P(A \cap B)}{P(B)} = 1 - \frac{P(A \cap B)}{P(B)} = 1 - P(A|B).$$

Sind  $A_1$  und  $A_2$  zwei disjunkte Ereignisse, so gilt allgemein

$$P((A_1 \cup A_2) \cap B) = P(A_1 \cap B) + P(A_2 \cap B).$$

Somit ist

$$\begin{aligned} P(A_1 \cup A_2|B) &= \frac{P((A_1 \cup A_2) \cap B)}{P(B)} \\ &= \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)} = P(A_1|B) + P(A_2|B). \end{aligned}$$

Bedingte Wahrscheinlichkeiten verhalten sich also bezüglich des ersten Argumentes wie gewöhnliche Wahrscheinlichkeiten.

**Beispiel:**

In einem Hut befinden sich drei Spielkarten, von denen eine auf der Vorder- und Rückseite *weiß* ist, eine weitere, die auf der Vorder- und Rückseite *schwarz* ist und eine dritte, die auf der Vorderseite *weiß* und auf der Rückseite *schwarz* ist. Eine gezogene Karte zeigt auf der Vorderseite *schwarz*. Wie groß ist die Wahrscheinlichkeit, daß sie auf der Rückseite *weiß* ist? Wir nummerieren die Seiten der Spielkarten wie folgt:

1. Karte  $w_1/w_2$
2. Karte  $s_1/s_2$
3. Karte  $w_3/s_3$ .

Nun ist die Menge der möglichen Ereignisse

$$\Omega = \{s_1, s_2, s_3, w_1, w_2, w_3\}$$

und damit

$$P(\{s_i\}) = P(\{w_i\}) = \frac{1}{6} \quad \text{für } i = 1, 2, 3.$$

Die Wahrscheinlichkeit, daß eine gezogene Karte auf der Vorderseite *schwarz* ist, ist dann

$$P(\text{schwarz}) = \frac{\#\{s_1, s_2, s_3\}}{\#\Omega} = \frac{1}{2}$$

Weiter ist

$$P(\text{Rückseite weiß} \mid \text{Vorderseite schwarz}) = \frac{\#\{s_3\}}{\#\{s_1, s_2, s_3\}} = \frac{1}{3}.$$

**Beispiel:**

Ein elektrisches Gerät, enthält zwei Schaltkreise I und II. Schaltkreis I fällt dabei mit Wahrscheinlichkeit 0.1 aus. Fällt Schaltkreis I aus, so fällt Schaltkreis II mit Wahrscheinlichkeit 0.2 ebenfalls aus. Bleibt Schaltkreis I intakt, so fällt Schaltkreis II mit Wahrscheinlichkeit 0.05 aus. Mit welcher Wahrscheinlichkeit fallen beide Schaltkreise aus? Sei

$$A = \{\text{Schaltkreis I fällt aus}\}$$

$$B = \{\text{Schaltkreis II fällt aus}\}$$

Aus den obigen Angaben erhalten wir

$$P(A) = 0.1, P(B|A) = 0.2 \text{ und } P(B|A^c) = 0.05$$

Damit folgt  $P(A \cap B) = 0.1 \times 0.2 = 0.02$ .

Können wir auch  $P(A|B)$  berechnen? Dazu benötigen wir  $P(B)$ :

$$\begin{aligned} P(B) &= P(A \cap B) + P(A^c \cap B) \\ &= P(B|A)P(A) + P(B|A^c)P(A^c) \\ &= P(B|A)P(A) + P(B|A^c)(1 - P(A)) \\ &= 0.2 \times 0.1 + 0.05 \times 0.9 = 0.065. \end{aligned}$$

Es folgt

$$P(A|B) = \frac{0.02}{0.065} = 0.308.$$

Die Rechnung, die wir hier für  $P(B)$  gemacht haben, gilt allgemein, wie der folgende Satz bestätigt.

**Satz von der totalen Wahrscheinlichkeit:**

Sei  $\{A_1, \dots, A_n\}$  eine Partition von  $\Omega$ . Sei  $B$  ein weiteres Ereignis. Dann gilt:

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

Begründung:

$$\begin{aligned} P(B) &= P(B \cap \Omega) = P(B \cap (\cup_{i=1}^n A_i)) \\ &= P(\cup_{i=1}^n (B \cap A_i)) = \sum_{i=1}^n P(B \cap A_i) \\ &= \sum_{i=1}^n P(B|A_i)P(A_i). \end{aligned}$$

**Beispiel:** Ziehen ohne Zurücklegen

Von einem gut gemischten Skat-Blatt (bestehend aus 32 Karten) werden vom Stapel nacheinander zwei Karten gezogen. Was ist die Wahrscheinlichkeit, daß die zweite Karte "Kreuz oder Pik", d.h. "schwarz" ist? Vermutung: Sie ist gleich  $\frac{1}{2}$ .

$$\begin{aligned} &P(2. Karte \text{ "schwarz"}) \\ &= P(2. Karte \text{ "schwarz"} \mid 1. Karte \text{ "schwarz"}) \cdot P(1. Karte \text{ "schwarz"}) \\ &\quad + P(2. Karte \text{ "schwarz"} \mid 1. Karte \text{ "rot"}) \cdot P(1. Karte \text{ "rot"}) \\ &= \frac{1}{2} \cdot \frac{15}{31} + \frac{1}{2} \cdot \frac{16}{31} = \frac{1}{2} \cdot \frac{15 + 16}{31} = \frac{1}{2}. \end{aligned}$$

Eine direkte Folgerung aus dem Satz von der totalen Wahrscheinlichkeit ist die Bayessche Formel. Sie ist von grundlegender Bedeutung und beschreibt, wie neues Wissen zu verwerten ist. Sie wird deswegen auch oft als Lernformel bezeichnet.

**Bayessche Formel:**

$\{A_1, \dots, A_n\}$  sei eine Partition von  $\Omega$ . Sei  $B$  ein weiteres Ereignis. Dann gilt:

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

**Denn:**

$$P(A_i | B) = \frac{P(B | A_i)P(A_i)}{P(B)} = \frac{P(B | A_i)P(A_i)}{\sum_{j=1}^m P(B | A_j)P(A_j)}$$

Dabei wurde in der letzten Gleichung der Satz von der totalen Wahrscheinlichkeit verwendet.

**Beispiel:** Farbenblindheit

Farbenblindheit ist eine typische Männerkrankheit.  $M$  stehe für männlich,  $W$  für weiblich und  $fb$  für farbenblind. Die Zahlenverhältnisse seien wie folgt:

$$P(M) = P(W) = \frac{1}{2}, \quad P(fb|M) = \frac{1}{12}, \quad P(fb|W) = \frac{1}{288}.$$

Wie groß ist die Wahrscheinlichkeit, daß eine Person männlich ist, wenn sie farbenblind ist?

$$P(M|fb) = \frac{P(fb|M)P(M)}{P(fb|M)P(M) + P(fb|W)P(W)} = \frac{\frac{1}{12} \cdot \frac{1}{2}}{\frac{1}{12} \cdot \frac{1}{2} + \frac{1}{288} \cdot \frac{1}{2}} = \frac{24}{25}.$$

**Beispiel:**

Zuverlässigkeit von Prüfverfahren. Ein bestimmter Chip wird in Massenproduktion hergestellt. Dabei wird jeder Chip vor Auslieferung getestet. Die Produktion hat eine Ausschußrate von 0.01; daß heißt ein einzelner Chip ist mit Wahrscheinlichkeit 0.01 defekt. Das Prüfverfahren besitzt folgende Eigenschaften:

- bei einem fehlerfreien Chip zeigt das Prüfverfahren mit Wahrscheinlichkeit 0.1 fälschlich einen Fehler an.
- bei einem fehlerbehafteten Chip zeigt das Prüfverfahren mit Wahrscheinlichkeit 0.05 fälschlich keinen Fehler an.

Chips bei denen das Prüfverfahren einen Fehler anzeigt werden aussortiert und die verbleibenden Chips werden ausgeliefert. Mit welcher Wahrscheinlichkeit ist ein ausgelieferte Chip auch wirklich fehlerfrei? Sei

$A = \{\text{Chip ist fehlerfrei}\}$  und  $B = \{\text{Prüfverfahren zeigt Fehler an}\}$ .

Wir wissen  $P(A) = 0.99$ ,  $P(B|A) = 0.1$  und  $P(B|A^c) = 0.95$ . Mit  $A_1 = A$  und  $A_2 = A^c$  liefert die Formel von Bayes

$$\begin{aligned} P(A|B^c) &= \frac{P(B^c|A)P(A)}{P(B^c|A)P(A) + P(B^c|A^c)P(A^c)} \\ &= \frac{0.9 \times 0.99}{0.9 \times 0.99 + 0.05 \times 0.01} \\ &= 0.999. \end{aligned}$$

Die Wahrscheinlichkeit, daß ein aussortierter Chip auch wirklich defekt ist ergibt sich mit analoger Rechnung als

$$P(A^c|B) = \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.1 \times 0.99} = 0.0876.$$

**Beispiel:** Welche Urne (Box) ?

3 Urnen mit weißen und schwarzen Kugeln seien gegeben: Es wird eine Urne zufällig ausgewählt und dann eine Kugel daraus gezogen; die Kugel wird gezeigt, nicht aber die Urne. Man rate, aus welcher Urne gezogen wurde. Bayes-Formel:  $P(\text{Urne } i \mid \text{weiß}) = ?$

$$P(\text{Urne } i \text{ und weiß}) = P(\text{weiß} \mid \text{Urne } i)P(\text{Urne } i) = \frac{1}{3} \cdot \frac{i}{i+1}$$

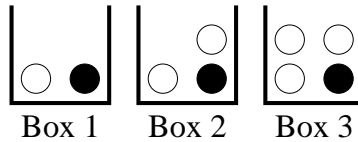


Abbildung 2.1: Urnen 1-3

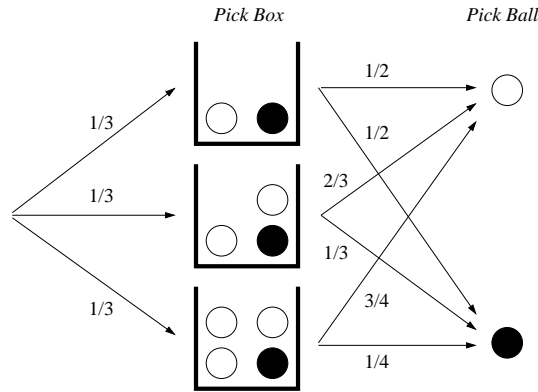


Abbildung 2.2: Wahrscheinlichkeiten

$$P(\text{weiß}) = \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{3}{4} = \frac{23}{36}$$

Dann folgt

$$P(\text{Urne 3} \mid \text{weiß}) = \frac{\frac{1}{3} \cdot \frac{3}{4}}{\frac{23}{36}} = \frac{\frac{3}{12}}{\frac{23}{36}} = \frac{9}{23}$$

Weiter sind:

$$P(\text{Urne 2} \mid \text{weiß}) = \frac{8}{23}$$

$$P(\text{Urne 1} \mid \text{weiß}) = \frac{6}{23}$$

**Definition 2.6.2** Zwei Ereignisse  $A$  und  $B$  heißen (stochastisch) unabhängig, falls

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

Sind  $A$  und  $B$  unabhängig und gilt  $P(B) > 0$ , so ist

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

Sind  $A$  und  $B$  unabhängig und gilt  $P(A) > 0$ , so ist

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

Stochastische Unabhängigkeit zweier Ereignisse bedeutet also, daß Kenntnis über das Eintreten des einen Ereignisses keine Information hinsichtlich des Eintretens des anderen

Ereignisses liefert.

Sind  $A$  und  $B$  unabhängig, so sind auch  $A$  und  $B^c$  unabhängig. Es gilt dann nämlich

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) \\ &= P(A)P(B^c). \end{aligned}$$

Betrachtet man mehr als zwei Ereignisse, so werden die Forderungen komplizierter.

**Definition 2.6.3** *Drei Ereignisse  $A$ ,  $B$  und  $C$  heißen unabhängig, falls die folgenden vier Gleichungen gelten:*

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

Gelten die ersten drei Gleichungen, so sind die Paare  $A$  und  $B$ ,  $A$  und  $C$  und  $B$  und  $C$  jeweils unabhängig. Man spricht in diesem Fall von paarweiser Unabhängigkeit. Die dritte Gleichung stellt eine zusätzliche Forderung dar und folgt nicht aus den ersten drei Gleichungen.

**Beispiel:**

Sei  $\Omega = \{0, 1\} \times \{0, 1\}$  und

$$P(\{(0, 0)\}) = P(\{(1, 0)\}) = P(\{(0, 1)\}) = P(\{(1, 1)\}) = 1/4.$$

Dies entspricht also dem zweimaligen Werfen einer fairen Münze. Sei

$A = \{(0, 1), (0, 0)\}$  (Erster Wurf liefert Wappen)

$B = \{(0, 1), (1, 1)\}$  (Zweiter Wurf liefert Zahl)

$C = \{(0, 0), (1, 1)\}$  (Beide Würfel liefern dasselbe Ergebnis)

Dann gilt  $|A| = |B| = |C| = 2$  und  $|A \cap B| = |A \cap C| = |B \cap C| = 1$ . Daraus folgt  $P(A) = P(B) = P(C) = 1/2$  und  $P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/4$ . Außerdem ist  $A \cap B \cap C = \emptyset$  und somit

$$P(A \cap B \cap C) = 0 \neq 1/8 = P(A)P(B)P(C).$$

Eine Menge von  $n$  Ereignissen heißt (stochastisch) unabhängig, falls für jede Teilmenge der  $n$  Ereignissen gilt: die Wahrscheinlichkeit des Schnittereignisses ergibt sich als Produkt der Einzelwahrscheinlichkeiten. Für  $n = 3$  sind dies vier Gleichungen; bei  $n = 4$  ergeben sich bereits elf Gleichungen.

Man kann Unabhängigkeit aber auch wie folgt definieren. Die Ereignisse  $A_1, \dots, A_n$  heißen unabhängig, falls

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

gilt. Dabei sind  $B_i$  entweder gleich  $A_i$  oder  $A_i^c$  für  $i = 1, \dots, n$ . Dies sind  $2^n$  Gleichungen. Beide Definitionen sind äquivalent.

Es folgen nun einige Beispiele, die die Nützlichkeit des Begriffs der Unabhängigkeit verdeutlichen.

**Beispiel:** Funktion einer Maschine

Eine Maschine besteht aus zwei unabhängig voneinander arbeitenden Teilen. Wenn beide Teile fehlerhaft arbeiten oder wenn die Maschine falsch bedient wird, entsteht Ausschussware. Dabei sind Bedienungsfehler unabhängig von der Arbeitsweise der Maschine. Die Wahrscheinlichkeit, daß der erste Teil fehlerhaft arbeitet, ist 0,05; beim zweiten Teil ist sie 0,08; die Wahrscheinlichkeit für einen Bedienungsfehler ist 0,02.

Mit welcher Wahrscheinlichkeit wird Ausschussware produziert?

$$\begin{aligned} A &= \{1. \text{ Teil fehlerfrei}\} & P(A) &= 0,95 \\ B &= \{2. \text{ Teil fehlerfrei}\} & P(B) &= 0,92 \\ C &= \{\text{kein Bedienfehler}\} & P(C) &= 0,98 \end{aligned}$$

$$\begin{aligned} P(\text{Defekt}) &= P(A^c \cap B^c \cap C) + P(C^c) \\ &= P((A^c \cap B^c)) P(C) + P(C^c) \\ &= P(A^c)P(B^c)P(C) + P(C^c) \\ &= 0,05 \cdot 0,08 \cdot 0,98 + 0,02 \\ &= 0,00392 + 0,02 = 0,02392 \\ &\approx 2,4\% \end{aligned}$$

**Beispiel:** Wahrscheinlichkeit, daß im  $k$ -ten Wurf erstmals eine “6” gewürfelt wird.

Erstmals im  $k$ -ten Wurf eine “6” zu werfen, bedeutet, in den vorangegangenen  $k - 1$  Würfeln keine “6” , dann aber eine “6” zu werfen. Folglich ist

$$P(\text{erstmalig “6” im } k\text{-ten Wurf}) = \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6} =: p_k, \quad k \geq 1.$$

Es muß natürlich  $\sum_{k \geq 1} p_k = 1$  gelten, was man leicht mit Hilfe der geometrische Reihe

$$\sum_{k=1}^{\infty} \left(\frac{5}{6}\right)^{k-1} = 6 \text{ einsieht.}$$

**Gamblers Rules**

Angenommen man spielt ein Spiel sehr oft hintereinander und dessen Gewinnchance ist  $1/N$ . Wie oft muß man spielen, damit man mit mindestens 50% Wahrscheinlichkeit

wenigstens einmal gewinnt?

$$\begin{aligned}
 P(\text{kein Gewinn in } n \text{ Spielen}) &= \left(1 - \frac{1}{N}\right)^n \\
 P(\text{Gewinn in } n \text{ Spielen}) &= 1 - \left(1 - \frac{1}{N}\right)^n \geq \frac{1}{2} \\
 &\Leftrightarrow \left(1 - \frac{1}{N}\right)^n \leq \frac{1}{2} \\
 &\Leftrightarrow n \log \left(1 - \frac{1}{N}\right) \leq \log \frac{1}{2}.
 \end{aligned}$$

Sei  $n^* = \lceil \log(\frac{1}{2}) / \log(1 - \frac{1}{N}) \rceil$ ; dabei ist  $\lceil x \rceil$  die kleinste ganze Zahl größer als  $x$  für  $x \in \mathbb{R}$ . Da  $\log(1 + z) \sim z$  für  $z \rightarrow 0$  ist die rechte Seite von  $n^*$  asymptotisch gleich

$$\log \left(\frac{1}{2}\right) / \left(-\frac{1}{N}\right) = N \log 2.$$

Da  $\log(2) \approx 0,69 \approx 2/3$  ist, lautet die Regel,  $n$  muß größer als  $2/3 N$  sein. Für  $N = 1, 2, \dots, 27$  gilt tatsächlich  $n^* = \lceil 2/3 N \rceil$ . Im Fall des Würfels mit Erfolg beim Werfen einer "6" ist

$$n^* = \left\lceil \log \left(\frac{1}{2}\right) / \left(-\frac{5}{6}\right) \right\rceil = \lceil 3,8 \rceil = 4.$$

### Beispiel: Fluß in einem Leiter

Angenommen für jeden von den Schaltern in dem folgenden Schaltkreis ist die Wahrscheinlichkeit, daß der Schalter geschlossen ist  $p_i$  und daß er offen ist  $q_i = 1 - p_i$ ,  $i = 1, \dots, 5$ . Man berechne die Wahrscheinlichkeit, daß ein Strom durch den Schaltkreis fließt unter der Annahme, daß die Schalter unabhängig sind.

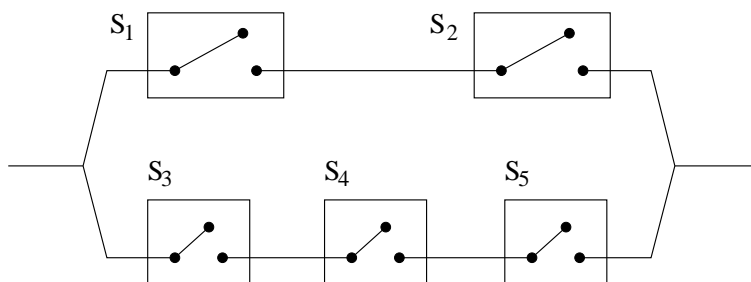


Abbildung 2.3: Schaltkreis

$$\begin{aligned}
 P(\text{Strom fließt}) &= P(\text{Strom fließt oben entlang}) + P(\text{Strom fließt unten entlang}) \\
 &\quad - P(\text{Strom fließt sowohl oben als auch unten entlang})
 \end{aligned}$$

Dabei ist  $P(\text{Strom fließt oben}) = p_1 \cdot p_2$  und  $P(\text{Strom fließt unten}) = p_3 \cdot p_4 \cdot p_5$ , wegen Unabhängigkeit.  $P(\text{Strom fließt oben und unten}) = p_1 \cdot p_2 \cdot p_3 \cdot p_4 \cdot p_5$

$$P(\text{Strom fließt}) = p_1 \cdot p_2 + p_3 \cdot p_4 \cdot p_5 - p_1 \cdot p_2 \cdot p_3 \cdot p_4 \cdot p_5$$

## 2.7 Kombination von Ereignissen

Seien  $E_1, E_2, \dots, E_k$  Ereignisse. Dann gilt

$$P(E_1 \cap E_2 \cap \dots \cap E_k) = P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2) \dots P(E_k|E_1 \cap E_2 \cap \dots \cap E_{k-1}).$$

Ist jeweils  $E_i$  von  $E_1 \cap \dots \cap E_{i-1}$  unabhängig, so folgt

$$P(E_1 \cap E_2 \cap \dots \cap E_k) = P(E_1)P(E_2) \dots P(E_k).$$

### Beispiel:

Geburtstagsproblem.  $k$  Personen befinden sich in einem Raum. Mit welcher Wahrscheinlichkeit haben mindestens zwei von ihnen am selben Tag Geburtstag? Wir treffen folgende vereinfachende Annahmen:

- Das Jahr hat 365 Tage.
- Jeder Tag kommt mit gleicher Wahrscheinlichkeit als Geburtstag in Frage.
- Es besteht keine Abhängigkeit zwischen den Geburtstagen verschiedener Personen (also keine Zwillinge!).

Der Einfachheit halber denken wir uns die Personen von 1 bis  $k$  numeriert und stellen uns vor, daß wir sie der Reihe nach befragen. Sei

$$\Omega = \{1, \dots, 365\}^k = \{\omega | \omega = (\omega_1, \dots, \omega_k); \omega_i \in \{1, \dots, 365\}\}.$$

Dabei ist  $\omega_i$  der Geburtstag der  $i$ -ten Person.

Es ist einfacher, die Wahrscheinlichkeit des Gegenereignisses  $D$  *keine zwei Personen haben am selben Tag Geburtstag* zu bestimmen. Die Überlegungen des Abschnittes 2.5 liefern sofort (Ziehen mit und ohne Zurücklegen)

$$|\Omega| = 365^k \quad \text{und} \quad |D| = \prod_{j=0}^{k-1} (365 - j)$$

und damit

$$P(D) = \prod_{j=0}^{k-1} \frac{365 - j}{365}.$$

Die Wahrscheinlichkeit, daß mindestens zwei Personen am selben Tag Geburtstag haben, beträgt somit

$$1 - \prod_{j=0}^{k-1} \frac{365 - j}{365}.$$

Für  $k = 23$  haben wir mit Wahrscheinlichkeit 0.506 mindestens zwei Personen am selben Tag Geburtstag. Bei  $k = 65$  beträgt die Wahrscheinlichkeit bereits 0.998.

Die obige Formel läßt sich auch mittels bedingter Wahrscheinlichkeiten herleiten. Sei dazu

$$\begin{aligned} D_j &= \{(j+1)\text{-te Person hat an einem anderen Tag Geburtstag als die Personen} \\ &\quad 1 \text{ bis } j\} \\ &= \{\omega \in \Omega \mid \omega_{j+1} \neq \omega_i \text{ für } 1 \leq i \leq j\}. \end{aligned}$$

Dann gilt

$$P(D_1) = P(\omega_1 \neq \omega_2) = \frac{365 \times 364}{365 \times 365} = \frac{364}{365}.$$

Sei nun  $j \geq 2$ . Auf dem Ereignis  $D_1 \cap \dots \cap D_{j-1}$  haben die Personen  $1, \dots, j$  an  $j$  verschiedenen Tagen Geburtstag. Damit ergibt sich die bedingte Wahrscheinlichkeit des Ereignisses  $D_j$  gegeben  $D_1 \cap \dots \cap D_{j-1}$  zu

$$P(D_j \mid D_1 \cap \dots \cap D_{j-1}) = \frac{365 - j}{365} = 1 - \frac{j}{365}.$$

Wir erhalten nun

$$\begin{aligned} &P(D_1 \cap D_2 \cap \dots \cap D_{k-1}) \\ &= P(D_1)P(D_2 \mid D_1)P(D_3 \mid D_1 \cap D_2) \dots P(D_{k-1} \mid D_1 \cap D_2 \cap \dots \cap D_{k-2}) \\ &= \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \dots \left(1 - \frac{k-1}{365}\right) \\ &= \prod_{j=1}^{k-1} \left(1 - \frac{j}{365}\right) = \prod_{j=0}^{k-1} \left(1 - \frac{j}{365}\right). \end{aligned}$$

Für größere Werte von  $k$  bietet sich folgende Näherung an. Es gilt  $\log(1 - h) \approx -h$  und somit

$$\begin{aligned} \log(P(D_1 \cap D_2 \cap \dots \cap D_{k-1})) &= \sum_{j=1}^{k-1} \log(1 - j/365) \approx -(1/365) \sum_{j=1}^{k-1} j \\ &= -\frac{k(k-1)}{2 \times 365}. \end{aligned}$$

Die Wahrscheinlichkeit, daß mindestens zwei Personen am selben Tag Geburtstag haben, ist daher näherungsweise

$$1 - e^{-\frac{k(k-1)}{2 \times 365}}.$$

Die Näherung ist sehr gut. Für  $k = 23$  liefert sie 0.500 im Vergleich zu dem exakten Wert 0.506.

# Kapitel 3

## Zufallsvariable und Erwartungswerte

### 3.1 Grundlagen

Eine Zufallsvariable ist ein Zahlenwert als Ergebnis eines Zufallsexperimentes.

**Definition 3.1.1** Sei  $\Omega$  ein Grundraum. Eine Abbildung  $X$

$$X : \Omega \longrightarrow \mathbf{R}$$

heißt **Zufallsvariable**.

Sei  $X$  eine diskrete Zufallsvariable mit Wertebereich  $x_1, x_2, \dots$ . Die **Verteilung von  $X$**  wird durch die Punktwahrscheinlichkeiten  $p_i$  mit

$$p_i = P(X = x_i) = P(\{\omega | X(\omega) = x_i\})$$

beschrieben.

**Bemerkung:** Es gilt  $\sum_{i=1}^{\infty} p_i = 1$ .

**Beispiel:**

Dreimaliges Werfen einer fairen Münze. Wieder stehe 1 für Zahl und 0 für Wappen. Also  $\Omega = \{0, 1\}^3$ . Sei für  $\omega = (\omega_1, \omega_2, \omega_3) \in \Omega$

$$X(\omega) = \omega_1 + \omega_2 + \omega_3.$$

Der Wertebereich von  $X$  ist  $\{0, 1, 2, 3\}$ . So ist etwa  $X((0, 0, 0)) = 0$  und  $X((0, 1, 1)) = 2$ .  $X$  zählt die Anzahl der Würfe mit Ergebnis Zahl. Es gilt

$$P(X = 0) = 1/8, \quad P(X = 1) = 3/8, \quad P(X = 2) = 3/8, \quad P(X = 3) = 1/8$$

**Definition 3.1.2** Ist  $X$  eine Zufallsvariable mit Werten  $x_1, x_2, \dots$ , so heißt

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i)$$

**Erwartungswert von  $X$ , falls**

$$\sum_{i=1}^{\infty} |x_i| P(X = x_i) < \infty.$$

**Beispiel:**

Für das vorangegangene Beispiel gilt

$$E(X) = 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) = \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = \frac{12}{8}.$$

**Beispiel:**

Einmaliges Werfen eines Würfels.  $\Omega = \{1, 2, \dots, 6\}$ . Sei  $X(i) = i$  für  $i = 1, \dots, 6$ . Dann ist  $P(X = i) = 1/6$  und  $E(X) = \frac{1}{6}(1 + 2 + \dots + 6) = \frac{3 \cdot 7}{6} = 3,5$ .

**Beispiel:**

Wie lange muß man beim wiederholten Würfeln im Mittel auf die erste Sechs warten? Allgemeiner: Gegeben sei eine Folge unabhängiger Versuche mit zwei Ausgängen: Mißerfolg (mit 0 bezeichnet) und Erfolg (mit 1 bezeichnet). In jedem Versuch sei die Wahrscheinlichkeit für einen Erfolg  $p$  und die Wahrscheinlichkeit für einen Mißerfolg  $q = 1 - p$ . Sei  $X$  die Anzahl der Versuche bis zum ersten Erfolg (einschließlich dieses Versuchs!). Wie groß ist der Erwartungswert von  $X$ ?

$X = k$  gilt genau dann, wenn die ersten  $k - 1$  Versuche Mißerfolge waren und der  $k$ -te Versuch ein Erfolg ist. Damit gilt

$$+P(X = k) = \underbrace{q \times \dots \times q}_{(k-1) \text{ mal}} \times p = q^{k-1}p.$$

Der Wertebereich von  $X$  ist  $\{1, 2, 3, \dots\}$ . Die Verteilung von  $X$  wird als *geometrische Verteilung* bezeichnet. Diese Bezeichnung rührt von der geometrischen Reihe

$$1 + q + q^2 + \dots = \sum_{j=0}^{\infty} q^j = \frac{1}{1 - q}$$

her. Damit sehen wir

$$\sum_{k=1}^{\infty} P(X = k) = p \sum_{k=1}^{\infty} q^{k-1} = p \sum_{j=0}^{\infty} q^j = \frac{p}{1 - q} = 1.$$

Der Erwartungswert von  $X$  ergibt sich zu

$$E(X) = \sum_{k=1}^{\infty} kq^{k-1}p = p \sum_{k=1}^{\infty} kq^{k-1}.$$

Die Reihe wertet man mit Hilfe folgenden Tricks aus:

$$\sum_{j=0}^{\infty} q^j = \frac{1}{1 - q} \Rightarrow \sum_{j=1}^{\infty} j q^{j-1} = \frac{d}{dq} \left( \sum_{j=0}^{\infty} q^j \right) = \frac{1}{(1 - q)^2}.$$

Dies liefert

$$E(X) = p \frac{1}{p^2} = \frac{1}{p}.$$

Bei einem einzelnen Würfelwurf erscheint mit Wahrscheinlichkeit  $p = 1/6$  die Sechs. Man benötigt deshalb im Mittel 6 Würfe, um eine Sechs zu erhalten.

Ist  $g$  eine Funktion von  $\mathbf{R}$  nach  $\mathbf{R}$  und  $X$  eine diskrete Zufallsvariable, so ist auch  $g(X)$  eine diskrete Zufallsvariable. Für den Erwartungswert von  $g(X)$  gilt

$$E(g(X)) = \sum_i g(x_i)P(X = x_i)$$

falls

$$\sum_i |g(x_i)|P(X = x_i) < \infty.$$

Es genügt also die Verteilung von  $X$  zu kennen, um  $E(g(X))$  zu berechnen.

Eine wichtige Eigenschaft des Erwartungswertes ist die Linearität. Für den Spezialfall  $g(x) = ax + b$  erhalten wir nämlich  $E(aX + b) = aE(X) + b$ .

Allgemeiner gilt für Zufallsvariablen  $X, Y$  mit existierenden Erwartungswerten  $E(X)$  und  $E(Y)$ , daß

$$E(aX + bY) = aE(X) + bE(Y)$$

gilt. Diese Eigenschaft heißt **Linearität des Erwartungswertes**.

Um diese Formel einzusehen, ist eine Darstellung des Erwartungswertes über dem Grundraum  $\Omega$  nützlich. Es gilt:

$$E(g(X)) = \sum_{\omega \in \Omega} g(X(\omega))P(\{\omega\}).$$

Dies zeigt man folgendermaßen:

$$\begin{aligned} \sum_{\omega} g(X(\omega))P(\{\omega\}) &= \sum_i \sum_{\omega|X(\omega)=x_i} g(X(\omega))P(\{\omega\}) \\ &= \sum_i \sum_{\omega|X(\omega)=x_i} g(x_i) P(\{\omega\}) \\ &= \sum_i g(x_i) \sum_{\omega|X(\omega)=x_i} P(\{\omega\}) \\ &= \sum_i g(x_i)P(\{\omega|X(\omega) = x_i\}) \\ &= E(g(X)) \end{aligned}$$

Nun läßt sich die Linearität des Erwartungswertes leicht zeigen. Für  $a = b = 1$  gilt

$$\begin{aligned} E(X + Y) &= \sum_{\omega} (X(\omega) + Y(\omega)) P(\{\omega\}) \\ &= \sum_{\omega} X(\omega)P(\{\omega\}) + \sum_{\omega} Y(\omega)P(\{\omega\}) \\ &= E(X) + E(Y). \end{aligned}$$

**Definition 3.1.3** Ist  $X$  eine diskrete Zufallsvariable mit Erwartungswert  $E(X)$ , so heißt

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

die **Varianz** von  $X$ .

Zur Berechnung der Varianz bietet sich oft die sogenannte *Verschiebungsformel* an:

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

Die Varianz mißt auf quadratischer Skala. Dies führt zur folgenden Definition für linear Skalen.

**Definition 3.1.4** Ist  $X$  eine diskrete Zufallsvariable mit endlicher Varianz  $\text{Var}(X)$ , so heißt  $\sigma(X) = \sqrt{\text{Var}(X)}$  die **Standardabweichung** von  $X$ .

Die Verschiebungsformel sieht man folgendermaßen ein ( $\mu = E(X)$ )

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^{\infty} (x_i - \mu)^2 P(X = x_i) \\ &= \sum_{i=1}^{\infty} (x_i^2 - 2x_i\mu + \mu^2) P(X = x_i) \\ &= \sum_{i=1}^{\infty} x_i^2 P(X = x_i) - 2\mu \sum_{i=1}^{\infty} x_i P(X = x_i) + \mu^2 \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - (E(X))^2. \end{aligned}$$

**Beispiel:**

Ist  $X$  das Ergebnis bei einmaligen Würfeln, so ist  $\text{Var}(X) = \frac{105}{36} = 2,91\bar{6}$  und  $\sigma(X) = 1,708$ .

**Beispiel:**

Ist  $X$  geometrisch verteilt, also  $P(X = k) = q^{k-1}p$  für  $k = \{1, 2, 3, \dots\}$ , so erhält man

$$\text{Var}(X) = \frac{q}{p^2} \quad \text{und} \quad \sigma(X) = \frac{\sqrt{q}}{p}.$$

(Hinweis: hierzu berechnet man zunächst  $\sum_k k(k-1)q^{k-2}$ )

## 3.2 Bernoulli-Verteilung und Binomialverteilung

Eine Zufallsvariable  $Y$  heißt *Bernoulli-verteilt*, falls sie nur die Werte 0 und 1 annimmt. Dann gilt

$$P(Y = 1) = p \quad \text{und} \quad P(Y = 0) = q = 1 - p$$

für ein  $0 \leq p \leq 1$ . Für den Erwartungswert der Bernoulli-Verteilung ergibt sich

$$E(Y) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

Weiter ist

$$E(Y^2) = 0 \cdot (1 - p) + 1 \cdot p = p$$

und mit der Verschiebungsformel folgt

$$\text{Var}(Y) = p - p^2 = pq.$$

Die Bernoulli-Verteilung ist der Grundbaustein für die Binomialverteilung. Wir betrachten nun  $n$  unabhängige Versuche mit den beiden Ausgängen Erfolg (entspricht 1) und Mißerfolg (entspricht 0). In jedem Versuch sei die Wahrscheinlichkeit für einen Erfolg  $p$  und die Wahrscheinlichkeit für einen Mißerfolg  $q = 1 - p$ . Wir interessieren uns nun für die Anzahl  $X$  der Erfolge bei  $n$  Versuchen. Formal läßt sich das wie folgt fassen.

$$\Omega = \{0, 1\}^n = \{(\omega_1, \dots, \omega_n) \mid \omega_i \in \{0, 1\}\}$$

Für  $\omega \in \Omega$  sei

$$\begin{aligned} (+) \quad P(\{\omega\}) &= P(\{(\omega_1, \dots, \omega_n)\}) \\ &= \prod_{i=1}^n P(\{\omega_i\}) \\ &= p^{\sum_{i=1}^n \omega_i} (1 - p)^{\sum_{i=1}^n (1 - \omega_i)} \end{aligned}$$

**Beispiel:**  $n = 4$

$$\omega = (1, 0, 1, 1)$$

$$P(\{(1, 0, 1, 1)\}) = p^3(1 - p)$$

Sei nun  $A \subset \Omega$ . Dann ist

$$P(A) := \sum_{\omega \in A} P(\{\omega\}).$$

Sei  $X(\omega) = \omega_1 + \omega_2 + \dots + \omega_n$  Dann ist wegen (+)

$$\begin{aligned} p_k &= P(\{\omega \mid X(\omega) = k\}) \\ &= \sum_{\omega \mid X(\omega) = k} P(\{\omega\}) \\ &\stackrel{(+)}{=} \sum_{\omega \mid X(\omega) = k} p^{\sum \omega_i} (1 - p)^{\sum (1 - \omega_i)} \\ &= \sum_{\omega \mid \sum \omega_i = k} p^k (1 - p)^{(n - k)} \\ &= \binom{n}{k} p^k (1 - p)^{n - k}, \end{aligned}$$

mit  $0 \leq k \leq n$ . Die letzte Gleichung folgt, da es  $\binom{n}{k}$  0-1 Folgen der Länge  $n$  gibt mit genau  $k$  Einsen.

Sei  $X_i(\omega) = \omega_i$  das Ereignis des  $i$ -ten Versuches. Da  $\omega_i$  entweder 0 oder 1 ist, sind  $X_i$  Bernoulli-verteilt mit

$$P(\{\omega \mid X_i(\omega) = \omega_i\}) = P(\{\omega_i\}) = \begin{cases} p & \text{für } \omega_i = 1 \\ 1 - p & \text{für } \omega_i = 0 \end{cases}$$

Nun ist  $X = \sum_{i=1}^n X_i$ . Dann folgt

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

Denn alle  $X_i$  haben denselben Erwartungswert  $p$ . Außerdem ist  $\text{Var}(X) = np(1 - p)$ .

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

$$= E\left(\left(\sum_{i=1}^n X_i\right)^2\right) - (np)^2$$

$$\begin{aligned} \text{Aber } E\left(\left(\sum_{i=1}^n X_i\right)^2\right) &= E\left[\left(\sum_{i=1}^n X_i\right)\left(\sum_{j=1}^n X_j\right)\right] \\ &= E\left(\sum_{i=1}^n X_i^2\right) + E\left(\sum_{i \neq j} X_i X_j\right) \\ &= E\left(\sum_{i=1}^n X_i^2\right) + \sum_{i \neq j} E(X_i X_j) \\ &= np + n(n-1)p^2, \quad \text{da } E(X_i X_j) = 1 \cdot p^2 = p^2 \text{ ist.} \end{aligned}$$

Insgesamt folgt  $\text{Var}(X) = np(1 - p)$  und  $\sigma(X) = \sqrt{np(1 - p)}$ .

Es besteht ein enger Zusammenhang zwischen der Binomialverteilung und dem Ziehen mit Zurücklegen. In einer Urne seien  $W$  weiße Kugeln und  $S$  schwarze Kugeln. Aus dieser Urne werde  $n$  mal eine Kugel gezogen und sofort wieder zurückgelegt. Die Wahrscheinlichkeit, daß bei einer einzelnen Ziehung eine weiße Kugel gezogen wird, ist

$$\frac{W}{W + S}$$

und die Wahrscheinlichkeit, daß eine schwarze Kugel gezogen wird, ist

$$\frac{S}{W + S}.$$

Aufgrund des Zurücklegens sind die einzelnen Versuche unabhängig und wir erhalten für die Anzahl der weißen Kugeln bei  $n$  Ziehungen

$$P(X = k) = \binom{n}{k} \left(\frac{W}{W + S}\right)^k \left(\frac{S}{W + S}\right)^{n-k}$$

für  $k = 0, 1, 2, \dots, n$  sowie

$$E(X) = n \frac{W}{W + S} \quad \text{und} \quad \text{Var}(X) = n \frac{W}{W + S} \frac{S}{W + S}.$$

### 3.3 Hypergeometrische Verteilung

In vielen Anwendungen, etwa bei der Qualitätskontrolle mittels Stichproben, wird ohne Zurücklegen gezogen. Wir betrachten deshalb nun die folgende Situation. In einer Urne seien  $W$  weiße Kugeln und  $S$  schwarze Kugeln. Aus dieser Urne werde  $n$  mal jeweils eine Kugel gezogen und danach *nicht* zurückgelegt. Offensichtlich muß dann  $n \leq W + S = N$  gelten. Sei  $X$  die Anzahl der weißen Kugeln in der gesamten Stichprobe. Wir bestimmen nun die Verteilung  $P(X = k)$  auf zwei Weisen.

Für die erste Herleitung denken wir uns die Kugeln von 1 bis  $N$  nummeriert und nehmen an, die Kugeln seien prinzipiell unterscheidbar. Die Kugeln mit den Nummern 1 bis  $W$  seien weiß und die Kugeln mit den Nummern  $W + 1$  bis  $N$  seien schwarz. Es gibt  $\binom{N}{n}$  verschiedene Möglichkeiten, aus den  $N$  Kugeln eine  $n$ -elementige Teilmenge auszuwählen. Um eine Stichprobe mit  $k$  weiße Kugeln zu erhalten, muß man  $k$  Elemente aus der Menge  $\{1, \dots, W\}$  auswählen und  $n - k$  Elemente aus der Menge  $\{W + 1, \dots, N\}$  auswählen. Dies ergibt nach dem Kombinationsprinzip

$$\binom{W}{k} \binom{S}{n-k}$$

viele Möglichkeiten. Damit erhalten wir

$$P(X = k) = \frac{\binom{W}{k} \binom{S}{n-k}}{\binom{W+S}{n}}$$

für alle  $k$  mit  $0 \leq k \leq W$  und  $0 \leq n - k \leq S$ . (Offensichtlich können wir nicht mehr als  $W$  weiße oder mehr als  $S$  schwarze Kugeln ziehen!) Die obige Verteilung von  $X$  heißt *Hypergeometrische Verteilung*. Für den Erwartungswert gilt (ohne Beweis)

$$E(X) = n \frac{W}{W + S}$$

und für die Varianz gilt (ohne Beweis)

$$\text{Var}(X) = n \frac{W}{W + S} \frac{S}{W + S} \left( \frac{W + S - n}{W + S - 1} \right).$$

Man sieht, daß der Erwartungswert derselbe ist wie beim *Ziehen mit Zurücklegen*. Die Varianz ist um den Faktor

$$\frac{W + S - n}{W + S - 1}$$

kleiner als beim *Ziehen mit Zurücklegen*. Dies erklärt sich aus der Abhängigkeit der einzelnen Ziehungen.

#### Beispiel:

Ein Kunde weiß, daß im Mittel 10% der Lieferungen von Bauteilen einer Firma defekt sind. Dieser Anteil ist vertraglich akzeptiert. Um sich gegen einen höheren Anteil an defekten Bauteilen abzusichern, entnimmt der Kunde jeder Einheit zu 50 Bauteilen eine Stichprobe vom Umfang 5 (selbstverständlich ohne Zurücklegen!) und testet diese 5 Bauteile. Der Kunde lehnt die Einheit ab, falls sich unter den fünf gezogenen mehr als ein defektes

Bauteil befindet. Mit welcher Wahrscheinlichkeit lehnt er eine Einheit ab, die 8 defekte Bauteile enthält?

Hier ist  $N = 50$ ,  $n = 5$ ,  $W = 8$ ,  $S = 42$ . Damit ist

$$\begin{aligned} P(X > 1) &= 1 - (P(X = 0) + P(X = 1)) \\ &= 1 - \frac{\binom{8}{0} \binom{42}{5}}{\binom{50}{5}} - \frac{\binom{8}{1} \binom{42}{4}}{\binom{50}{5}} \\ &= 1 - 0.4015 - 0.4226 \\ &= 0.1759. \end{aligned}$$

Wir wollen nun noch eine weitere Herleitung der Hypergeometrischen Verteilung betrachten. Dabei gehen wir ähnlich wie bei der Binomialverteilung vor. Es zeigt sich nämlich, daß – trotz der Abhängigkeit zwischen den Ziehungen – die Reihenfolge, in der weiße und schwarze Kugeln gezogen werden, *keine* Rolle spielt. Man sieht dies am besten an einem Beispiel: sei  $W = 3$ ,  $S = 5$  und  $n = 3$ . Die Wahrscheinlichkeit erst zwei weiße und dann eine schwarze Kugel zu ziehen ist

$$\frac{3}{8} \times \frac{2}{7} \times \frac{5}{6} = \frac{3 \times 2 \times 5}{8 \times 7 \times 6}.$$

Die Wahrscheinlichkeit erst eine weiße, dann eine schwarze und dann noch eine weiße Kugel zu ziehen ist

$$\frac{3}{8} \times \frac{5}{7} \times \frac{2}{6} = \frac{3 \times 5 \times 2}{8 \times 7 \times 6}.$$

Die zwei Ausdrücke ergeben denselben Wert. Offensichtlich hat sich nur die Reihenfolge der Zahlen im Zähler geändert. Eine Möglichkeit  $k$  weiße Kugeln zu erhalten besteht darin, erst  $k$  weiße Kugeln in Folge zu ziehen und anschließend  $n - k$  schwarze Kugeln in Folge zu ziehen. Die Wahrscheinlichkeit für dieses Elementarereignis ist

$$\frac{W}{W+S} \times \frac{W-1}{W+S-1} \times \cdots \times \frac{W-k+1}{W+S-k+1} \times \frac{S}{W+S-k} \times \frac{S-1}{W+S-k-1} \times \cdots \times \frac{S-(n-k)+1}{W+S-n+1}.$$

Die  $k$  weißen Kugeln und  $n - k$  schwarzen Kugeln können natürlich auch in einer anderen Reihenfolge auftreten; dies ändert jedoch nichts an der Wahrscheinlichkeit des betreffenden Elementarereignisses. Insgesamt gibt es  $\binom{n}{k}$  viele verschiedene Elementarereignisse mit  $k$  gezogenen weißen Kugeln. Damit erhalten wir

$$\begin{aligned} P(X = k) &= \binom{n}{k} \frac{\prod_{j=0}^{k-1} (W - j) \prod_{j=0}^{n-k-1} (S - j)}{\prod_{j=0}^{n-1} (W + S - j)} \\ &= \binom{n}{k} \frac{W!}{(W - k)!} \frac{S!}{(S - n + k)!} \frac{(W + S - n)!}{(W + S)!} \\ &= \frac{\binom{W}{k} \binom{S}{n-k}}{\binom{W+S}{n}}. \end{aligned}$$

Man rechnet die letzte Gleichung leicht direkt nach.

Ist  $N = W + S$  groß im Verhältnis zu  $n$ , so sollte der Unterschied zwischen Ziehen mit und Ziehen ohne Zurücklegen kaum bemerkbar sein. Wir betrachten nun den Grenzübergang  $S, W \rightarrow \infty$  mit

$$\frac{W}{W + S} \rightarrow p \in (0, 1).$$

Dann gilt auch  $S/(W + S) \rightarrow q$  und  $N \rightarrow \infty$ . Gleichzeitig lassen wir  $k$  und  $n$  fest. Terme der Form

$$\frac{W - j}{W + S - j}$$

streben dann für  $0 \leq j \leq k - 1$  gegen  $p$  und Terme der Form

$$\frac{S - j}{W + S - k - j}$$

streben für  $0 \leq j \leq n - k - 1$  gegen  $q$ . Damit strebt  $P(X = k)$  gegen

$$\binom{n}{k} p^k q^{n-k}.$$

Die Wahrscheinlichkeiten der Hypergeometrischen Verteilung streben also gegen die Wahrscheinlichkeiten der Binomialverteilung. Die folgende Tabelle veranschaulicht diese Grenzwertaussage. Dabei ist  $n = 5$ .

$k$	$W=20, S=10$	$W=100, S=50$	$W=200, S=100$	$W=1000, S=500$	$p = 2/3$
0	0.002	0.004	0.004	0.004	0.004
1	0.029	0.0039	0.040	0.041	0.041
2	0.160	0.164	0.164	0.165	0.165
3	0.360	0.335	0.332	0.330	0.329
4	0.340	0.331	0.330	0.329	0.329
5	0.109	0.127	0.129	0.131	0.132

### 3.4 Poisson-Verteilung und das Gesetz der kleinen Wahrscheinlichkeiten

Wir werden nun einen Grenzwertsatz für die Binomialverteilung kennenlernen. Sei  $X$  binomialverteilt. Also

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{für } 0 \leq k \leq n.$$

Wir betrachten in diesem Abschnitt den Fall  $p_n \rightarrow 0$  und  $n \rightarrow \infty$ , so daß  $np_n \rightarrow \mu$ . Wir haben also eine große Anzahl von Versuchen vorliegen mit jeweils sehr kleiner Erfolgswahrscheinlichkeit. Dies erklärt den Namen "Gesetz der kleinen Wahrscheinlichkeiten". Es gilt

$$P(X = 0) = (1 - p_n)^n = \left(1 - \frac{np_n}{n}\right)^n \approx \left(1 - \frac{\mu}{n}\right)^n \approx e^{-\mu}.$$

Dies folgt aus

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\mu}{n}\right)^n = e^{-\mu}.$$

Damit haben wir schon eine Näherungsformel für  $P(X = 0)$ . Um auch für die weiteren Wahrscheinlichkeiten eine Approximation zu erhalten, betrachten wir zunächst das Verhältnis

$$R(k) = \frac{P(X = k)}{P(X = k - 1)}.$$

Es gilt

$$R(k) = \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\binom{n}{k-1} p^{k-1} (1-p)^{n-k+1}} = \frac{\frac{n!}{k!(n-k)!} p}{\frac{n!}{(k-1)!(n-k+1)!} (1-p)} = \frac{n-k+1}{k} \frac{p}{1-p}.$$

Damit folgt bei  $n \rightarrow \infty$  und  $p_n \rightarrow 0$  mit  $np_n \rightarrow \mu$  (bei festem  $k$ )

$$R(k) = \frac{np_n}{k} \left(1 - \frac{k-1}{n}\right) \frac{1}{q_n} \rightarrow \frac{\mu}{k}.$$

Also  $R(k) \approx \mu/k$ . Damit gilt weiter

$$\begin{aligned} P(X = k) &= P(X = 0)R(1)R(2) \dots R(k) \\ &\approx e^{-\mu} \frac{\mu}{1} \frac{\mu}{2} \dots \frac{\mu}{k} \\ &= e^{-\mu} \frac{\mu^k}{k!}. \end{aligned}$$

Diese Näherung heißt Poisson-Näherung.

Die Zahlen  $e^{-\mu} \mu^k / k!$  sind für  $k = 0, 1, \dots$  positiv. Darüberhinaus gilt

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

Damit legen diese Zahlen eine Verteilung auf  $\{0, 1, 2, \dots\}$  fest. Diese Verteilung heißt Poisson-Verteilung. Eine Zufallsvariable  $Z$  heißt Poisson-verteilt mit Parameter  $\lambda$ , falls

$$P(Z = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

für  $k = 0, 1, 2, \dots$  gilt. Der Erwartungswert  $E(Z)$  ist dann gleich

$$E(Z) = \lambda.$$

Diesen erhält man wie folgt:

$$E(Z) = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Ähnlich sieht man  $E(Z(Z-1)) = \lambda^2$  und erhält hieraus  $E(Z^2) = \lambda^2 + \lambda$  und schließlich

$$\text{Var}(Z) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Bei der Poisson-Verteilung stimmen also Erwartungswert und Varianz überein.

**Beispiel:** Defekte in einer Produktionseinheit

Ein Produktionsprozess führt zu 1% defekten Einheiten. Wie groß ist die Wahrscheinlichkeit 2 oder mehr defekte unter 100 produzierten Teilen zu finden? Die Verteilung der Defekten ist binomialverteilt mit  $n = 100$  und  $p = \frac{1}{100}$ :

$$\begin{aligned} P(0 \text{ oder } 1 \text{ Defektes}) &= \left(1 - \frac{1}{100}\right)^{100} + 100 \cdot \frac{1}{100} \left(1 - \frac{1}{100}\right)^{99} \\ &= \left(\frac{99}{100}\right)^{100} + \left(\frac{99}{100}\right)^{99} \\ &= 0,36603234 + 0,36972964 \\ &= 0,735762 \end{aligned}$$

Näherungsweise ist die Verteilung der Defekten poissonisch mit Parameter 1. Dann ist

$$P(0 \text{ oder } 1 \text{ Defektes}) = 2 \cdot e^{-1} = 0,735758$$

Damit erhält man in beiden Rechnungen

$$P(2 \text{ oder mehr Defekte}) = 0,26424.$$

Der Unterschied in der Approximation erscheint erst in der 6. Stelle hinter dem Komma.

Ein Datenbeispiel, für das die Poisson-Verteilung sehr gut paßt, ist die Statistik der Hufschlagtoten in der Preußischen Armee von 1875 bis 1894. Dabei ergeben sich 196 Tote in 280 Regimentsjahren. Dies ergibt als beste Anpassung eine Poisson-Verteilung mit  $\mu = \frac{196}{280} = 0,7$ .

**Number of Deaths by Horsekicks  
in the Prussian Army  
from 1875-1894 for 14 Corps**

Year	G*	I*	II	III	IV	V	VI*	VII	VIII	IX	X	XI*	XIV	XV	Total
1875								1	1				1		3
1876	2				1								1	1	5
1877	2						1	1			1		2		7
1878	1	2	2	1	1						1		1		9
1879				1	1	2	2		1			2	1		10
1880		3	2	1	1	1				2	1	4	3		18
1881	1			2	1			1		1					6
1882	1	2					1		1	1	2	1	4	1	14
1883			1	2		1	2	1		1		3			11
1884	3		1					1			2		1	1	9
1885							1			2		1		1	5
1886	2	1			1	1	1			1		1	3		11
1887	1	1	2	1			3	2	1	1		1	2		15
1888		1	1			1	1					1	1		6
1889			1	1		1	1			1	2	2		2	11
1890	1	2		2		1	1	2		2	1	1	2	2	17
1891				1	1	1		1	1		3	3	1		12
1892	1	3	2		1	1	3		1	1		1	1		15
1893		1				1		2			1	3			8
1894	1								1		1	1			4
Total	16	16	12	12	8	11	17	12	7	13	15	25	24	8	196

\* G indicates Guard Corps

G, I, VI and XI Corps' organization differ from the others

### 3.5 Gesetz der Großen Zahlen

**Definition 3.5.1** Die Zufallsvariablen  $X_1, X_2, \dots, X_n$  mit dem diskreten Wertebereich  $X(\Omega) = \{z_1, z_2, \dots\}$  heißen unabhängig, falls für alle Werte  $z_1, z_2, \dots, z_n$  aus  $X(\Omega)$  gilt

$$P(X_1 = z_1, \dots, X_n = z_n) = \prod_{j=1}^n P(X_j = z_j)$$

**Beispiel:**

Sei  $\Omega = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_i \in \{0, 1\}\}$ ,  $X_i(\omega) = \omega_i$  und

$$P(\{\omega \mid X_i(\omega) = 1\}) = p, \quad P(\{\omega \mid X_i(\omega) = 0\}) = 1 - p.$$

Sind  $X_1, X_2, \dots, X_n$  unabhängig, so folgt mit  $\varepsilon_i \in \{0, 1\}$  für  $i = 1, \dots, n$

$$P(X_i = \varepsilon_i, i = 1, \dots, n) = \prod_{i=1}^n P(X_i = \varepsilon_i) = p^{\sum \varepsilon_i} (1-p)^{\sum (1-\varepsilon_i)} = p^k (1-p)^{n-k},$$

falls  $\sum_{i=1}^n \varepsilon_i = k$  ist. Dann ist  $X = X_1 + \dots + X_n$  binomialverteilt mit Parametern  $n$  und  $p$  (siehe Abschnitt 3.3).

Es gilt nun die **folgende Stabilitätseigenschaft**:

Sind  $X$  und  $Y$  unabhängig und binomialverteilt mit Parametern  $n, p$  bzw.  $m, p$ , so ist  $X + Y$  binomialverteilt mit Parametern  $n + m, p$ . Dies sieht man so:

$$\begin{aligned} P(X + Y = l) &= P(\cup_{i=0}^l \{X = i, Y = l - i\}) \\ &= \sum_{i=0}^l P(\{X = i, Y = l - i\}) \\ &= \sum_{i=0}^l P(\{X = i\})P(\{Y = l - i\}) \\ &= \sum_{i=0}^l \binom{n}{i} p^i (1-p)^{n-i} \binom{m}{l-i} p^{l-i} (1-p)^{m-l+i} \\ &= p^l (1-p)^{n+m-l} \sum_{i=0}^l \binom{n}{i} \binom{m}{l-i} \\ &= p^l (1-p)^{n+m-l} \binom{n+m}{l} \end{aligned}$$

Eine entsprechende Aussage gilt für Poisson-verteilte Zufallsvariablen.

Betrachten wir nun nochmals binomialverteilte Zufallsvariablen  $S_n$  mit  $P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$ . Der Erwartungswert von  $S_n$  ist  $E(S_n) = np$ . Folglich ist  $E(\frac{S_n}{n}) = p$ . Außerdem ist  $\text{Var}(\frac{S_n}{n}) = p(1-p)/n$ ; d.h. sie geht gegen 0 für  $n \rightarrow \infty$ . Folglich vermutet man  $S_n/n \rightarrow p$ . Dies ist ein Spezialfall des folgenden grundlegenden Sachverhaltes.

### Das Gesetz der Großen Zahlen:

- 1.) Sei  $n \in \mathbb{N}$ . Seien  $X_1, X_2, \dots, X_n$  unabhängige Zufallsvariablen mit derselben Verteilung und endlichem Erwartungswert  $E(X_1)$  und endlicher Varianz  $\text{Var}(X_1)$ . Dann gilt für jedes  $\varepsilon > 0$

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - E(X_1)\right| > \varepsilon\right) \leq \frac{\text{Var}(X_1)}{n\varepsilon^2}.$$

- 2.) Es gilt:

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

für alle  $\varepsilon > 0$ .

### Bemerkungen:

1. Sind die Zufallsvariablen  $X_1, X_2, \dots, X_n$  Bernoulli-verteilt mit Parameter  $p$ , so gilt  $E(X_1) = p$  und man hat

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = p.$$

2. Speziell beim Würfeln setze man

$$X_i = \begin{cases} 1 & \text{falls } Y_i = 6 \\ 0 & \text{falls } Y_i \neq 6 \end{cases}$$

wenn  $Y_i$  das Ergebnis des  $i$ -ten Wurfes bedeutet. Dann ist  $p = 1/6$  und man hat

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \rightarrow \frac{1}{6};$$

in Worten: Die relative Häufigkeit der Sechsen in  $n$  Würfeln konvergiert für  $n \rightarrow \infty$  gegen  $1/6$ .

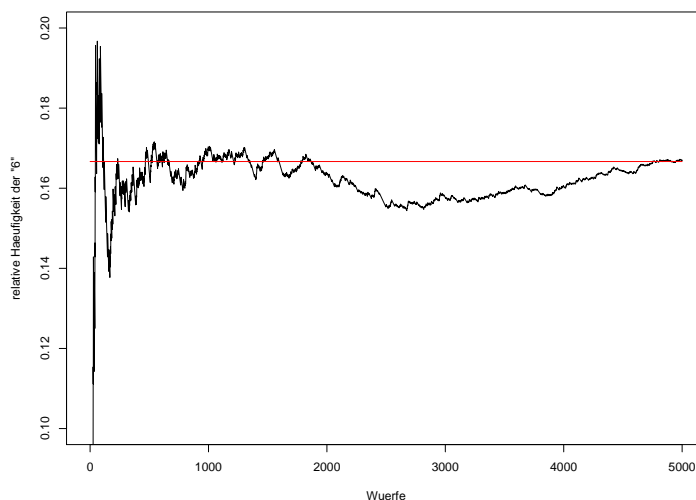


Abbildung 3.1: relative Häufigkeit

### Die Empirische Verteilung

Seien  $X_1, X_2, \dots, X_n$  unabhängige Zufallsvariablen mit gleicher Verteilung  $Q$  und Wertebereich  $X(\Omega) = \{z_1, z_2, \dots, z_k\}$ . Sei  $F_n(z) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i=z\}}$  für  $z \in X(\Omega)$ . Dann gilt für  $n \rightarrow \infty$ :  $F_n(z) \rightarrow Q(z)$  für alle  $z \in X(\Omega)$ .

$F_n(z)$  heißt *empirische Verteilung*, in Englisch: sample distribution.

Beachte:  $\sum_z F_n(z) = 1$ .

**Beweis:**

Da die  $X_i$  unabhängig sind, sind auch  $1_{\{X_i=z\}}$ ,  $i = 1, \dots, n$  unabhängig. Das Gesetz der Großen Zahlen liefert:  $F_n(z) \rightarrow E1_{\{X_1=z\}} = P(X_1 = z) = Q(z)$ .

□

**Beispiel: Würfeln!**

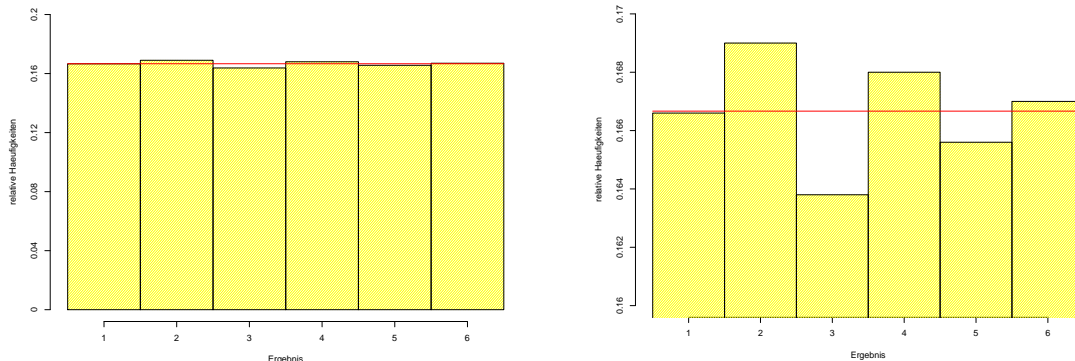


Abbildung 3.2: Empirische Verteilung bei n = 5000

**Tschebychev-Ungleichung:**

Sei  $Y$  Zufallsvariable mit endlichem Erwartungswert  $E(Y)$  und endlicher Varianz  $\text{Var}(Y)$ . Dann gilt für jedes  $a > 0$

$$P(|Y - E(Y)| > a) \leq \frac{\text{Var}(Y)}{a^2}$$

**Denn:**

$$\begin{aligned} a^2 P(|Y - E(Y)| > a) &\leq a^2 E1_{\{|Y-E(Y)|>a\}} \\ &\leq E[(Y - E(Y))^2 1_{\{|Y-E(Y)|>a\}}] \\ &\leq \text{Var}(Y) \end{aligned}$$

Dabei verwendet man, daß  $P(Y \in A) = E1_A(Y)$  gilt. Dies sieht man leicht ein, denn

$$P(Y \in A) = \sum_{y_i \in A} P(Y = y_i) = \sum_{y_i \in \Omega} 1_A(y_i) P(Y = y_i) = E1_A(Y).$$

**Beweis des GGZ:**

Wende Tschebychev-Ungleichung auf  $\bar{X}_n = S_n/n$  an ! Dann gilt

$$\begin{aligned} E(\bar{X}_n) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} \sum_{i=1}^n E(X_i) \\ &= \frac{nE(X_1)}{n} = E(X_1) \end{aligned}$$

Außerdem ist

$$\begin{aligned}
\text{Var}(\bar{X}_n) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\
&= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \\
&= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \\
&= \frac{n \text{Var}(X_1)}{n^2} = \frac{\text{Var}(X_1)}{n}
\end{aligned}$$

Die dritte Gleichung folgt mit der Formel von Bienaymé. Sie wird später bewiesen. Einsetzen liefert

$$P(|\bar{X}_n - E(X_1)| \geq \varepsilon) \leq \frac{\text{Var}(X_1)}{\varepsilon^2 n}$$

□

**Definition 3.5.2**  $X$  und  $Y$  seien Zufallsvariablen mit  $E(X^2) < \infty$  und  $E(Y^2) < \infty$ .

1. Dann heißt  $\text{Kov}(X, Y) = E((X - E(X))(Y - E(Y)))$  die Kovarianz von  $X$  und  $Y$ .

2. Die Größe

$$\rho(X, Y) = \frac{\text{Kov}(X, Y)}{\sigma(X)\sigma(Y)}$$

heißt Korrelationskoeffizient.

3.  $X$  und  $Y$  heißen unkorreliert, falls  $\text{Kov}(X, Y) = 0$ .

**Bemerkungen:**

1. Für die Kovarianz gilt die Verschiebungsformel

$$\text{Kov}(X, Y) = E(X \cdot Y) - E(X)E(Y).$$

2. Sind  $X$  und  $Y$  unabhängig, so gilt  $\text{Kov}(X, Y) = 0$  oder äquivalent

$$E(X \cdot Y) = E(X)E(Y).$$

**Denn:**

$$\begin{aligned}
E(X \cdot Y) &= \sum_{i,j} x_i y_j P(X = x_i, Y = y_j) \\
&= \sum_{i,j} x_i y_j P(X = x_i) P(Y = y_j) \\
&= \left(\sum_i x_i P(X = x_i)\right) \left(\sum_j y_j P(Y = y_j)\right) \\
&= E(X) \cdot E(Y).
\end{aligned}$$

**Beachte:**

Es kann  $\text{Kov}(X, Y) = 0$  sein, ohne daß  $X$  und  $Y$  unabhängig sind.

3. Mit einer ähnlichen Rechnung wie in Abschnitt 3.2 erhält man

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Kov}(X_i, X_j).$$

Sind  $X_1, \dots, X_n$  unabhängig, ist der Kovarianz-Term gleich 0 und es gilt die Formel von Bienaymé

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i).$$

## 3.6 Der zentrale Grenzwertsatz

Seien  $X_1, X_2, \dots, X_n$  unabhängige Zufallsvariablen, alle mit derselben Verteilung und mit endlichem Erwartungswert  $E(X_1) = \mu$  und endlicher Varianz  $\text{Var}(X_1) = \sigma^2$ . Sei  $S_n = \sum_{i=1}^n X_i$ . Wir interessieren uns für das Verhalten von

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n}\sigma} \quad \text{für wachsendes } n.$$

Es gilt  $E(S_n^*) = 0$  und  $\text{Var}(S_n^*) = 1$ . Man sagt  $S_n^*$  ist standardisiert. Nun gilt:

$$\lim_{n \rightarrow \infty} P(a \leq S_n^* \leq b) = \Phi(b) - \Phi(a) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

Diese Aussage wird ‘‘zentraler Grenzwertsatz’’ genannt. Sie besagt, daß für große  $n$   $S_n^*$  näherungsweise normalverteilt ist.

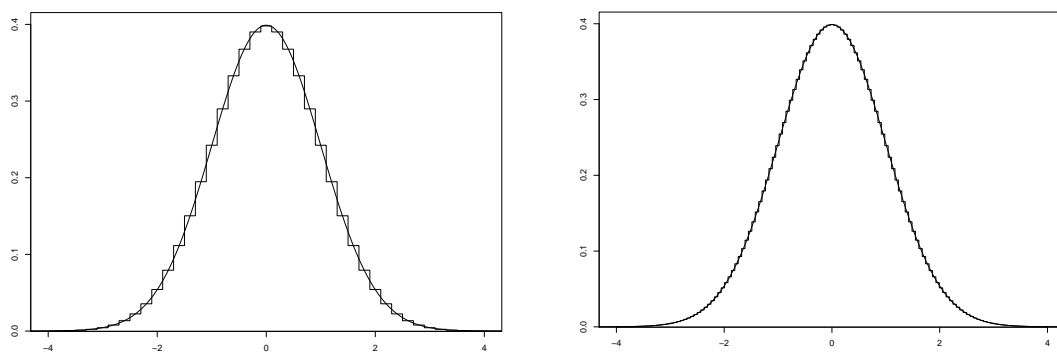


Abbildung 3.3: Approximation der Binomialverteilung durch die Normalverteilung  
( $n=100$ ,  $n=1000$ )

Anschaulich heißt dies, die Fläche unter dem Histogramm der Verteilung von  $S_n^*$  zwischen  $a$  und  $b$  konvergiert für  $n \rightarrow \infty$  gegen die Fläche unter der Kurve von  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$  zwischen  $a$  und  $b$ .  $\varphi$  heißt Gaußsche Glockenkurve (Sie findet sich auch auf der Vorderseite unseres letzten 10 DM Scheins).



Abbildung 3.4: 10 DM Schein

Die Funktion  $\Phi(z) = \int_{-\infty}^z \varphi(x) dx$  ist gut tabelliert und in vielen hochwertigen Taschenrechnern festverdrahtet. Sie läßt sich nicht elementar berechnen. Es gilt  $\Phi(-z) = 1 - \Phi(z)$  für  $z > 0$ . Dies folgt direkt aus der Symmetrie von  $\varphi(x)$  um  $1/2$ .

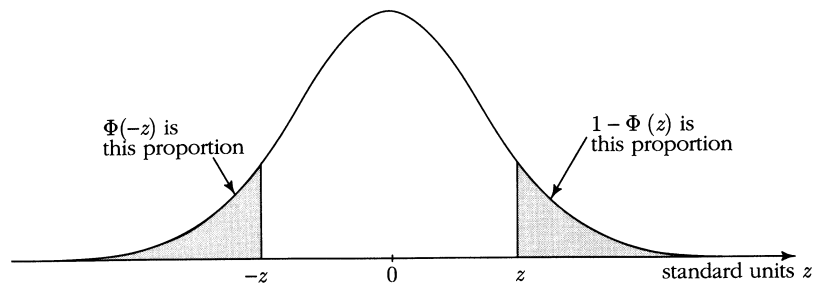


Abbildung 3.5: Symmetrie der Normalverteilung

Es gilt  $\Phi(z) - \Phi(-z) = 2\Phi(z) - 1$ .

Hier sind fünf Werte dieser Funktion, die man sich für Rechenzwecke merken kann.

$z$	$\Phi(z) - \Phi(-z)$
1	0,683
2	0,954
3	0,997
4	$6,3 \cdot 10^{-5}$
5	$5,7 \cdot 10^{-7}$

Eine altbekannter Spezialfall des zentralen Grenzwertsatzes ist die Approximation der Binomialverteilung für den Fall nicht seltener Ereignisse:

### Der Satz von de Moivre-Laplace

Seien  $X_1, X_2, \dots, X_n$  unabhängige Bernoulli-verteilte Zufallsvariablen mit Parameter  $p$ .

Sei  $S_n = \sum_{i=1}^n X_i$ . Dann gilt für  $-\infty \leq a < b \leq \infty$

$$\lim_{n \rightarrow \infty} P \left( a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b \right) = \Phi(b) - \Phi(a).$$

**Bemerkung:**

Beachte  $S_n$  ist binomialverteilt. Der Satz wird oft so verwendet:

$$(+) \quad P(a \leq S_n \leq b) = \sum_{k=a}^b \binom{n}{k} p^k (1-p)^{n-k} \approx \Phi \left( \frac{b - np}{\sqrt{np(1-p)}} \right) - \Phi \left( \frac{a - np}{\sqrt{np(1-p)}} \right).$$

Dazu beachte man:  $S_n \leq b \Leftrightarrow \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{b - np}{\sqrt{np(1-p)}}$ .

Für kleine  $n$  ist die Approximation (+) nicht sehr gut und es empfiehlt sich eine Stetigkeitskorrektur, die darin besteht die approximierende Normalverteilung der Kästchenbreite anzupassen.

$$P(a \leq S_n \leq b) \approx \Phi \left( \frac{b + \frac{1}{2} - \mu}{\sigma} \right) - \Phi \left( \frac{a - \frac{1}{2} - \mu}{\sigma} \right)$$

mit  $\mu = np$  und  $\sigma^2 = np(1-p)$ .

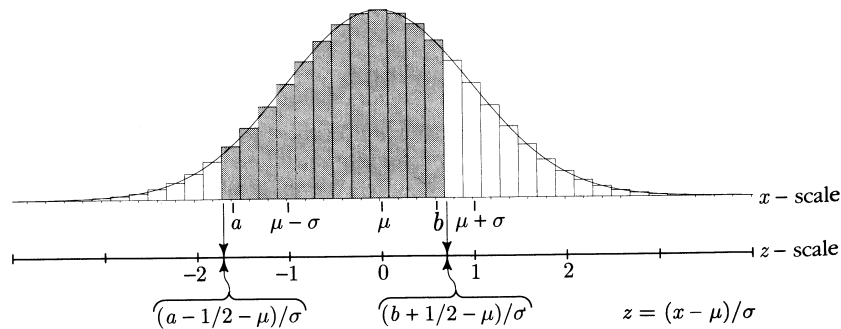


Abbildung 3.6

Den Vorteil zeigt das folgende extreme Beispiel:

$$n = 100, p = \frac{1}{2}$$

$$P(S_n = 50) \approx \Phi\left(\frac{50 + \frac{1}{2} - 50}{5}\right) - \Phi\left(\frac{50 - \frac{1}{2} - 50}{5}\right) = \Phi(0,1) - \Phi(-0,1) = 0,0796$$

Der exakte Wert ist 0,07959, während die Approximation mit (+) einfach 0 ergibt.

Es folgt nun ein Beispiel, bei dem  $n$  so groß ist, daß man die Stetigkeitskorrektur außer acht lassen kann.

**Beispiel 1:**

1) Wie groß ist die Wahrscheinlichkeit, daß in 6000 Würfeln eines Würfels die "6"

- a) mehr als 1100 mal auftritt,
- b) mindestens 950 mal und höchstens 1050 mal auftritt?

$$\begin{aligned} \text{Zu a)} \quad P(S_{6000} > 1100) &= P(S_{6000} - 1000 > 100) \\ &= P \left( \frac{S_{6000} - 1000}{\sqrt{6000 \cdot \frac{1}{6} \cdot \frac{5}{6}}} > \frac{100}{\sqrt{6000 \cdot \frac{1}{6} \cdot \frac{5}{6}}} \right) \\ &= P(S_{6000}^* > \sqrt{12}) \\ &\approx 1 - \Phi(\sqrt{12}) = 0,00028 \end{aligned}$$

$$\begin{aligned} \text{Zu b)} \quad P(|S_{6000} - 1000| \leq 50) &= P\left(|S_{6000}^*| \leq \frac{50}{\sqrt{\frac{5000}{6}}}\right) \\ &\approx [\Phi(\sqrt{3}) - \Phi(-\sqrt{3})] = 0,917 \end{aligned}$$

### Beispiel 2:

Wieviele Wähler muß man befragen, um mit einer Abweichung von höchstens 1% das Wahlergebnis einer Partei  $A$  mit 90% Sicherheit vorherzusagen?

Hier ist das Binomialmodell angebracht. Sei  $\hat{p}_n = \frac{S_n}{n}$  der Anteil der Befürworter der Partei  $A$  unter den Befragten.  $S_n$  ist binomialverteilt mit Parameter  $n$  und  $p_0$ .  $p_0$  soll bis auf 1% bestimmt werden, d.h.

$$(*) \quad P_{p_0}(|\hat{p}_n - p_0| \leq 0,01) \geq 0,90.$$

Wie groß muß  $n$  sein, damit Aussage (\*) für alle  $p_0$  gilt?

Sei  $p_1 = p_0 - 0,01$ ,  $p_2 = p_0 + 0,01$ , dann gilt

$$\begin{aligned} P_{p_0}(p_1 \leq \hat{p}_n \leq p_2) &= P_{p_0}\left(\frac{\sqrt{n}(p_1 - p_0)}{\sqrt{p_0(1 - p_0)}} \leq S_n^* \leq \frac{\sqrt{n}(p_2 - p_0)}{\sqrt{p_0(1 - p_0)}}\right) \\ &\approx \Phi\left(\frac{\sqrt{n}(p_2 - p_0)}{\sqrt{p_0(1 - p_0)}}\right) - \Phi\left(\frac{\sqrt{n}(p_1 - p_0)}{\sqrt{p_0(1 - p_0)}}\right) \\ &= 1 - 2\Phi\left(-\frac{\sqrt{n}(p_2 - p_0)}{\sqrt{p_0(1 - p_0)}}\right) \\ &= 1 - 2\Phi\left(-\frac{\sqrt{n} \cdot 0,01}{\sqrt{p_0(1 - p_0)}}\right) \stackrel{!}{=} 0,9 \end{aligned}$$

Dies ist äquivalent zu :

$$0,1 = 2\Phi\left(-\frac{\sqrt{n} \cdot 0,01}{\sqrt{p_0(1 - p_0)}}\right)$$

Nun folgt

$$-\Phi^{-1}(0,05) = \frac{\sqrt{n} \cdot 0,01}{\sqrt{p_0(1 - p_0)}}$$

und

$$\Phi^{-1}(0,05)^2 = \frac{n \cdot 0,01^2}{p_0(1 - p_0)}.$$

Damit muß gelten:

$$n_0 \geq (\Phi^{-1}(0,05)^2 / 0,01^2) p_0(1 - p_0)$$

Beachte:  $p(1 - p) \leq \frac{1}{4}$  mit Gleichheit für  $p = \frac{1}{2}$ ,  $\Phi(0,05) = -1,645$ .

Für  $p_0 = \frac{1}{2}$  ist  $n_0 = 6764$ , für  $p_0 = 0,1$  ist  $n_0 = 2435$ .

# Kapitel 4

## Zufallsvariable und Wahrscheinlichkeitsdichten

### 4.1 Dichten und Verteilungsfunktionen

Bis jetzt haben wir nur diskrete Verteilungen mit Punktmassen kennengelernt. Viele Vorgänge – etwa die (verrauschte) Messung eines Signals – lassen sich besser durch kontinuierliche Verteilungen beschreiben. Um solche Verteilungen einzuführen, ist der folgende Begriff hilfreich.

**Definition 4.1.1** Eine Zufallsvariable  $X$  ist eine meßbare Abbildung  $X : \Omega \rightarrow \mathbb{R}$ . Die Funktion  $F : \mathbb{R} \rightarrow [0, 1]$  mit

$$F(x) = P(X \leq x) = P(\{\omega | X(\omega) \leq x\})$$

heißt die Verteilungsfunktion von  $X$ . Meßbar heißt dabei, daß die rechte Seite der Gleichung für alle  $x$  erklärt ist.

Eigenschaften einer Verteilungsfunktion  $F$ :

- i)  $F$  ist wachsend.
- ii)  $F$  ist rechtsseitig stetig.
- iii)  $\lim_{x \rightarrow -\infty} F(x) = 0$  und  $\lim_{x \rightarrow \infty} F(x) = 1$ .

Es gilt

$$P(a < X \leq b) = F(b) - F(a)$$

und

$$P(X > a) = 1 - F(a).$$

**Definition 4.1.2** Sei  $X$  eine Zufallsvariable. Dann heißt  $X$  stetig verteilt mit Dichte  $f$ , falls für die Verteilungsfunktion  $F$  von  $X$

$$F(x) = \int_{-\infty}^x f(u) du$$

gilt. Die Funktion  $f$  heißt Wahrscheinlichkeitsdichte (Dichtefunktion) von  $X$ .

Eigenschaften einer Dichtefunktion  $f$ :

i)  $\int_{-\infty}^{\infty} f(u)du = 1.$

ii)  $f(u) \geq 0$  für alle  $u \in \mathbf{R}.$

Es gilt

$$P(a < X \leq b) = \int_a^b f(u)du$$

und damit

$$P(X = b) = \lim_{h \downarrow 0} P(b - h < X \leq b) = \lim_{h \downarrow 0} \int_{b-h}^b f(u)du = 0.$$

Es gibt also keine Punktmassen. Die Dichte  $f$  läßt sich (falls  $f$  stetig ist) folgendermaßen verstehen:

$$P(x < X \leq x + h) = \int_x^{x+h} f(u)du \approx hf(x).$$

Man schreibt dafür gerne mit Differentialen  $P(X \in dx) = f(x)dx$ . Ist die Dichte  $f$  in einer Umgebung von  $x = x_0$  stetig, so gilt

$$f(x_0) = F'(x_0).$$

Für stetige Dichten  $f$  ist die zugehörige Verteilungsfunktion  $F$  folglich eine Stammfunktion.

**Definition 4.1.3** Sei  $X$  eine stetig verteilte Zufallsvariable mit Dichte  $f$ . Dann heißt

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

der Erwartungswert von  $X$ , falls

$$\int_{-\infty}^{\infty} |x|f(x)dx < \infty.$$

**Definition 4.1.4** Sei  $X$  eine stetig verteilte Zufallsvariable mit Dichte  $f$  und Erwartungswert  $E(X) = \mu$ . Dann heißt

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$$

die Varianz von  $X$ . Ist  $\text{Var}(X) < \infty$ , so heißt  $\sigma(X) = \sqrt{\text{Var}(X)}$  die Standardabweichung von  $X$ .

Auch hier gilt die Verschiebungsformel

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

Ist  $g$  eine Funktion von  $\mathbf{R}$  nach  $\mathbf{R}$  und  $X$  eine stetig verteilte Zufallsvariable mit Dichte  $f$ , so ist  $g(X)$  Zufallsvariable und der Erwartungswert  $E(g(X))$  von  $g(X)$  durch

$$\int_{-\infty}^{\infty} g(x)f(x)dx$$

gegeben, falls

$$\int_{-\infty}^{\infty} |g(x)|f(x)dx < \infty$$

gilt.

**Beispiel:** Gleichverteilung auf  $[a, b]$

Hier ist  $f(x) = \frac{1}{b-a}1_{[a,b]}(x)$ , falls  $a < b$  ist. Es gilt:

$$F(x) = \frac{x-a}{b-a} \quad \text{für } a \leq x \leq b$$

$$F(x) = 0 \quad \text{für } x < a$$

$$F(x) = 1 \quad \text{für } x > b.$$

$$\begin{aligned} E(X) &= \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \frac{1}{2} x^2 \Big|_a^b \\ &= \frac{b^2 - a^2}{2(b-a)} \\ &= \frac{a+b}{2} \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= \frac{1}{b-a} \int_a^b x^2 dx - \left( \frac{a+b}{2} \right)^2 \\ &= \frac{\frac{1}{3}(b^3 - a^3)}{(b-a)} - \left( \frac{a+b}{2} \right)^2 \\ &= \frac{1}{12}(b-a)^2. \end{aligned}$$

## 4.2 Die Normalverteilung

Eine stetige Zufallsvariable  $X$  heißt normalverteilt mit Mittelwert  $\mu$  und Varianz  $\sigma^2$  (kurz  $N(\mu, \sigma^2)$ -verteilt), falls für die zugehörige Dichte  $f$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

gilt. Dabei ist  $\pi = 3.141592\dots$  die Kreiszahl und  $e$  die Eulersche Konstante  $e = 2.718281\dots$ . Ist  $X$   $N(\mu, \sigma^2)$ -verteilt, so ist  $aX + b$  gemäß  $N(a\mu + b, a^2\sigma^2)$ -verteilt. Insbesondere ist  $Z$  mit

$$Z = \frac{X - \mu}{\sigma}$$

$N(0, 1)$ -verteilt. Diese Verteilung heißt Standard Normalverteilung. Die Dichte der Standard Normalverteilung wird üblicherweise mit  $\phi$  bezeichnet.

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

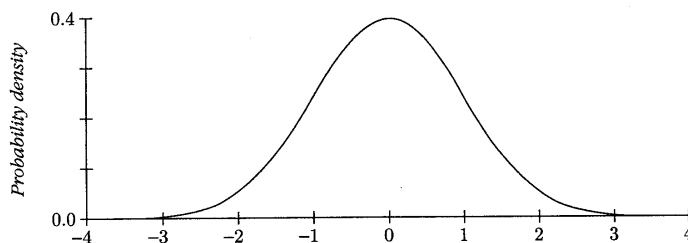


Abbildung 4.1: Die Standard Normal Dichte

Die zugehörige Verteilungsfunktion wird mit  $\Phi$  bezeichnet. Also

$$\Phi(x) = \int_{-\infty}^x \phi(u) du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du.$$

Es gilt

$$\Phi(-x) = 1 - \Phi(x)$$

und

$$P(-z \leq Z \leq z) = 2\Phi(z) - 1.$$

Die Funktion  $\Phi$  läßt sich nicht explizit darstellen, sondern muß numerisch berechnet werden. Es gilt

$$P(-1 \leq Z \leq 1) \approx 0.68,$$

$$P(-2 \leq Z \leq 2) \approx 0.95$$

und

$$P(-3 \leq Z \leq 3) \approx 0.997.$$

Wie schon oben erwähnt, ist sie überall tabelliert und in jedem guten Taschenrechner verfügbar. Ist  $X$   $N(\mu, \sigma^2)$ -verteilt, so gilt für die zugehörige Verteilungsfunktion  $F$

$$F(x) = P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Damit erhalten wir insbesondere die sogenannten 1- $\sigma$ , 2- $\sigma$  und 3- $\sigma$  Regeln:

$$P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.68,$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95$$

und

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.997.$$

Wie bereits früher gezeigt, spielt die Normalverteilung als approximierende Verteilung eine zentrale Rolle.

### 4.3 Exponential- und Gamma-Verteilung

Wir wollen nun ein wichtiges Beispiel für eine stetige Verteilung mit Dichte näher kennenlernen. Eine Zufallsvariable  $T$  heißt exponentialverteilt mit Parameter  $\lambda > 0$ , falls sie die Dichte

$$f(t) = \begin{cases} 0 & \text{für } t \leq 0 \\ \lambda e^{-\lambda t} & \text{für } t > 0 \end{cases}$$

besitzt. Für die Verteilungsfunktion  $F$  von  $T$  gilt dann

$$F(t) = \int_{-\infty}^t f(u) du = \begin{cases} 0 & \text{für } t \leq 0 \\ \lambda \int_0^t e^{-\lambda u} du & \text{für } t > 0. \end{cases}$$

Damit erhalten wir

$$F(t) = \begin{cases} 0 & \text{für } t \leq 0 \\ 1 - e^{-\lambda t} & \text{für } t > 0. \end{cases}$$

Für den Erwartungswert von  $T$  gilt

$$E(T) = \frac{1}{\lambda}.$$

Die Varianz ist gleich  $\frac{1}{\lambda^2}$ .

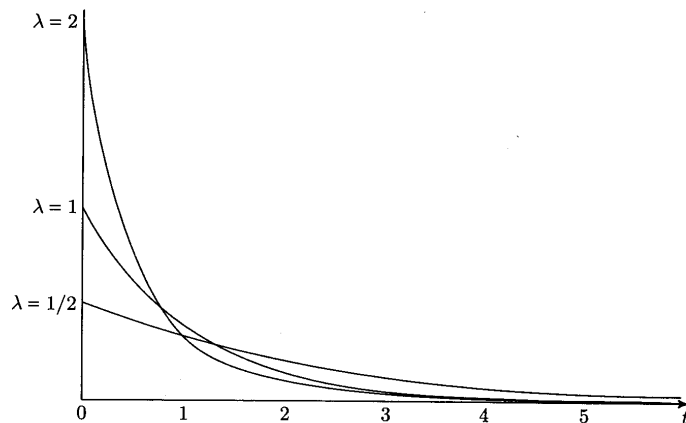


Abbildung 4.2: Die Exponentialdichten für  $\lambda = 0.5, 1, 2$

Wir wollen nun die Bedeutung von  $\lambda$  überlegen. Sei  $T$  die Lebensdauer einer technischen Komponente und sei  $T$  exponentialverteilt mit Parameter  $\lambda$ . Dann ist

$$\begin{aligned} P(T \leq t + \Delta \mid T > t) &= 1 - P(T > t + \Delta \mid T > t) \\ &= 1 - e^{-\lambda \Delta} \\ &= 1 - \left[ 1 - \lambda \Delta + \frac{1}{2} \lambda^2 \Delta^2 + \dots \right] \\ &\cong \lambda \Delta \quad \text{für } \Delta \text{ klein.} \end{aligned}$$

Dies besagt, daß die Ausfallrate für kleines  $\Delta$  näherungsweise proportional ist zu  $\lambda$ .

**Beispiel 1:**

Die mittlere Lebenszeit eines Transistors ist 100 Stunden. Was ist die Wahrscheinlichkeit, daß der Transistor länger als 50 Stunden hält?

$$\frac{1}{\lambda} = 100, \quad \lambda = 0,01$$

$$P(T > 50) = e^{-\lambda \cdot 50} = e^{-0,5} = 0,606.$$

**Beispiel 2:** "Radioaktiver Zerfall"

Jedes Atom hat unabhängig von den anderen eine exponentialverteilte Lebensdauer  $T$ .

$$P(T > t) = e^{-\lambda t}$$

Die Halbwertszeit einer radioaktiven Substanz ist diejenige Zeit  $h$ , für die

$$e^{-\lambda h} = \frac{1}{2} \quad \text{oder} \quad h = \frac{\log(2)}{\lambda}$$

gilt.

**Beispiel:** Strontium 90,  $h=28$  Jahre

$$\lambda = \log(2)/h = 0,0248/\text{Jahr} \quad \frac{1}{\lambda} = 40,4 \text{ Jahre.}$$

Die Wahrscheinlichkeit bei Strontium 90, daß ein Atom mehr als 50 Jahre nicht zerfällt, ist

$$P(T > t) = e^{-\lambda \cdot 50} = 0,29.$$

Dies ist auch der Anteil von Strontium 90, der nach 50 Jahren noch vorhanden ist. Dies folgt aus dem Gesetz der Großen Zahlen. Bis 99 % von Strontium 90 zerfallen ist, vergehen 186 Jahre.

Die Exponentialverteilung ist durch die Eigenschaft der sogenannten „Gedächtnislosigkeit“ ausgezeichnet. Damit ist folgendes gemeint: Für  $0 < s < t$  gilt

$$\begin{aligned} P(T > t \mid T > s) &= \frac{P(T > t)}{P(T > s)} = \frac{1 - F(t)}{1 - F(s)} \\ &= \frac{e^{-\lambda t}}{e^{-\lambda s}} = e^{-\lambda(t-s)} \\ &= P(T > t - s). \end{aligned}$$

Nehmen wir etwa an, daß die Brenndauer von Glühbirnen einer bestimmten Sorte einer Exponentialverteilung folgt, so ist  $P(T > t \mid T > s)$  die bedingte Wahrscheinlichkeit, daß eine Glühbirne, die seit dem Zeitpunkt 0 in Betrieb ist, zum Zeitpunkt  $t$  noch nicht durchgebrannt ist, gegeben die Kenntnis, daß sie zum Zeitpunkt  $s$  noch intakt war. Die Wahrscheinlichkeit  $P(T > t - s)$  gibt die Wahrscheinlichkeit an, daß eine Glühbirne, die zum Zeitpunkt  $s$  in Betrieb genommen wurde, zum Zeitpunkt  $t = s + (t - s)$  noch intakt ist. Damit bedeutet die obige Gleichung, daß sich eine zum Zeitpunkt  $s$  noch intakte Birne in der Zukunft wie eine zum Zeitpunkt  $s$  neue Birne verhält. Das heißt, es findet keine Abnutzung oder Alterung statt. Im folgenden Abschnitt wird die Möglichkeit der Alterung genauer diskutiert.

Eng verwandt mit der Exponentialverteilung ist die Klasse der Gamma-Verteilungen. Dies ist eine sehr flexible Klasse von Wahrscheinlichkeitsverteilungen, auch in Hinblick auf Lebensdauern.

**Definition 4.3.1** Die  $\Gamma$ -Funktion ist gegeben durch

$$\Gamma(\alpha) := \int_0^\infty y^{\alpha-1} e^{-y} dy \quad \text{für } \alpha > 0$$

**Bemerkung:**

Für  $\alpha \in \mathbb{N}$  gilt  $\Gamma(\alpha + 1) = \alpha!$ . Dies läßt sich mit partieller Integration beweisen.

**Definition 4.3.2** Die Gamma-Verteilung zu den Parametern  $\alpha$  und  $\beta$ , genannt  $\mathcal{G}(\alpha, \beta)$  hat die Wahrscheinlichkeitsdichte

$$f_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} 1_{[0,\infty)}(x).$$

**Beispiele:**

- 1) Für  $\alpha = 1$  und  $\beta = \lambda$  ergibt sich die Exponentialverteilung mit Parameter  $\lambda$ .
- 2) Sind  $T_1, T_2, \dots, T_r$  unabhängig (zur Definition siehe den nächsten Abschnitt) und exponentialverteilt mit Parameter  $\lambda$ , so ist  $\sum_{i=1}^r T_i$  Gamma-verteilt mit den Parametern  $\alpha = r$  und  $\beta = \lambda$ . Der Beweis folgt weiter unten.

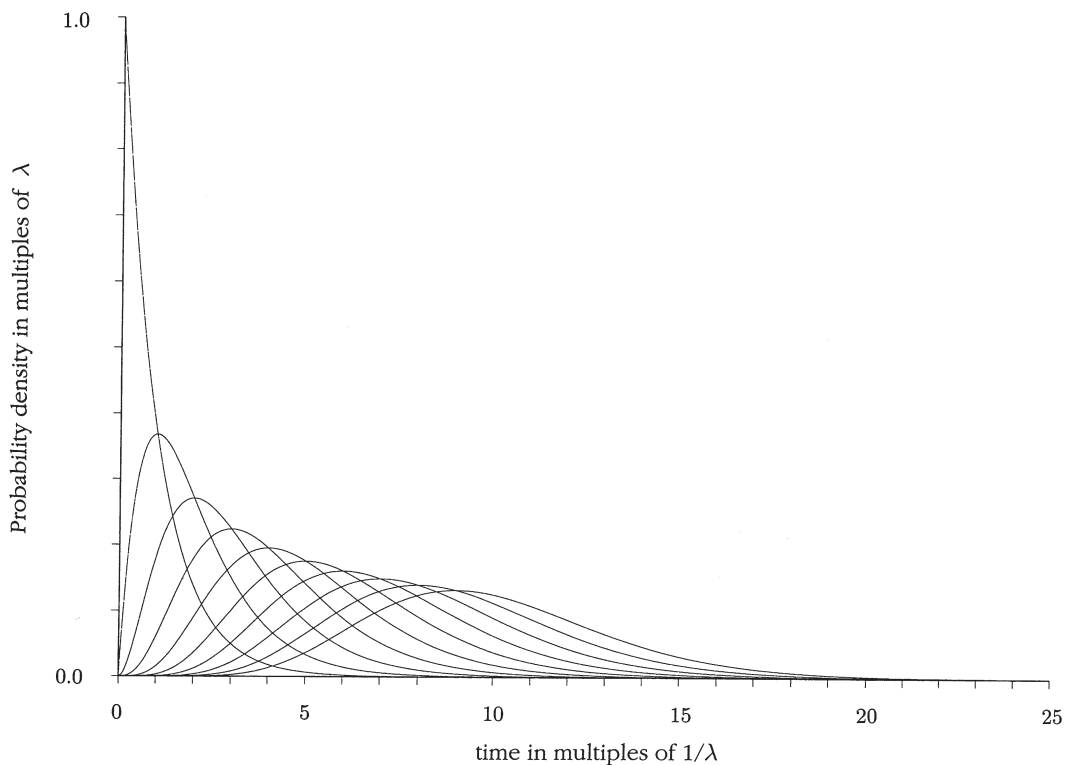


Abbildung 4.3: Die Dichte der Gamma-Verteilung für  $r = 1, \dots, 10$

## 4.4 Lebenszeiten

Wir wollen uns nun mit einem etwas anderen Zugang zur Beschreibung von zufälligen Ausfallzeiten befassen. Dabei beschränken wir uns auf den Fall, daß diese stetig verteilt sind.

**Definition 4.4.1** Sei  $T$  eine positive und stetige Zufallsvariable mit Dichte  $f$ . Dann heißt

$$G(t) = 1 - F(t) = \int_t^{\infty} f(u) du$$

die Zuverlässigkeitsfunktion von  $T$ . Die Funktion

$$\lambda(t) = \frac{f(t)}{G(t)}$$

heißt Ausfallrate.

Es besteht folgender Zusammenhang zwischen der Zuverlässigkeitsfunktion  $G$  und der Ausfallrate  $\lambda$ :

$$G(t) = \exp \left\{ - \int_0^t \lambda(u) du \right\}.$$

Die Größe  $\int_0^t \lambda(u) du$  heißt kumulierte Ausfallrate. Man sieht die obige Identität folgendermaßen ein:

$$\frac{d}{dt} \log(G(t)) = \frac{\frac{d}{dt} G(t)}{G(t)} = - \frac{f(t)}{G(t)} = -\lambda(t).$$

Bei der Exponentialverteilung ergibt sich

$$G(t) = e^{-\lambda t} \quad \text{und} \quad \lambda(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

Bisweilen ist die folgende Formel für  $E(T)$  von Nutzen:

$$E(T) = \int_0^{\infty} G(t) dt.$$

Man kann diese Gleichung mit Hilfe partieller Integration aus  $E(T) = \int_0^{\infty} t f(t) dt$  herleiten.

Eigenschaften von Ausfallraten:

i)  $\lambda(t) \geq 0$  für alle  $t \geq 0$ .

ii)  $\int_0^{\infty} \lambda(u) du = +\infty$ .

Durch Vorgabe einer Ausfallrate läßt sich eine Verteilung spezifizieren.

**Beispiel:** (Weibull-Verteilung) Sei  $\alpha > 0$  und  $\lambda > 0$ . Sei

$$\lambda(t) = \lambda \alpha t^{\alpha-1}.$$

Dann gilt

$$G(t) = \exp \left\{ - \int_0^t \lambda(u) du \right\} = \exp \{ -\lambda t^\alpha \}.$$

Für die zugehörige Dichte  $f$  ergibt sich

$$f(t) = -G'(t) = \lambda \alpha t^{\alpha-1} e^{-\lambda t^\alpha}.$$

Für  $0 < \alpha < 1$  fällt die Ausfallrate mit wachsender Zeit (Verjüngungseffekt) und für  $\alpha > 1$  wächst die Ausfallrate mit wachsender Zeit (Alterungseffekt).

Sei  $0 < s < t$ . Für die bedingte Wahrscheinlichkeit von  $s < T \leq t$  gegeben  $T > s$  erhält man

$$\begin{aligned} P(s < T \leq t \mid T > s) &= \frac{P(s < T \leq t)}{P(T > s)} = \frac{F(t) - F(s)}{1 - F(s)} \\ &= \frac{1}{1 - F(s)} \int_s^t f(u) du = \int_s^t \frac{f(u)}{1 - F(s)} du. \end{aligned}$$

Darüberhinaus gilt

$$P(T > t \mid T > s) = \frac{1 - F(t)}{1 - F(s)}.$$

## 4.5 Gemeinsame Verteilung mehrerer Zufallsvariablen

Wir betrachten nun  $n$  Zufallsvariablen  $X_1, \dots, X_n$  simultan und interessieren uns für die *gemeinsame Verteilung*. Dabei werden wir den Fall  $n = 2$  besonders behandeln und uns im wesentlichen auf stetige Verteilungen mit Dichten konzentrieren. Mit der gemeinsamen Verteilung meinen wir

$$P(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n),$$

wobei  $I_1, \dots, I_n$  beliebige Intervalle aus  $\mathbf{R}$  sind. Wir fassen gewissermaßen  $X_1, \dots, X_n$  als Zufallselement aus  $\mathbf{R}^n$  auf.

**Definition 4.5.1** Die Zufallsvariablen  $X_1, \dots, X_n$  heißen *unabhängig*, falls für alle Intervalle  $I_1, \dots, I_n$  aus  $\mathbf{R}$

$$P(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n) = \prod_{k=1}^n P(X_k \in I_k)$$

*gilt.*

**Definition 4.5.2** Die Zufallsvariablen  $X_1, \dots, X_n$  heißen *gemeinsam stetig verteilt* mit Dichte  $f$ , falls für alle Intervalle  $I_1, \dots, I_n$  aus  $\mathbf{R}$  gilt

$$P(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n) = \int_{I_n} \dots \int_{I_1} f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Die Funktion  $f$  heißt *gemeinsame Dichte* von  $X_1, \dots, X_n$ .

Ist  $g$  eine (stetige) Funktion von  $\mathbf{R}^n$  nach  $\mathbf{R}$  und sind  $X_1, \dots, X_n$  Zufallsvariablen, dann ist  $g(X_1, \dots, X_n)$  ebenfalls eine Zufallsvariable. Sind die Zufallsvariablen  $X_1, \dots, X_n$  gemeinsam stetig verteilt mit Dichte  $f$ , so erhält man den zugehörigen Erwartungswert  $E(g(X_1, \dots, X_n))$  durch

$$E(g(X_1, \dots, X_n)) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \dots dx_n,$$

falls

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |g(x_1, \dots, x_n)| f(x_1, \dots, x_n) dx_1 \dots dx_n < \infty.$$

Wir betrachten nun den Fall  $n = 2$  ausführlicher. Sei  $X = X_1$  und  $Y = X_2$ . Sind  $X$  und  $Y$  gemeinsam stetig verteilt mit Dichte  $f$ , so gilt

$$P(X \in I_1, Y \in I_2) = \int_{I_1} \left( \int_{I_2} f(x, y) dy \right) dx = \int_{I_2} \left( \int_{I_1} f(x, y) dx \right) dy.$$

Insbesondere gilt für beliebige  $a$  und  $b$

$$P(X \leq a, Y \leq b) = \int_{-\infty}^a \left( \int_{-\infty}^b f(x, y) dy \right) dx = \int_{-\infty}^b \left( \int_{-\infty}^a f(x, y) dx \right) dy.$$

Damit folgt

$$\begin{aligned} P(X \leq a) &= P(X \leq a, Y < \infty) = \int_{-\infty}^a \left( \int_{-\infty}^{\infty} f(x, y) dy \right) dx \\ &= \int_{-\infty}^a f_X(x) dx \end{aligned}$$

mit

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy.$$

Die Dichte von  $X$  ist also durch die sogenannte *Randdichte*  $f_X$  gegeben. Analog erhält man

$$P(Y \leq b) = \int_{-\infty}^b f_Y(y) dy \quad \text{mit} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Die folgende Definition ist die gleiche wie für diskrete Zufallsvariable.

**Definition 4.5.3**  $X$  und  $Y$  seien Zufallsvariablen mit

$$E(X^2) < \infty \quad \text{und} \quad E(Y^2) < \infty.$$

Dann heißt

$$\text{Kov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

die Kovarianz von  $X$  und  $Y$ . Die Größe

$$\rho(X, Y) = \frac{\text{Kov}(X, Y)}{\sigma(X)\sigma(Y)}$$

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

Für die Kovarianz gilt folgende Verschiebungsformel

$$\text{Kov}(X, Y) = E(XY) - E(X)E(Y).$$

Es gilt stets  $-1 \leq \rho(X, Y) \leq 1$ . Ist  $X = Y$ , so gilt

$$\text{Kov}(X, Y) = \text{Kov}(X, X) = \text{Var}(X).$$

Damit ist  $\rho(X, X) = 1$ .

Als Beispiel wollen wir die 2-dimensionale Normalverteilung betrachten. Bei ihr ist die Dichte der gemeinsamen Verteilung der Zufallsvariablen  $X$  und  $Y$  gegeben durch

$$\begin{aligned} (+) \quad f(x, y) &= \frac{1}{2\pi(\det \Sigma)^{1/2}} \\ &\cdot \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} - \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right) \end{aligned}$$

die Dichten der Randverteilungen sind

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-(x-\mu_X)^2/2\sigma_X^2} \\ f_Y(y) &= \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-(y-\mu_Y)^2/2\sigma_Y^2} \end{aligned}$$

Außerdem ist  $\rho = \text{Kor}(X, Y)$  und  $\sigma_{XY} := \text{Kov}(X, Y) = \rho\sigma_X\sigma_Y$ . Die Matrix  $\Sigma$  ist gegeben durch

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}.$$

Es gilt natürlich

$$\sigma_{XY} = \int_{\mathbb{R}} \int_{\mathbb{R}} (y - \mu_X)(y - \mu_Y) f(x, y) dx dy$$

Man erkennt, daß gilt

$$f(x, y) = f_X(x)f_Y(y),$$

d.h. daß  $X$  und  $Y$  unabhängig sind, falls  $\rho = 0$  bzw.  $\sigma_{XY} = 0$  sind.

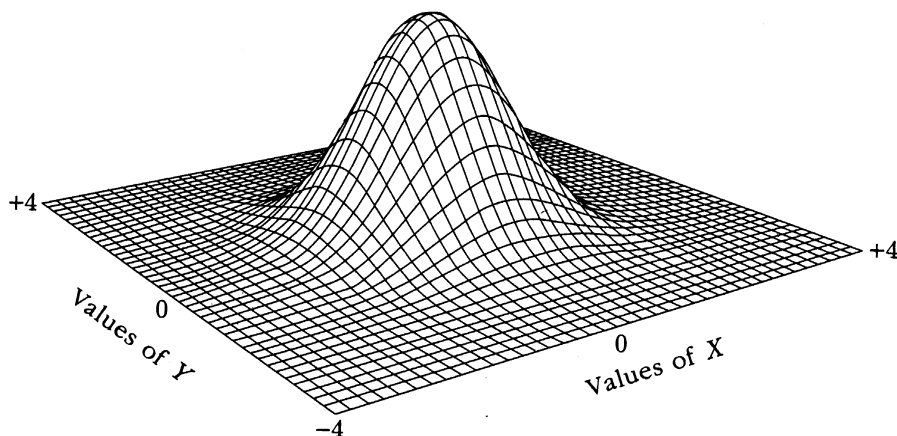


Abbildung 4.4: 2-dimensionale Normalverteilung

Das folgende Beispiel gilt allgemein.

**Beispiel:** (Lineares Filtern)

Seien  $X$  ein gesendetes und  $Y$  das zugehörige empfangene Signal. Gesucht wird eine lineare Rekonstruktion  $aY + b$  mit minimaler mittlerer quadratischer Abweichung

$$E([X - (aY + b)]^2).$$

Das Minimum wird für

$$a = \rho(X, Y) \frac{\sigma(X)}{\sigma(Y)} \quad \text{und} \quad b = E(X) - aE(Y)$$

erreicht. Es gilt nämlich

$$\begin{aligned} E([X - (aY + b)]^2) &= E(X^2) - 2E[(aY + b)X] + E[(aY + b)^2] \\ &= E(X^2) - 2aE(XY) - 2bE(X) + a^2E(Y^2) \\ &\quad + 2abE(Y) + b^2. \end{aligned}$$

Ableiten nach  $a$  und  $b$  und 0 setzen ergibt die Gleichungen

$$\begin{aligned} aE(Y^2) - E(XY) + bE(Y) &= 0 \\ b - E(X) + aE(Y) &= 0. \end{aligned}$$

Hieraus folgt zunächst

$$b = E(X) - aE(Y)$$

und

$$a \underbrace{[E(Y^2) - (E(Y))^2]}_{=\text{Var}(Y)} - \underbrace{[E(XY) - E(X)E(Y)]}_{=\text{Kov}(X,Y)} = 0.$$

Damit ist

$$a = \frac{\text{Kov}(X, Y)}{\text{Var}(Y)} = \frac{\sigma(X)}{\sigma(Y)} \rho(X, Y).$$

Wir kommen nun zurück auf die Unabhängigkeit von Zufallsvariablen und wollen uns mit der Verteilung von deren Summen beschäftigen. Zunächst eine wichtige Tatsache. Sind die Zufallsvariablen  $X_1, X_2, \dots, X_n$  unabhängig und jeweils stetig verteilt mit Dichten  $f_1, f_2, \dots, f_n$ , so gilt für die gemeinsame Dichte

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f_i(x_i).$$

Denn:

$$\begin{aligned} P(X_1 \leq \alpha_1, \dots, X_n \leq \alpha_n) &= \prod_{i=1}^n P(X_i \leq \alpha_i) \\ &= \prod_{i=1}^n \int_{-\infty}^{\alpha_i} f_i(x_i) dx_i \\ &= \int_{-\infty}^{\alpha_1} \dots \int_{-\infty}^{\alpha_n} \prod_{i=1}^n f_i(x_i) dx_1 \dots dx_n \end{aligned}$$

Sind nun  $X$  und  $Y$  unabhängig mit Dichten  $f$  und  $g$  verteilt. Dann hat  $Z = X + Y$  eine Verteilung mit Dichte  $(f * g)(t) = \int_{-\infty}^{+\infty} f(t-s)g(s) ds$ . (Man nennt  $f * g$  das Faltungsprodukt von  $f$  und  $g$ .) Dies sieht man so:

$$\begin{aligned} P(Z \leq \alpha) &= P(X + Y \leq \alpha) \\ &= \int_{\{(x,y)|x+y \leq \alpha\}} f(x)g(y) dx dy \\ &= \int_{\{(s,t)|t \leq \alpha\}} f(t-s)g(s) ds dt \quad \text{mit } t = x + y \text{ und } s = y \\ &= \int_{-\infty}^{\alpha} \left( \int_{-\infty}^{+\infty} f(t-s)g(s) ds \right) dt \end{aligned}$$

**Beispiel:**

Seien  $X$  und  $Y$  normalverteilt,  $X$  nach  $N(0, v_1)$  und  $Y$  nach  $N(0, v_2)$ . Sind  $X$  und  $Y$  unabhängig, so ist  $X + Y$  normalverteilt nach  $N(0, v_1 + v_2)$ . Die Dichte der Verteilung  $X + Y$  ergibt sich nach etwas Rechnung aus der Faltungsformel als

$$\begin{aligned} f(z) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi v_2}} e^{-(z-x)^2/2v_2} \frac{1}{\sqrt{2\pi v_1}} e^{-x^2/2v_1} dx \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{(2\pi)^2 v_1 v_2}} e^{-z^2/2(v_1+v_2)} e^{-\frac{1}{2}\left(x - \frac{v_1}{v_1+v_2}z\right)^2 / \frac{v_1 v_2}{v_1+v_2}} dx \\ &= \frac{1}{\sqrt{2\pi(v_1 + v_2)}} e^{-\frac{z^2}{2(v_1+v_2)}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \sqrt{\frac{v_1 + v_2}{v_1 v_2}} e^{-\frac{1}{2}\left(x - \frac{v_1}{v_1+v_2}z\right)^2 / \frac{v_1 v_2}{v_1+v_2}} dx \end{aligned}$$

Das Integral auf der rechten Seite ist gleich 1, so daß weiter gilt

$$= \frac{1}{\sqrt{2\pi(v_1 + v_2)}} e^{-z^2/2(v_1+v_2)}.$$

Dies ist aber die Behauptung.

**Beispiel:**

$$\begin{aligned} f(x) &= g(x) = \lambda e^{-\lambda x} 1_{[0, \infty)}(x) \\ (f * g)(z) &= \int_0^{\infty} \lambda e^{-\lambda(z-x)} 1_{\{x < z\}} \lambda e^{-x} dx \\ &= \lambda^2 e^{-\lambda z} \int_0^z dx \\ &= \lambda^2 e^{-\lambda z} \cdot z, \quad \text{falls } z > 0 \end{aligned}$$

Allgemein gilt:

$$(f^*)^n(z) = \lambda^n \frac{z^{n-1}}{(n-1)!} e^{-\lambda z}, \quad \text{falls } z > 0 \text{ gilt.}$$

Dies beweist man mit Induktion:  $(f^*)^n(z)$  ist die Dichte einer Gamma-Verteilung mit Parameter  $n$  und  $\lambda$ . Der Fall  $n = 2$  wurde schon gezeigt. Sei die Aussage nun richtig für beliebiges  $n$ , wir zeigen, daß sie für  $n + 1$  gilt.

$$\begin{aligned}
 (f^*)^{n+1}(z) &= f * (f^*)^n(z) \\
 &= \int_{-\infty}^{\infty} f(z-x)(f^*)^n(x)dx \\
 &= \int_{-\infty}^{\infty} \lambda e^{-\lambda(z-x)} 1_{[0,\infty]}(z-x) \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} 1_{[0,\infty]}(x) dx \\
 &= \lambda^{n+1} e^{-\lambda z} \int_{-\infty}^{\infty} \frac{x^{n-1}}{(n-1)!} 1_{[0,z)}(x) dx \\
 &= \lambda^{n+1} e^{-\lambda z} \int_0^z \frac{x^{n-1}}{(n-1)!} dx \quad \text{für } z \geq 0 \text{ und } = 0 \text{ und } z < 0 \\
 &= \lambda^{n+1} \frac{z^n}{n!} e^{-\lambda z} 1_{[0,\infty]}(z).
 \end{aligned}$$

Eine Anwendung dieser Formel ist die Herleitung des Poisson-Prozesses.

## Der Poisson-Prozess

Wir interessieren uns nun für die Anzahl der Atome einer radioaktiven Substanz, die in einem gewissen Zeitintervall  $[0, t]$  zerfallen. Wir nennen diese Größe  $N_t$ . Sei  $T_1$  die Zeitspanne bis das erste Atom zerfallen ist,  $T_2$  die Zeitspanne, die nach  $T_1$  vergeht bis das zweite Atom zerfallen ist,  $T_3$  die Zeitspanne bis das dritte Atom nach  $T_1 + T_2$  zerfallen ist, u.s.w. Wir nehmen an, daß alle  $T_1, T_2, T_3$  u.s.w. unabhängig sind. Sei  $S_n = \sum_{i=1}^n T_i$ . Dann ist  $N_t = \max\{k \geq 1 \mid S_k \leq t\}$ . Was ist die Verteilung von  $N_t$  ?

Unter Verwendung des vorangegangenen Resultats erhalten wir:

$$\begin{aligned}
 P(N_t = k) &= P(S_k \leq t, S_{k+1} > t) \\
 &= P(S_{k+1} > t) - P(S_k > t) \\
 &= \int_t^{\infty} \lambda^{k+1} \frac{s^k}{k!} e^{-\lambda s} ds - \int_t^{\infty} \lambda^k \frac{s^{k-1}}{(k-1)!} e^{-\lambda s} ds \\
 &= -\lambda^k \frac{s^k}{k!} e^{-\lambda s} \Big|_t^{\infty} \\
 &= \frac{(\lambda t)^k}{k!} e^{-\lambda t}
 \end{aligned}$$

Folglich gilt, daß  $N_t$  Poisson-verteilt ist mit Parameter  $\lambda \cdot t$ . Man nennt  $N_t$ ;  $t \geq 0$  Poisson-Prozess zur Intensität  $\lambda$ . Es gilt  $E(N_t) = \lambda \cdot t$  und  $\text{Var}(N_t) = \lambda \cdot t$ .

## 4.6 Bedingte Verteilungen und Bedingte Erwartungen

Wir wenden uns nun der *bedingten* Verteilung von  $X$  gegeben  $Y = y$  zu. Diese läßt sich über ihre Dichte festlegen. Zur Motivation betrachten wir zunächst kurz den Fall, daß  $X$

und  $Y$  beide diskret sind. Wir nehmen an, daß  $X$  und  $Y$  jeweils Werte in  $\{1, 2, \dots, n\}$  annehmen. Dann ist die gemeinsame Verteilung von  $X$  und  $Y$  durch die Wahrscheinlichkeiten

$$p_{ij} = P(X = i, Y = j)$$

festgelegt. Die Randverteilung von  $X$  ergibt sich zu

$$P(X = i) = \sum_{j=1}^n P(X = i, Y = j) = \sum_{j=1}^n p_{ij}.$$

Die Randverteilung von  $Y$  ergibt sich zu

$$P(Y = j) = \sum_{i=1}^n P(X = i, Y = j) = \sum_{i=1}^n p_{ij}.$$

Die bedingte Wahrscheinlichkeit von  $X = i$  gegeben  $Y = j$  ergibt sich damit zu

$$P(X = i|Y = j) = \frac{P(X = i, Y = j)}{P(Y = j)} = \frac{p_{ij}}{\sum_{i=1}^n p_{ij}}.$$

Dies motiviert die folgenden Überlegungen für den Fall, daß  $X$  und  $Y$  gemeinsam stetig verteilt mit Dichte  $f$  sind. Sei

$$f_X(x|Y = y) = \frac{f(x, y)}{f_Y(y)} = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dx}.$$

Dabei lassen wir nur  $y$  Werte mit  $f_Y(y) > 0$  zu. Offensichtlich ist  $f_X(x|Y = y) \geq 0$  und

$$\int_{-\infty}^{\infty} f_X(x|Y = y) dx = \int_{-\infty}^{\infty} \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dx} dx = \frac{\int_{-\infty}^{\infty} f(x, y) dx}{\int_{-\infty}^{\infty} f(x, y) dx} = 1.$$

Damit ist  $f_X(\cdot|Y = y)$  eine Dichte; die sogenannte *bedingte Dichte* von  $X$  gegeben  $Y = y$ . Analog erhält man die bedingte Dichte  $f_Y(\cdot|X = x)$  von  $Y$  gegeben  $X = x$  zu

$$f_Y(y|X = x) = \frac{f(x, y)}{f_X(x)} = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dy}.$$

Es gilt also stets: bedingte Dichte = gemeinsame Dichte geteilt durch Randdichte. Damit läßt sich die gemeinsame Dichte auch folgendermaßen darstellen

$$f(x, y) = f_X(x) f_Y(y|X = x) = f_Y(y) f_X(x|Y = y).$$

Der Erwartungswert der bedingten Verteilung von  $X$  gegeben  $Y = y$  wird mit  $E(X|Y = y)$  bezeichnet und ist gleich

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f_X(x|Y = y) dx = \frac{\int_{-\infty}^{\infty} x f(x, y) dx}{\int_{-\infty}^{\infty} f(x, y) dx}.$$

Analog gilt für den bedingten Erwartungswert  $E(Y|X = x)$  von  $Y$  gegeben  $X = x$

$$E(Y|X = x) = \frac{\int_{-\infty}^{\infty} y f(x, y) dy}{\int_{-\infty}^{\infty} f(x, y) dy}.$$

Allgemeiner gilt

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

und

$$E(g(Y)|X = x) = \int_{-\infty}^{\infty} g(y) f_Y(y|X = x) dy.$$

**Beispiel:** Sind  $X$  und  $Y$  unabhängig, so gilt

$$f(x, y) = f_X(x) f_Y(y).$$

Damit erhält man

$$f_X(x|Y = y) = \frac{f(x, y)}{f_Y(y)} = f_X(x).$$

Die bedingte Verteilung stimmt also bei Unabhängigkeit mit der (unbedingten) Randverteilung überein. Deshalb ist für alle  $y$

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

Der Erwartungswert  $E(X|Y = y)$  von  $X$  gegeben  $Y = y$  ist eine Funktion von  $y$ . Setzt man für  $y$  die Zufallsgröße  $Y$  ein (oder anders gesagt: wendet man die Funktion auf  $Y$  an), so erhält man wieder eine Zufallsgröße, die mit  $E(X|Y)$  bezeichnet wird. Es gilt die *Glättungsregel*

$$E(X) = E(E(X|Y)) = \int_{-\infty}^{\infty} E(X|Y = y) f_Y(y) dy.$$

Darüberhinaus gilt die sogenannte *Einsetzungsregel*

$$E(Xh(Y)|Y = y) = h(y)E(X|Y = y)$$

beziehungsweise

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

**Beispiel:** Sind  $X$  und  $Y$  unabhängig, so erhält man

$$E(XY|Y) = YE(X|Y) = YE(X)$$

und weiter

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

**Beispiel (Filtern):** Seien  $X$  ein gesendetes und  $Y$  das zugehörige empfangene Signal. Gesucht wird eine Rekonstruktion  $g(Y)$  mit minimaler mittlerer quadratischer Abweichung

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

Das Minimum wird mit

$$g(y) = E(X|Y = y)$$

erreicht. Es gilt nämlich

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

Bei festem  $Y = y$  wird  $-2E(X|Y = y)g + g^2$  durch  $g = E(X|Y = y)$  minimiert. Dieses Resultat ist Grundlage der *Filtertheorie*.

Nun kehren wir zurück zu normalverteilten Zufallsvariablen. Seien  $X$  und  $Z$  unabhängige Zufallsgrößen, die beide  $N(0, 1)$ -verteilt sind. Dann ist

$$E(X) = 0 = E(Z)$$

und

$$1 = \text{Var}(X) = E(X^2) - (E(X))^2 = E(X^2)$$

sowie

$$1 = \text{Var}(Z) = E(Z^2) - (E(Z))^2 = E(Z^2).$$

Aus der Unabhängigkeit folgt  $E(XZ) = E(X)E(Z) = 0$ . Sei  $-1 < \rho < 1$  und

$$Y = \rho X + \sqrt{1 - \rho^2}Z.$$

Wir untersuchen nun die gemeinsame Verteilung von  $X$  und  $Y$ . Offensichtlich gilt

$$E(Y) = \rho E(X) + \sqrt{1 - \rho^2}E(Z) = 0.$$

Weiterhin ist

$$\begin{aligned} E(Y^2) &= E\left(\rho^2 X^2 + 2\rho\sqrt{1 - \rho^2}XZ + (1 - \rho^2)Z^2\right) \\ &= \rho^2 E(X^2) + 2\rho\sqrt{1 - \rho^2}E(XZ) + (1 - \rho^2)E(Z^2) \\ &= \rho^2 + (1 - \rho^2) \\ &= 1. \end{aligned}$$

Also gilt  $\text{Var}(Y) = 1$ . Für die Kovarianz von  $X$  und  $Y$  erhalten wir

$$\begin{aligned} \text{Kov}(X, Y) &= E(XY) - E(X)E(Y) = E[X(\rho X + \sqrt{1 - \rho^2}Z)] \\ &= \rho E(X^2) + \sqrt{1 - \rho^2}E(XZ) = \rho. \end{aligned}$$

Die bedingte Verteilung von  $Y$  gegeben  $X = x$  ist gleich der Verteilung von  $\rho x + \sqrt{1 - \rho^2}Z$  und somit eine Normalverteilung mit Mittelwert  $\rho x$  und Varianz  $1 - \rho^2$ . Damit gilt für die bedingte Dichte  $f_Y(\cdot | X = x)$

$$(*) \quad f_Y(y|X = x) = \frac{1}{\sqrt{2\pi(1 - \rho^2)}} \exp\left(-\frac{(y - \rho x)^2}{2(1 - \rho^2)}\right).$$

Der bedingte Erwartungswert  $E(Y|X = x)$  von  $Y$  gegeben  $X = x$  ist deshalb gleich  $\rho x$ . Damit ist  $E(Y|X) = \rho X$ . Für die gemeinsame Dichte  $f$  von  $X$  und  $Y$  erhalten wir gemäß der Formel für die bedingte Dichte

$$\begin{aligned} f(x, y) &= f_X(x)f_Y(y|X = x) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{(x^2 - 2\rho xy + y^2)}{2(1-\rho^2)}\right). \end{aligned}$$

Diese Formel stimmt mit der der Formel (+) auf Seite 53 überein, wenn man  $\mu_X = \mu_Y = 0$  und  $\sigma_X = \sigma_Y = 1$  setzt. Dieser Ausdruck ist symmetrisch in  $x$  und  $y$ . Wir können also die Rollen von  $X$  und  $Y$  vertauschen. Damit gilt auch:

- a) die bedingte Verteilung von  $X$  gegeben  $Y = y$  ist  $N(\rho y, 1 - \rho^2)$ .
- b)  $E(X|Y = y) = \rho y$ .

Setzt man nun  $X = \frac{U-\mu_U}{\sigma_U}$  und  $Y = \frac{V-\mu_V}{\sigma_V}$ , so erhält man für  $U$  und  $V$  die Dichte

$$\frac{1}{2\pi\sqrt{\sigma_U^2\sigma_V^2(1-\rho^2)}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[ \frac{(u-\mu_U)^2}{\sigma_U^2} - 2\rho\frac{(u-\mu_U)(v-\mu_V)}{\sigma_U\sigma_V} + \frac{(v-\mu_V)^2}{\sigma_V^2} \right]\right).$$

Wegen  $V = \sigma_V Y + \mu_V$  erhält man mit (\*) die bedingte Dichte

$$f_V(v|U = u) = \frac{1}{\sqrt{2\pi\sigma_V^2(1-\rho^2)}} \exp\left(-\frac{1}{2} \frac{\left(v - \left(\rho\frac{u-\mu_U}{\sigma_U}\sigma_V + \mu_V\right)\right)^2}{\sigma_V^2(1-\rho^2)}\right).$$

D.h., die bedingte Verteilung von  $V$  gegeben  $U = u$  ist eine Normalverteilung mit Mittelwert  $\rho\frac{u-\mu_U}{\sigma_U}\sigma_V + \mu_V$  und Varianz  $\sigma_V^2(1-\rho^2)$ . Insbesondere ist

$$E(V|U = u) = \rho\frac{u-\mu_U}{\sigma_U}\sigma_V + \mu_V.$$

Aus der Gleichung

$$f_U(u) = \frac{f(u, v)}{f_V(v|U = u)}$$

ergibt sich, daß  $U$   $N(\mu_U, \sigma_U^2)$ -verteilt ist. Analog folgt, daß  $V$   $N(\mu_V, \sigma_V^2)$ -verteilt ist. Für unkorrelierte  $U$  und  $V$  gilt speziell  $f_V(v|U = u) = f_V(v)$ . Dies bedeutet, daß unkorrelierte normalverteilte Zufallsgrößen unabhängig sind.

**Beispiel:** Der Statistiker Karl Pearson hat die Körpergrößen bei 1078 Vater-Sohn Paaren bestimmt. Er stellte fest, daß die Väter im Mittel 5 Fuß und 9 Inch und die Söhne im Mittel 5 Fuß und 10 Inch groß sind. Die Standardabweichung beträgt jeweils 2 Inch und die Korrelation ist 0.5. Man kann in guter Näherung annehmen, daß die Körpergrößen in einem Vater-Sohn Paar gemeinsam normalverteilt sind. Wie groß ist dann im Mittel der

Sohn eines Vaters mit 6 Fuß und 2 Inch Körpergröße?

Sei  $U$  die Körpergröße des Vaters und  $V$  die Körpergröße des Sohnes. Sei

$$X = \frac{U - \mu_U}{\sigma_U} \quad \text{und} \quad Y = \frac{V - \mu_V}{\sigma_V}.$$

Für  $U = 6$  Fuß und 2 Inch erhalten wir  $X = 2.5$  (12 Inch sind 1 Fuß). Da  $X$  und  $Y$  gemeinsam standard normalverteilt sind, ist  $E(Y|X = x) = \rho x$ . Damit ist  $E(Y|X = 2.5) = 1.25$ . Dies ergibt

$$E(V|U = u) = \rho \frac{u - \mu_U}{\sigma_U} \sigma_V + \mu_V = 2.5 \text{ Inch} + 5 \text{ Fuß } 10 \text{ Inch} = 6 \text{ Fuß } 0.5 \text{ Inch}.$$

Der Sohn eines außergewöhnlich großen Vaters ist also im Mittel kleiner als sein Vater. Dieses Phänomen wurde von Galton als *Regression zum Mittelwert* bezeichnet.

## 4.7 Empirische Verteilungsfunktion und Quantile

**Definition 4.7.1** Sei  $X$  eine Zufallsvariable und  $0 < p < 1$ . Dann heißt  $q_p$  Quantil der Ordnung  $p$  oder  $p$ -Quantil, falls

$$P(X < q_p) \leq p \leq P(X \leq q_p)$$

gilt.  $q_{1/2}$  heißt Median,  $q_{1/4}$  heißt erstes Quartil,  $q_{3/4}$  heißt drittes Quartil.

Man beachte, daß  $p$ -Quantile im allgemeinen nicht eindeutig bestimmt sind.

Ist  $F$  die Verteilungsfunktion von  $X$ , so gilt

$$P(X < q) = \lim_{h \downarrow 0} P(X \leq q - h) = \lim_{h \downarrow 0} F(q - h) = F(q-).$$

Dabei bezeichnet  $F(q-)$  den linksseitigen Grenzwert von  $F$  an der Stelle  $q$ . Dieser existiert stets, da  $F$  wachsend ist. Die definierende Gleichung für ein  $p$ -Quantil ist äquivalent zu

$$F(q-) \leq p \leq F(q).$$

Ist  $X$  stetig verteilt mit Dichte  $f$ , so ist  $F$  stetig und damit gilt für ein  $p$ -Quantil

$$F(q) = p.$$

Ist die Dichte  $f$  strikt positiv, so ist die Verteilungsfunktion  $F$  strikt monoton wachsend und die Gleichung  $F(q) = p$  besitzt eine eindeutige Lösung.

**Beispiel:**

Sei  $X$  exponentialverteilt mit Parameter  $\lambda$ . Dann gilt  $E(X) = 1/\lambda$  und

$$P(X \leq 1/\lambda) = \lambda \int_0^{1/\lambda} e^{-\lambda x} dx = 1 - e^{-1} \approx 0.632 > 0.5.$$

Der Median der Exponentialverteilung ergibt sich durch Lösen der Gleichung

$$1 - e^{-\lambda q} = 0.5.$$

Damit erhalten wir, daß der Median gleich

$$\log(2) \frac{1}{\lambda} \approx 0.69 \frac{1}{\lambda} < E(X)$$

ist. Die linke Seite wird auch oft als Halbwertszeit bezeichnet, wenn man an den radioaktiven Zerfall von Materie denkt.

**Definition 4.7.2** Seien  $X_1, X_2, \dots, X_n, \dots$  unabhängig identisch verteilt mit Verteilungsfunktion  $F$ . Dann heißt  $\hat{F}_n(z) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq z\}}$  empirische Verteilungsfunktion nach  $n$  Beobachtungen.

Es gilt:  $\hat{F}_n(z) \rightarrow F(z)$  für alle  $z$ . Denn:

$$\begin{aligned} E\hat{F}_n(z) &= \frac{1}{n} \sum_{i=1}^n E1_{\{X_i \leq z\}} \\ &= \frac{1}{n} \sum_{i=1}^n P(X_i \leq z) \\ &= P(X_1 \leq z) \\ &= F(z). \end{aligned}$$

Aber  $\hat{F}_n(z)$  ist arithmetisches Mittel der unabhängig identisch verteilten Zufallsvariablen  $1_{\{X_i \leq z\}}$ . Aus dem Gesetz der großen Zahlen folgt die Konvergenz.

Sei nun  $\hat{q}_{p,n}$  das  $p$ -Quantil der empirischen Verteilungsfunktion  $\hat{F}_n$ . Dann folgt aus der Konvergenz der empirischen Verteilungsfunktion  $\hat{q}_{p,n} \rightarrow q_p$  für alle  $0 < p < 1$ .

Schließlich lassen sich arithmetisches Mittel und empirische Varianz auch als Funktionale der empirischen Verteilungsfunktion auffassen:

$$\begin{aligned} \text{a)} \quad \bar{X}_n &= \frac{1}{n} \sum_{i=1}^n X_i = \int x \hat{F}_n(dx) \\ \text{b)} \quad \hat{\sigma}_n^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \int (x - \bar{X}_n)^2 \hat{F}_n(dx). \end{aligned}$$

Die Konvergenz von  $\bar{X}_n \rightarrow E(X_1)$  und  $\hat{\sigma}_n^2 \rightarrow \sigma^2$  folgt unter passenden Zusatzvoraussetzungen.

Wie lassen sich nun Datensätze mit unbekannter zugrundeliegender Verteilung darstellen?

Gängig sind die folgenden drei Möglichkeiten:

1. als empirische Verteilungsfunktion,
2. als Histogramm, d.h. als Balkendiagramm,
3. als Box-Plot.

Der Box-Plot besteht aus einer *Box* um den Median mit 1. und 3. Quartil als Begrenzung nach unten oder oben. Sei  $\widehat{IQR} = \widehat{q}_{3/4} - \widehat{q}_{1/4}$  das interquartile range. Weiterhin gehen Linien von den oberen und unteren Enden der Box zu den adjacent Werten, die wie folgt definiert sind. Der größte Wert, der kleiner ist als  $\widehat{q}_{3/4} + 1.5 \widehat{IQR}$ , ist der obere adjacent Wert. Der kleinste Wert, der größer ist als  $\widehat{q}_{1/4} - 1.5 \widehat{IQR}$ , ergibt den unteren adjacent Wert. Liegt ein Beobachtungswert außerhalb von den adjacent Werten, gilt er als Ausreißer und wird einzeln dargestellt.

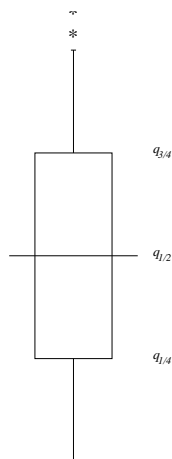


Abbildung 4.5: Box-Plot

Box-Plots sind sehr gut geeignet für einfache und mehrfache Vergleiche mehrerer Datensätze.

**Beispiel:**

Studiendauern in Semester im Diplom-Studiengang Mathematik vor 10 Jahren.

Fachsemester	6	10	11	12	13	14	15	16	18	20	22	29
	1	1	1	3	4	4	2	3	1	3	1	1

Der Median ist  $\text{med}_{25} = 14$  und das arithmetische Mittel ist  $\overline{X}_{25} = 15,12$ . Die Standardabweichung beträgt  $\hat{\sigma}_{25} = 4,57$ . Histogramm und Boxplot für obigen Datensatz zeigen die beiden

folgenden Graphiken.

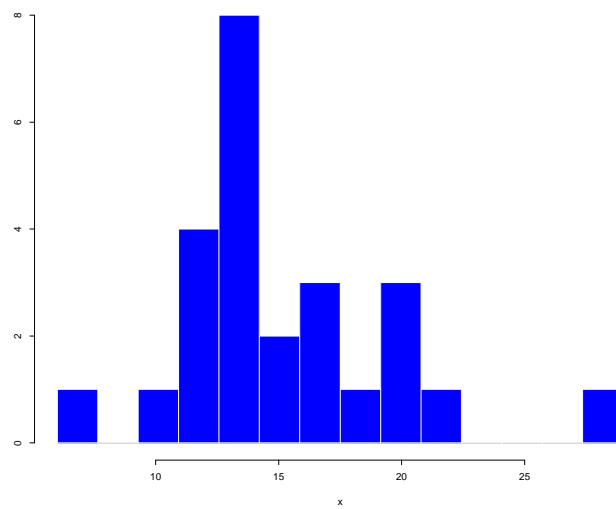


Abbildung 4.6: Histogramm

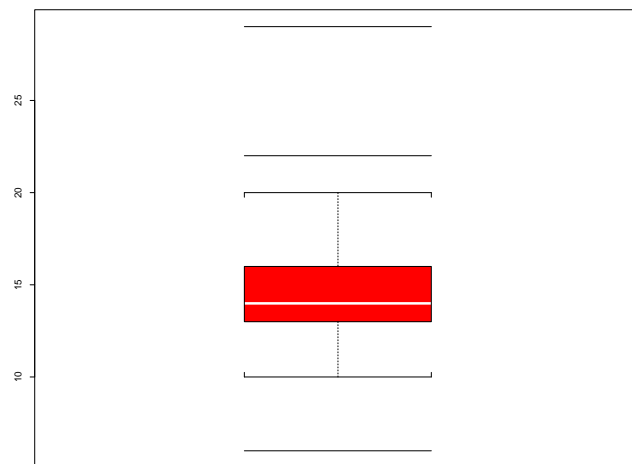


Abbildung 4.7: Box-Plot

# Kapitel 5

## Statistik

### 5.1 Maximum-Likelihood Methode

Wir betrachten nun folgendes allgemeine parametrische Schätzproblem. Die Verteilung  $F$  der Beobachtungen  $X_1, \dots, X_n$  sei bis auf einen Parameter  $\theta \in \Theta \subset \mathbf{R}^p$  bekannt. Sei  $X$  eine Zufallsvariable, die dieser Verteilung  $F$  folgt. Dann gilt

$$P(X \in I) = P_\theta(X \in I)$$

für ein  $\theta \in \Theta$  und beliebige Teilintervalle  $I$  von  $\mathbf{R}$ . Ziel ist die Bestimmung von  $\theta$ . Zunächst beschränken wir uns auf den Fall  $p = 1$ .

**Beispiel 1:**

$X$  ist Poisson-verteilt zum Parameter  $\theta$ . Also

$$P(X = x) = \frac{\theta^x}{x!} e^{-\theta}$$

für  $k = 0, 1, 2, \dots$

**Beispiel 2:**

$X$  ist exponentialverteilt zum Parameter  $\theta$ . Damit ist  $X$  stetig verteilt mit Dichte

$$f_\theta(x) = \begin{cases} 0 & \text{für } x \leq 0 \\ \theta e^{-\theta x} & \text{für } x > 0 \end{cases}.$$

Ein zentraler Begriff der parametrischen Statistik ist die sogenannte *Likelihood-Funktion*  $L(x_1, \dots, x_n; \theta)$ . Dies ist die Dichte bzw. die Wahrscheinlichkeitsfunktion der gemeinsamen Verteilung der Beobachtungen als Funktion von  $\theta$ . Wir betrachten zunächst den Fall, daß die Verteilung von  $X$  diskret ist mit

$$P_\theta(X = x) = p_\theta(x)$$

für alle  $x$  im Wertebereich von  $X$ . Dabei soll der Wertebereich nicht von  $\theta$  abhängen. Dann ergibt sich die Likelihood zu

$$L(x_1, \dots, x_n; \theta) = \prod_{k=1}^n p_\theta(x_k).$$

Im ersten Beispiel gilt also

$$L(x_1, \dots, x_n; \theta) = \frac{\theta^{\sum_{k=1}^n x_k}}{\prod_{k=1}^n (x_k!)} e^{-n\theta}.$$

Ist nun die Verteilung von  $X$  stetig mit Dichte  $f_\theta$ , so ist die Likelihood gleich

$$L(x_1, \dots, x_n; \theta) = \prod_{k=1}^n f_\theta(x_k).$$

Im zweiten Beispiel gilt also

$$L(x_1, \dots, x_n; \theta) = \theta^n e^{-\theta \sum_{k=1}^n x_k}.$$

Bei Beobachtungen  $x_1, \dots, x_n$  ist die Likelihood eine Funktion des unbekanntes Parameters  $\theta$ . Die Grundidee der Maximum-Likelihood Methode besteht darin, die Likelihood über  $\Theta$  zu maximieren und als Schätzwert

$$\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$$

eine Maximalstelle zu wählen. Also

$$L(x_1, \dots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, \dots, x_n; \theta).$$

$\hat{\theta}(X_1, \dots, X_n)$  heißt dann Maximum-Likelihood Schätzer. Die Maximierung von  $L(x_1, \dots, x_n; \theta)$  ist offensichtlich äquivalent zur Maximierung von

$$\log L(x_1, \dots, x_n; \theta).$$

In vielen Anwendungen ist  $\log L$  eine konkave Funktion in  $\theta$ . Dies gewährleistet dann Existenz und Eindeutigkeit des Maximum-Likelihood Schätzers.

Im ersten Beispiel erhalten wir

$$\log L(x_1, \dots, x_n; \theta) = -n\theta + \log(\theta) \sum_{k=1}^n x_k - \sum_{k=1}^n \log(x_k!).$$

Damit ist

$$\frac{\partial}{\partial \theta} \log L = -n + \frac{1}{\theta} \sum_{k=1}^n x_k$$

und

$$\frac{\partial^2}{\partial \theta^2} \log L = -\frac{1}{\theta^2} \sum_{k=1}^n x_k < 0.$$

Der Maximum-Likelihood Schätzer  $\hat{\theta}$  ergibt sich zu

$$\hat{\theta}(X_1, \dots, X_n) = \bar{X}_n = (X_1 + \dots + X_n)/n.$$

Im zweiten Beispiel erhalten wir

$$\log L(x_1, \dots, x_n; \theta) = n \log \theta - \theta \sum_{k=1}^n x_k.$$

Damit ist

$$\frac{\partial}{\partial \theta} \log L = \frac{n}{\theta} - \sum_{k=1}^n x_k$$

und

$$\frac{\partial^2}{\partial^2 \theta} \log L = -\frac{n}{\theta^2} < 0.$$

Der Maximum-Likelihood Schätzer  $\hat{\theta}$  ergibt sich zu

$$\hat{\theta}(X_1, \dots, X_n) = \frac{n}{\sum_{k=1}^n X_k}.$$

Die Maximum-Likelihood Methode läßt sich ohne weiteres auf den Fall von mehreren Parametern ( $p > 1$ ) ausdehnen.

### Beispiel 3: Normalverteilung

Seien  $X_1, X_2, \dots, X_n$  unabhängig und identisch normalverteilt mit unbekanntem Mittelwert  $\mu$  und unbekannter Varianz  $\sigma^2$ . Hier ist dann  $\theta = (\mu, \sigma^2)$  und

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \theta) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i - \mu)^2 / 2\sigma^2} \\ &= \frac{1}{\sqrt{(2\pi)^n \sigma^{2n}}} e^{-\sum_{i=1}^n (x_i - \mu)^2 / 2\sigma^2} \end{aligned}$$

Wir betrachten nun

$$\begin{aligned} l_\theta(x_1, \dots, x_n) &= \log L(x_1, \dots, x_n; \theta) \\ &= -n \log \sigma - n \log(\sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

Notwendig für ein Maximum ist

a)  $\frac{\partial}{\partial \mu} l_\theta|_{\hat{\theta}} = 0$

b)  $\frac{\partial}{\partial \sigma} l_\theta|_{\hat{\theta}} = 0.$

Die Gleichung a)

$$0 = \frac{\partial}{\partial \mu} l_\theta(x_1, \dots, x_n) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)$$

liefert den Schätzwert  $\hat{\mu}(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i.$

Die Gleichung b)

$$0 = \frac{\partial}{\partial \sigma} l_\theta = -\frac{n}{\sigma} + \frac{1}{2\sigma^3} \sum_{i=1}^n (x_i - \mu)^2$$

liefert nach Einsetzen von  $\hat{\mu}$  den Schätzwert

$$\hat{\sigma}^2(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

## 5.2 Methode der Momente

Eine Alternative zum Maximum-Likelihood Ansatz stellt die Methode der Momente dar. Die Verteilung  $F$  der Beobachtungen sei wieder bis auf einen *Parameter*  $\theta \in \Theta \subset \mathbf{R}^p$  bekannt. Dann hängt der Erwartungswert von  $X^r$  von  $\theta$  ab. Für  $r = 1, 2, \dots, p$  sei

$$m_r(\theta) = E_\theta(X^r)$$

und

$$\hat{m}_r = \frac{1}{n} \sum_{k=1}^n X_k^r$$

das *empirische*  $r$ -te Moment. Die Grundidee der Methode der Momente besteht darin,  $\theta$  so zu bestimmen, daß

$$m_r(\theta) = \hat{m}_r$$

für  $r = 1, 2, \dots, p$  gilt. Man beachte, daß man damit  $p$  Gleichungen bei  $p$  Unbekannten erhält.

**Beispiel:** Gamma-Verteilung

$X_1, \dots, X_n$  seien unabhängig und jeweils stetig verteilt mit Dichte

$$f_{\alpha,\lambda}(x) = \begin{cases} 0 & \text{für } x \leq 0 \\ \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{für } x > 0 \end{cases}.$$

Dabei sind  $\alpha, \lambda > 0$  und  $\Gamma$  bezeichnet die Gammafunktion

$$\Gamma(u) = \int_0^\infty x^{u-1} e^{-x} dx.$$

Man kann zeigen, daß

$$\frac{\alpha}{\lambda} = \int_0^\infty x f_{\alpha,\lambda}(x) dx$$

und

$$\frac{\alpha}{\lambda^2} + \left(\frac{\alpha}{\lambda}\right)^2 = \int_0^\infty x^2 f_{\alpha,\lambda}(x) dx$$

gilt. Damit erhält man die Gleichungen

$$\begin{aligned} \hat{m}_1 &= \frac{\alpha}{\lambda} \\ \hat{m}_2 &= \frac{\alpha}{\lambda^2} + \left(\frac{\alpha}{\lambda}\right)^2. \end{aligned}$$

Aus der ersten Gleichung folgt  $\lambda = \alpha/\hat{m}_1$ . Setzen wir dies in die zweite Gleichung ein, so ergibt sich

$$\hat{m}_2 = \frac{(\hat{m}_1)^2}{\alpha} + (\hat{m}_1)^2.$$

Also ist

$$\begin{aligned} \hat{\alpha} &= \frac{(\hat{m}_1)^2}{\hat{m}_2 - (\hat{m}_1)^2} \\ \hat{\lambda} &= \frac{\hat{m}_1}{\hat{m}_2 - (\hat{m}_1)^2}. \end{aligned}$$

## 5.3 Konfidenzintervalle

Sei  $X$  eine Zufallsvariable, die nach  $F$  verteilt ist und sei

$$\mu = E(X) \quad \text{und} \quad \sigma^2 = \text{Var}(X).$$

Es ist naheliegend, den Mittelwert  $\mu$  durch das empirische Mittel  $\bar{X}_n$  mit

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$$

zu schätzen. Dieser Schätzer ist eine Zufallsvariable. Ein konkreter Schätzwert entsteht durch Einsetzen der Meßwerte, also durch Ersetzen von  $X_k$  durch  $x_k$ . Die Varianz  $\sigma^2$  kann man durch  $\hat{\sigma}_n^2$  mit

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

schätzen. Beide Schätzer haben die Eigenschaft der sogenannten Erwartungstreue, das heißt, es gilt

$$E(\bar{X}_n) = \mu \quad \text{und} \quad E(\hat{\sigma}_n^2) = \sigma^2.$$

Dies ist für beliebige Verteilungen richtig, solange Mittelwert und Varianz endlich sind. Die Behauptung für  $\hat{\sigma}_n^2$  sieht man folgendermaßen. Zunächst gilt

$$\begin{aligned} \sum_{k=1}^n (X_k - \bar{X}_n)^2 &= \sum_{k=1}^n \{X_k^2 - 2X_k\bar{X}_n + (\bar{X}_n)^2\} \\ &= \left\{ \sum_{k=1}^n X_k^2 \right\} - n(\bar{X}_n)^2. \end{aligned}$$

Damit folgt

$$\begin{aligned} E\left(\sum_{k=1}^n (X_k - \bar{X}_n)^2\right) &= nE(X_1^2) - nE[(\bar{X}_n)^2] \\ &= nE(X_1^2) - \frac{1}{n}E\left[\left(\sum_{k=1}^n X_k\right)^2\right] \\ &= nE(X_1^2) - \frac{1}{n}E\left[\sum_{k=1}^n \sum_{j=1}^n X_k X_j\right] \\ &= nE(X_1^2) - \frac{n}{n}E(X_1^2) - \frac{n(n-1)}{n}[E(X_1)]^2 \\ &= (n-1)\text{Var}(X_1). \end{aligned}$$

Für die Varianz des Schätzers  $\bar{X}_n$  gilt

$$\text{Var}(\bar{X}_n) = \frac{1}{n} \text{Var}(X_1 + \dots + X_n) = \frac{\sigma^2}{n}.$$

Ist  $E(X^4) < \infty$ , so gilt für die Varianz von  $\hat{\sigma}_n^2$

$$\text{Var}(\hat{\sigma}_n^2) = \frac{\sigma^4}{n(n-1)} \left( (n-1) \frac{\mu_4}{\sigma^4} - (n-3) \right)$$

mit  $\mu_4 = E[(X - \mu)^4]$ .

Die Schätzer  $\bar{X}_n$  und  $\hat{\sigma}_n^2$  besitzen beide die Eigenschaft der sogenannten *Konsistenz*. Es gilt nämlich für beliebiges  $\varepsilon > 0$  bei wachsendem Stichprobenumfang  $n$

- a)  $\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0$
- b)  $\lim_{n \rightarrow \infty} P(|\hat{\sigma}_n^2 - \sigma^2| > \varepsilon) = 0.$

Diese Aussagen folgen mit der Tschebychev-Ungleichung aus Abschnitt 3.5.

Wir beschäftigen uns nun mit der Konstruktion von Schätzintervallen. Solche Intervalle sollen den Parameter mit einer großen Wahrscheinlichkeit enthalten. Konstruiert werden sie hier mit Hilfe von Punktschätzern.

**Definition 5.3.1** Ein zufälliges Intervall  $I(X_1, \dots, X_n)$  heißt *Konfidenzintervall* für  $\theta$  zur Sicherheitswahrscheinlichkeit  $1 - \alpha$ , falls für alle  $\theta \in \Theta$  gilt

$$P_\theta(\theta \in I(X_1, \dots, X_n)) \geq 1 - \alpha.$$

Bei der Interpretation von Konfidenzintervallen ist Vorsicht angebracht. Der Parameter  $\theta$  wird nicht als zufällig angenommen. Die Größe  $1 - \alpha$  gibt die Wahrscheinlichkeit an, daß ein vorliegendes Konfidenzintervall  $\theta$  enthält. Diese Wahrscheinlichkeitsaussage bezieht sich auf wiederholtes Anwenden in folgendem Sinne. Hat man aus 100 unabhängigen Experimenten je ein 95 %-Konfidenzintervall gewonnen, so werden etwa 95 dieser Intervalle den wahren Wert enthalten.

Wir betrachten nun den Fall normalverteilter Beobachtungen. Seien also  $X_1, \dots, X_n$  unabhängig und  $N(\mu, \sigma^2)$ -verteilt.

Zunächst nehmen wir an, daß  $\mu$  unbekannt ist und  $\sigma^2$  bekannt. Wir schätzen  $\mu$  durch

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

$\bar{X}_n$  ist nun normalverteilt mit Mittelwert  $\mu$  und Varianz  $\frac{\sigma^2}{n}$ . Folglich ist

$$\frac{\sqrt{n} (\bar{X}_n - \mu)}{\sigma}$$

nach  $N(0, 1)$  verteilt. Sei  $k_\alpha$  das  $\alpha$ -Quantil der Standardnormalverteilung, das heißt  $\Phi(k_\alpha) = \alpha$ . Dann gilt  $-k_\alpha = k_{1-\alpha}$ . Es gilt nun

$$P \left( -k_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n} (\bar{X}_n - \mu)}{\sigma} \leq k_{1-\frac{\alpha}{2}} \right) = 1 - \alpha.$$

Nun ist die Aussage

$$-k_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq k_{1-\frac{\alpha}{2}}$$

äquivalent zu

$$\bar{X}_n - \frac{\sigma}{\sqrt{n}}k_{1-\frac{\alpha}{2}} \leq \mu \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}}k_{1-\frac{\alpha}{2}}.$$

Setzt man nun

$$I(X_1, X_2, \dots, X_n) = \left[ \bar{X}_n - \frac{\sigma}{\sqrt{n}}k_{1-\frac{\alpha}{2}}, \bar{X}_n + \frac{\sigma}{\sqrt{n}}k_{1-\frac{\alpha}{2}} \right]$$

so gilt

$$P(\mu \in I(X_1, X_2, \dots, X_n)) = 1 - \alpha, \quad \text{für alle } \mu \in \mathbb{R}.$$

Man hat also ein  $(1 - \alpha)$ -Konfidenzintervall für  $\mu$  konstruiert. Ganz ähnlich argumentiert man, wenn  $\sigma^2$  unbekannt ist. Man ersetzt lediglich  $\sigma$  durch  $\hat{\sigma}_n$  und argumentiert entsprechend. Formal geht es folgendermaßen.

Die Parameter  $\mu$  und  $\sigma^2$  seien beide unbekannt. Wir können sie durch

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$$

und

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

schätzen. Der grundlegende Ansatz, um ein (exaktes) Konfidenzintervall zu erhalten, liegt darin, eine geeignete Größe zu finden, deren Verteilung nicht von den unbekanntem Parametern abhängt. Man spricht von einer sogenannten *Pivot-Statistik*. Die Größe

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\hat{\sigma}_n}$$

hat diese Eigenschaft. Wir werden dies gleich nachweisen. Natürlich können wir diese Größe nicht wirklich bestimmen, da wir ja  $\mu$  nicht kennen. Trotzdem werden wir mit ihrer Hilfe ein Konfidenzintervall für  $\mu$  herleiten können. Sei für  $k = 1, \dots, n$

$$Y_k = \frac{X_k - \mu}{\sigma}.$$

Dann sind  $Y_1, \dots, Y_n$  unabhängig und jeweils standard normalverteilt. Es gilt

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\hat{\sigma}_n} = \frac{\sqrt{n}\bar{Y}_n}{\sqrt{\frac{1}{n-1} \sum_{k=1}^n (Y_k - \bar{Y}_n)^2}}$$

mit  $\bar{Y}_n = (1/n) \sum_{k=1}^n Y_k$ . Die Verteilung der Größe auf der rechten Seite hängt offensichtlich nicht von  $\mu$  und  $\sigma^2$  ab. Diese Verteilung heißt Student'sche  $t$ -Verteilung mit  $n-1$  Freiheitsgraden. Kurz:  $t_{n-1}$ -Verteilung. Sie ist symmetrisch um Null und besitzt die Dichte

$$\frac{\Gamma(n/2)}{\Gamma((n-1)/2)} \frac{1}{\sqrt{(n-1)\pi}} \left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}.$$

Sei für  $0 < \alpha < 1$   $t_{\alpha;n-1}$  das  $\alpha$ -Quantil dieser Verteilung. Dann ist

$$P \left( \frac{\sqrt{n}\bar{Y}_n}{\sqrt{\frac{1}{n-1} \sum_{k=1}^n (Y_k - \bar{Y}_n)^2}} \leq t_{\alpha;n-1} \right) = \alpha.$$

Sei  $z_\alpha$  das  $\alpha$ -Quantil der Standard Normalverteilung. Also  $\Phi(z_\alpha) = \alpha$ . Man kann zeigen, daß für  $0 < \alpha \leq 1/2$

$$t_{\alpha;n-1} \leq z_\alpha \leq 0$$

gilt und für  $1/2 \leq \alpha < 1$

$$0 \leq z_\alpha \leq t_{\alpha;n-1}.$$

Außerdem ist  $\lim_{n \rightarrow \infty} t_{\alpha;n-1} = z_\alpha$  für  $0 < \alpha < 1$ . Aus der Symmetrie der  $t_{n-1}$ -Verteilung folgt

$$P \left( -t_{1-\alpha/2;n-1} \leq \frac{\sqrt{n}\bar{Y}_n}{\sqrt{\frac{1}{n-1} \sum_{k=1}^n (Y_k - \bar{Y}_n)^2}} \leq t_{1-\alpha/2;n-1} \right) = 1 - \alpha$$

und somit auch

$$P \left( -t_{1-\alpha/2;n-1} \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{\hat{\sigma}_n} \leq t_{1-\alpha/2;n-1} \right) = 1 - \alpha.$$

Löst man die Ungleichungen

$$-t_{1-\alpha/2;n-1} \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{\hat{\sigma}_n} \leq t_{1-\alpha/2;n-1}$$

nach  $\mu$  auf, so erhält man

$$\mu \in \left[ \bar{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} t_{1-\alpha/2;n-1}, \bar{X}_n + \frac{\hat{\sigma}_n}{\sqrt{n}} t_{1-\alpha/2;n-1} \right].$$

Damit gilt

$$P \left( \mu \in \left[ \bar{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} t_{1-\alpha/2;n-1}, \bar{X}_n + \frac{\hat{\sigma}_n}{\sqrt{n}} t_{1-\alpha/2;n-1} \right] \right) = 1 - \alpha.$$

Ein  $(1 - \alpha)$ -Konfidenzintervall für  $\mu$  ist also durch

$$I(X_1, \dots, X_n) = \left[ \bar{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} t_{1-\alpha/2;n-1}, \bar{X}_n + \frac{\hat{\sigma}_n}{\sqrt{n}} t_{1-\alpha/2;n-1} \right]$$

gegeben. Dieses Intervall liegt symmetrisch um den Punktschätzer  $\bar{X}_n$ .

Die Breite des Intervalls wird wesentlich von  $n$  bestimmt; sie schrumpft etwa wie  $1/\sqrt{n}$ . Dies bedeutet, daß für doppelte Genauigkeit (also halbe Breite) schon viermal so viele Beobachtungen benötigt werden.

Um ein Konfidenzintervall für  $\sigma^2$  zu erhalten, betrachtet man die Pivot-Statistik

$$\frac{(n-1)\hat{\sigma}_n^2}{\sigma^2} = \frac{\sum_{k=1}^n (X_k - \bar{X}_n)^2}{\sigma^2}.$$

Die Verteilung dieser Größe hängt nicht von  $\mu$  oder  $\sigma^2$  ab. Diese Verteilung heißt  $\chi^2$ -Verteilung mit  $n - 1$  Freiheitsgraden. Kurz  $\chi_{n-1}^2$ -Verteilung. Sie ist auf  $(0, \infty)$  konzentriert und besitzt die Dichte

$$2^{\frac{1-n}{2}} \frac{1}{\Gamma((n-1)/2)} x^{\frac{n-1}{2}-1} e^{-\frac{x}{2}}.$$

Sie ist eine Gamma-Verteilung mit Parametern  $\alpha = \frac{n-1}{2}$  und  $\beta = \frac{1}{2}$ .

Sei für  $0 < \alpha < 1$   $\chi_{\alpha;n-1}^2$  das  $\alpha$ -Quantil der  $\chi_{n-1}^2$ -Verteilung.

Ein  $(1 - \alpha)$ -Konfidenzintervall für  $\sigma^2$  ist gegeben durch

$$\left[ \frac{(n-1)\hat{\sigma}_n^2}{\chi_{1-\alpha/2;n-1}^2}, \frac{(n-1)\hat{\sigma}_n^2}{\chi_{\alpha/2;n-1}^2} \right].$$

Konfidenzintervalle müssen nicht notwendigerweise an beiden Seiten beschränkt sein. Manchmal ist man an einseitigen Konfidenzaussagen interessiert. Für den Mittelwert  $\mu$  erhält man bei der Normalverteilung das einseitige untere  $(1 - \alpha)$ -Konfidenzintervall

$$\left[ \bar{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} t_{1-\alpha;n-1}, \infty \right)$$

und das einseitige obere  $(1 - \alpha)$ -Konfidenzintervall

$$\left( -\infty, \bar{X}_n + \frac{\hat{\sigma}_n}{\sqrt{n}} t_{1-\alpha;n-1} \right].$$

Ein einseitiges unteres Konfidenzintervall ist von Interesse, wenn man nur feststellen will, daß  $\mu$  nicht zu niedrig ist.

**Beispiel:**

Sei  $n = 8$ . Es liegen folgende Beobachtungen vor:  $x_1 = 464$ ,  $x_2 = 469.7$ ,  $x_3 = 469.2$ ,  $x_4 = 469.5$ ,  $x_5 = 461.8$ ,  $x_6 = 468.7$ ,  $x_7 = 469.5$ ,  $x_8 = 463.9$ . Dann gilt

$$\bar{x}_8 = \frac{1}{8} \sum_{k=1}^8 x_k = 467.09$$

und

$$\hat{\sigma}_8^2 = \frac{1}{7} \sum_{k=1}^8 (x_k - \bar{x}_8)^2 = 10.11.$$

Es ist weiter  $t_{0.975;7} = 2.365$ . Unter Normalverteilungsannahme erhält man damit als 95 % Konfidenzintervall für den Mittelwert  $\mu$

$$\left[ 467.09 - 2.365 \times \frac{\sqrt{10.11}}{\sqrt{8}}, 467.09 + 2.365 \times \frac{\sqrt{10.11}}{\sqrt{8}} \right] = [464.43, 469.75].$$

Mit  $\chi_{0.975;7}^2 = 16.01$  und  $\chi_{0.025;7}^2 = 1.69$  erhält man das folgende 95 %-Konfidenzintervall für die Varianz  $\sigma^2$

$$\left[ \frac{7 \times 10.11}{16.01}, \frac{7 \times 10.11}{1.69} \right] = [4.42, 41.88].$$

## 5.4 Hypothesentests

Wir beschäftigen uns in diesem Abschnitt mit dem Problem der Überprüfung von Hypothesen. Wir gehen wieder davon aus, daß die Verteilung der Beobachtungen  $X_1, \dots, X_n$  bis auf einen Parameter  $\theta \in \Theta$  bekannt ist. Eine Hypothese ist durch eine Aufteilung des Parameterbereichs  $\Theta$  in zwei Teilmengen  $\Theta_0$  und  $\Theta_1$  gegeben. Man spricht zum einen von der *Nullhypothese*  $H_0 : \theta \in \Theta_0$ . Das Gegenteil – die sogenannte *Alternative* – ist durch das Komplement  $\Theta_1 = \Theta \setminus \Theta_0$  festgelegt. Man schreibt  $H_1 : \theta \in \Theta_1$ . Ein statistischer Test ist eine (datenabhängige) Entscheidungsregel, die angibt, welche der beiden sich ausschließenden Hypothesen für wahr gehalten wird. Formal handelt es sich dabei um eine Abbildung  $\phi$  von den Beobachtungen  $(x_1, \dots, x_n)$  in die Menge  $\{0, 1\}$ . Dabei steht  $\phi = 1$  für die Entscheidung für die Alternative und  $\phi = 0$  für die Hypothese. Es bestehen folgende vier Möglichkeiten:

	Entscheidung für $H_0$	Entscheidung für $H_1$
$H_0$ liegt vor	richtige Entscheidung	<i>Fehler 1. Art</i>
$H_1$ liegt vor	<i>Fehler 2. Art</i>	richtige Entscheidung

Die Wahrscheinlichkeit eines Fehlers 1. Art bezeichnet man als Niveau des Testes. Besteht  $\Theta_0$  aus mehr als einem Element, so betrachtet man üblicherweise das Maximum des Fehlers 1. Art über  $\Theta_0$ . In der Regel versucht man diese Wahrscheinlichkeit festzulegen beziehungsweise zu beschränken. Streng genommen sind damit die Rollen von  $H_0$  und  $H_1$  nicht symmetrisch. Wir können mit einem Test die Nullhypothese widerlegen, sie aber nicht wirklich beweisen. Hierzu müßten wir auch die Wahrscheinlichkeit eines Fehlers 2. Art vollständig unter Kontrolle halten.

In der Regel gibt man Tests durch eine *Teststatistik*  $T = T(X_1, \dots, X_n)$  und Festlegung eines *kritischen Bereichs*  $K$  an. Der Test lehnt ab, falls  $T \in K$  gilt. Also

$$\phi(X_1, \dots, X_n) = \begin{cases} 0 & \text{falls } T \notin K \\ 1 & \text{falls } T \in K \end{cases} .$$

Wir betrachten nun den Fall einer *einfachen* Nullhypothese näher:  $\Theta_0 = \{\theta_0\}$ . Sei  $I(X_1, \dots, X_n)$  ein  $(1 - \alpha)$ -Konfidenzintervall für  $\theta$ . Sei

$$\phi(X_1, \dots, X_n) = \begin{cases} 0 & \text{falls } \theta_0 \in I(X_1, \dots, X_n) \\ 1 & \text{falls } \theta_0 \notin I(X_1, \dots, X_n) \end{cases} .$$

Dann gilt

$$P_{\theta_0}(\phi = 1) = P_{\theta_0}(\theta_0 \notin I(X_1, \dots, X_n)) = \alpha .$$

Damit hat der Test  $\phi$  das Niveau  $\alpha$ . So kommt man zu folgenden Tests.

**Zweiseitiger  $t$ -Test:**  $X_1, \dots, X_n$  seien unabhängig und jeweils  $N(\mu, \sigma^2)$ -verteilt.  $\mu$  und  $\sigma^2$  seien unbekannt. Sei

$$H_0 : \mu = \mu_0 \quad \text{und} \quad H_1 : \mu \neq \mu_0 .$$

Ein  $(1 - \alpha)$ -Konfidenzintervall für  $\mu$  ist durch

$$I(X_1, \dots, X_n) = \left[ \bar{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} t_{1-\alpha/2; n-1}, \bar{X}_n + \frac{\hat{\sigma}_n}{\sqrt{n}} t_{1-\alpha/2; n-1} \right]$$

gegeben. Damit erhält man den Niveau- $\alpha$  Test

$$\phi(X_1, \dots, X_n) = \begin{cases} 0 & \text{falls } \sqrt{n}|\bar{X}_n - \mu_0| \leq \hat{\sigma}_n t_{1-\alpha/2; n-1} \\ 1 & \text{falls } \sqrt{n}|\bar{X}_n - \mu_0| > \hat{\sigma}_n t_{1-\alpha/2; n-1} \end{cases}.$$

**Beispiel:**

Sei  $n = 8$ . Es liegen folgende Beobachtungen vor:  $x_1 = 464$ ,  $x_2 = 469.7$ ,  $x_3 = 469.2$ ,  $x_4 = 469.5$ ,  $x_5 = 461.8$ ,  $x_6 = 468.7$ ,  $x_7 = 469.5$ ,  $x_8 = 463.9$ . Sei

$$H_0 : \mu = 469 \quad \text{und} \quad H_1 : \mu \neq 469.$$

Ein 95% Konfidenzintervall für den Mittelwert  $\mu$  ist

$$[464.43, 469.75].$$

Der Wert 469 liegt in diesem Intervall und damit lehnt der Test zum Niveau 5% die Nullhypothese nicht ab.

Tests zu einseitigen Nullhypothesen, wie etwa  $H_0 : \theta \leq \theta_0$ , lassen sich mit Hilfe von einseitigen Konfidenzintervallen konstruieren. Sei

$$I(X_1, \dots, X_n) = [U(X_1, \dots, X_n), \infty)$$

ein einseitiges unteres  $(1 - \alpha)$ -Konfidenzintervall für  $\theta$ . Sei

$$\phi(X_1, \dots, X_n) = \begin{cases} 0 & \text{falls } \theta_0 \in I(X_1, \dots, X_n) \\ 1 & \text{falls } \theta_0 \notin I(X_1, \dots, X_n) \end{cases}.$$

Dann gilt für  $\theta \leq \theta_0$ :

$$\theta_0 \notin [U(X_1, \dots, X_n), \infty) \Rightarrow \theta \notin [U(X_1, \dots, X_n), \infty).$$

Deshalb ist

$$\begin{aligned} P_\theta(\phi(X_1, \dots, X_n) = 1) &= P_\theta(\theta_0 \notin (U[X_1, \dots, X_n), \infty)) \\ &\leq P_\theta(\theta \notin (U[X_1, \dots, X_n), \infty)) \\ &= 1 - P_\theta(\theta \in (U[X_1, \dots, X_n), \infty)) \leq 1 - (1 - \alpha) = \alpha. \end{aligned}$$

Das heißt, die Wahrscheinlichkeit eines Fehlers 1. Art ist kleiner gleich  $\alpha$  für alle  $\theta \leq \theta_0$ .

**Einseitiger t-Test:**  $X_1, \dots, X_n$  seien unabhängig und jeweils  $N(\mu, \sigma^2)$ -verteilt. Sei

$$H_0 : \mu \leq \mu_0 \quad \text{und} \quad H_1 : \mu > \mu_0.$$

Ein unteres  $(1 - \alpha)$ -Konfidenzintervall für  $\mu$  ist durch

$$I_n = \left[ \bar{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} t_{1-\alpha; n-1}, \infty \right)$$

gegeben. Ist  $\mu_0 \notin I_n$ , so verwirft man  $H_0$ . Damit erhält man den Test

$$\phi(X_1, \dots, X_n) = \begin{cases} 0 & \text{falls } \sqrt{n}(\bar{X}_n - \mu_0) \leq \hat{\sigma}_n t_{1-\alpha; n-1} \\ 1 & \text{falls } \sqrt{n}(\bar{X}_n - \mu_0) > \hat{\sigma}_n t_{1-\alpha; n-1} \end{cases}.$$

**Einseitiger  $\chi^2$ -Streuungstest:**  $X_1, \dots, X_n$  seien unabhängig und jeweils  $N(\mu, \sigma^2)$ -verteilt. Sei

$$H_0 : \sigma^2 \leq \sigma_0^2 \quad \text{und} \quad H_1 : \sigma^2 > \sigma_0^2.$$

Ein unteres  $(1 - \alpha)$ -Konfidenzintervall für  $\sigma^2$  ist durch

$$\left[ \frac{(n-1)\hat{\sigma}_n^2}{\chi_{1-\alpha; n-1}^2}, \infty \right)$$

gegeben. Damit erhält man den Test

$$\phi(X_1, \dots, X_n) = \begin{cases} 0 & \text{falls } \hat{\sigma}_n^2 \leq \sigma_0^2 \chi_{1-\alpha; n-1}^2 / (n-1) \\ 1 & \text{falls } \hat{\sigma}_n^2 > \sigma_0^2 \chi_{1-\alpha; n-1}^2 / (n-1) \end{cases}.$$

**Beispiel:**

Sei  $n = 8$ . Es liegen folgende Beobachtungen vor:  $x_1 = 464$ ,  $x_2 = 469.7$ ,  $x_3 = 469.2$ ,  $x_4 = 469.5$ ,  $x_5 = 461.8$ ,  $x_6 = 468.7$ ,  $x_7 = 469.5$ ,  $x_8 = 463.9$ . Sei

$$H_0 : \sigma^2 \leq 4 \quad \text{und} \quad H_1 : \sigma^2 > 4.$$

Es gilt  $\chi_{0.95; 7}^2 = 14.07$ . Ein unteres 95% Konfidenzintervall für die Varianz  $\sigma^2$  ist damit

$$\left[ \frac{7 \times 10.11}{14.07}, \infty \right) = [5.03, \infty).$$

Dieses Intervall enthält den Wert 4 nicht. Damit lehnt der Test die Nullhypothese zum Niveau 5% ab.

Wir betrachten noch einmal das zweiseitige Testproblem

$$H_0 : \theta = \theta_0 \quad \text{gegen} \quad H_1 : \theta \neq \theta_0$$

mit *einfacher* Nullhypothese  $H_0$ . Sei  $T$  die verwendete Teststatistik. Die Wahrscheinlichkeit eines Fehlers 1. Art (Signifikanzniveau) ist durch

$$\alpha = P_{\theta_0}(T \in K)$$

gegeben. Dabei ist  $K$  der kritische Bereich des Tests

$$\phi(X_1, \dots, X_n) = \begin{cases} 0 & \text{falls } T \notin K \\ 1 & \text{falls } T \in K \end{cases}.$$

Üblicherweise wird vor Anwendung des Tests das Signifikanzniveau  $\alpha$  festgelegt und ein entsprechender kritischer Bereich  $K_\alpha$  mit  $P_{\theta_0}(T \in K_\alpha) = \alpha$  gewählt. Ist die Verteilung von  $T$  unter der Nullhypothese stetig, so gibt es zu jeder möglichen Wahl von  $0 < \alpha < 1$  einen entsprechenden Ablehnbereich. In der Regel wählt man die kritischen Bereiche so, daß gilt

$$\tilde{\alpha} < \alpha \Rightarrow K_{\tilde{\alpha}} \subset K_\alpha.$$

Je kleiner das Signifikanzniveau ist, desto zögerlicher sind wir damit, die Nullhypothese abzulehnen; damit bekommt eine eventuelle Ablehnung mehr Bedeutung. Liegen konkrete

Beobachtungen  $(x_1, \dots, x_n)$  vor, so kann man die Menge aller Signifikanzniveaus betrachten, für die

$$T(x_1, \dots, x_n) \notin K_\alpha$$

gilt. Die obere Grenze  $p^* = \max\{\alpha \in (0, 1) \mid T(x_1, \dots, x_n) \notin K_\alpha\}$  heißt *P-Wert*. Für Signifikanzniveaus, die kleiner als der *P-Wert* sind, wird  $H_0$  nicht verworfen. Für Signifikanzniveaus, die größer als der *P-Wert* sind, wird  $H_0$  verworfen. Der *P-Wert* gibt damit das Signifikanzniveau an, zu dem die Nullhypothese gerade noch verworfen wird. Er beschreibt also wie stark die vorliegenden Beobachtungen im Widerspruch zur Nullhypothese stehen. Der *P-Wert* ist deshalb von zentraler Bedeutung, weil viele Statistikprogramme *P-Werte* ausgeben. Kennt man den zu konkreten Beobachtungen  $x_1, \dots, x_n$  gehörenden *P-Wert*, so kann man einen Test zum Niveau  $\alpha_0$  (etwa  $\alpha_0 = 0.05$ ) folgendermaßen durchführen:

- ist der *P-Wert* kleiner als  $\alpha_0$ , wird  $H_0$  abgelehnt.
- ist der *P-Wert* größer als  $\alpha_0$ , wird  $H_0$  nicht abgelehnt.

Betrachten wir das letzte Beispiel so ist

$$\frac{7 \cdot 10, 11}{\chi_{1-p^*;7}^2} = 4$$

und damit

$$\chi_{1-p^*;7}^2 = 17, 69.$$

Es folgt  $p^* = 0, 013$  ist der zugehörige *P-Wert*.

*P-Werte* und Signifikanzniveaus lassen sich auch für *zusammengesetzte* Nullhypothesen betrachten. In diesem Fall zieht man  $\alpha = \sup_{\theta \in \Theta_0} P_\theta(T \in K)$  heran.

## 5.5 Approximative Konfidenzintervalle und Hypothesentests

Es ist nicht immer möglich, exakte Konfidenzintervalle über eine Pivot-Statistik zu erhalten. In diesem Fall kann man versuchen, mittels des zentralen Grenzwertsatzes approximative Konfidenzintervalle zu gewinnen. Seien  $X_1, \dots, X_n$  unabhängige Beobachtungen mit identischer Verteilung. Wir nehmen wieder an, daß diese Verteilung bis auf einen unbekanntem Parameter  $\theta$  bekannt ist. Sei  $\mu(\theta)$  der zugehörige Mittelwert und  $\sigma^2(\theta)$  die zugehörige Varianz. Dann ist

$$\frac{\sqrt{n}(\bar{X}_n - \mu(\theta))}{\sigma(\theta)}$$

näherungsweise standard normalverteilt. Damit gilt

$$P\left(\left|\frac{\sqrt{n}(\bar{X}_n - \mu(\theta))}{\sigma(\theta)}\right| \leq z_{1-\alpha/2}\right) \approx 1 - \alpha.$$

Dabei ist  $z_{1-\alpha/2}$  das  $1 - \alpha/2$ -Quantil der Standard Normalverteilung. Löst man nun die Ungleichung

$$\left| \frac{\sqrt{n}(\bar{X}_n - \mu(\theta))}{\sigma(\theta)} \right| \leq z_{1-\alpha/2}$$

nach  $\theta$  auf, so erhält man ein näherungsweise  $(1 - \alpha)$ -Konfidenzintervall für  $\theta$ . Eine weitere Vereinfachung ergibt sich, falls man  $\sigma(\theta)$  durch einen Schätzer, etwa  $\hat{\sigma}_n$  ersetzt und die Gleichung

$$\left| \frac{\sqrt{n}(\bar{X}_n - \mu(\theta))}{\hat{\sigma}_n} \right| \leq z_{1-\alpha/2}$$

nach  $\theta$  auflöst. Als approximatives Konfidenzintervall für  $\mu(\theta)$  ergibt sich dann sofort

$$\left[ \bar{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} z_{1-\alpha/2}, \bar{X}_n + \frac{\hat{\sigma}_n}{\sqrt{n}} z_{1-\alpha/2} \right].$$

Ist  $\theta \rightarrow \mu(\theta)$  umkehrbar, kann man dies einfach nach  $\theta$  auflösen.

**Beispiel:**(Hufschlagtote)

Das approximative 95% Konfidenzintervall für  $n = 280$  mit  $\hat{\theta}_{280} = \frac{196}{280} = 0,7$  und  $\hat{\sigma}_{280} = 0,7$  sowie  $z_{0,975} = 1,96$  lautet:

$$\left[ \hat{\theta}_{280} \pm \frac{1,96 \hat{\sigma}_{280}}{\sqrt{280}} \right] = \left[ 0,7 \pm \frac{1,96 \cdot \sqrt{0,7}}{\sqrt{280}} \right] = [0,602, 0,798].$$

**Beispiel:** (Binomial-Verteilung)

$X_1, \dots, X_n$  seien unabhängig und jeweils Bernoulli-verteilt mit Erfolgswahrscheinlichkeit  $p$ , das heißt

$$P(X_k = 0) = 1 - p \quad \text{und} \quad P(X_k = 1) = p$$

für  $k = 1, \dots, n$ . Gesucht ist ein (näherungsweise) Konfidenzintervall für  $\theta = p$ . Es gilt  $\mu(p) = p$ ,  $\sigma^2(p) = p(1 - p)$ . Die Größe

$$\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1 - p)}}$$

ist näherungsweise  $N(0, 1)$ -verteilt. Für ein näherungsweise  $(1 - \alpha)$ -Konfidenzintervall muß man deshalb die quadratische Ungleichung

$$n(\bar{X}_n - p)^2 \leq z_{1-\alpha/2}^2 p(1 - p)$$

nach  $p$  auflösen. Dies ergibt ein Intervall mit unterer Grenze

$$\frac{n}{n + z_{1-\alpha/2}^2} \left( \bar{X}_n + \frac{z_{1-\alpha/2}^2}{2n} - z_{1-\alpha/2} \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n} + \left(\frac{z_{1-\alpha/2}}{2n}\right)^2} \right)$$

und oberer Grenze

$$\frac{n}{n + z_{1-\alpha/2}^2} \left( \bar{X}_n + \frac{z_{1-\alpha/2}^2}{2n} + z_{1-\alpha/2} \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n} + \left(\frac{z_{1-\alpha/2}}{2n}\right)^2} \right).$$

Ersetzt man andererseits  $\sigma(p) = p(1-p)$  durch  $\sigma(\bar{X}_n) = \sqrt{\bar{X}_n(1-\bar{X}_n)}$  (wir ersetzen also  $p$  durch den Schätzer  $\bar{X}_n$  in  $\sigma(p)$ ), so erhält man

$$\left[ \bar{X}_n - \frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}} z_{1-\alpha/2}, \bar{X}_n + \frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}} z_{1-\alpha/2} \right].$$

Mit Hilfe eines ähnlichen Argumentes erhält man auch einen Test für

$$H_0 : \mu(\theta) = \mu(\theta_0) \text{ gegen } H_1 : \mu(\theta) \neq \mu(\theta_0)$$

der annähernd das vorgegebene Niveau  $\alpha$  einhält. Unter der Nullhypothese ist

$$\frac{\sqrt{n}(\bar{X}_n - \mu(\theta_0))}{\sigma(\theta_0)}$$

näherungsweise Standard Normalverteilt. Dies motiviert folgenden Test

$$\phi(X_1, \dots, X_n) = \begin{cases} 0 & \text{falls } \sqrt{n}|\bar{X}_n - \mu(\theta_0)| \leq z_{1-\alpha/2}\sigma(\theta_0) \\ 1 & \text{falls } \sqrt{n}|\bar{X}_n - \mu(\theta_0)| > z_{1-\alpha/2}\sigma(\theta_0) \end{cases}.$$

**Beispiel:** (Binomial-Verteilung)

$X_1, \dots, X_n$  seien unabhängig und jeweils Bernoulli-verteilt mit Erfolgswahrscheinlichkeit  $p$ . Also

$$P(X_k = 0) = 1 - p \text{ und } P(X_k = 1) = p$$

für  $k = 1, \dots, n$ . Dann ist ein Test für

$$H_0 : p = p_0 \text{ gegen } H_1 : p \neq p_0,$$

der annähernd das vorgegebene Niveau  $\alpha$  einhält, gegeben durch

$$\phi(X_1, \dots, X_n) = \begin{cases} 0 & \text{falls } \sqrt{n}|\bar{X}_n - p_0| \leq z_{1-\alpha/2}\sqrt{p_0(1-p_0)} \\ 1 & \text{falls } \sqrt{n}|\bar{X}_n - p_0| > z_{1-\alpha/2}\sqrt{p_0(1-p_0)} \end{cases}.$$

Hängt  $\mu$  monoton von  $\theta$  ab, so kann man den obigen Ansatz auch entsprechend auf einseitige Testprobleme anwenden.

## 5.6 $\chi^2$ -Anpassungstest

Anpassungstests beantworten die Frage, ob eine bestimmte Verteilung  $F$  die vorliegenden Daten hinreichend gut beschreibt. Wir werden in diesem Abschnitt den sogenannten  $\chi^2$ -Anpassungstest kennenlernen. Seien  $X_1, \dots, X_n$  unabhängige Beobachtungen mit identischer Verteilung. Zunächst nehmen wir an, daß der Wertebereich der Zufallsgrößen die Menge  $\{1, \dots, m\}$  ist. Dann ist die Verteilung der  $X_i$  durch

$$P(X_1 = k) = \dots = P(X_n = k) = p_k, \quad k = 1, \dots, m$$

festgelegt. Wir suchen einen Test für

$$H_0 : p_k = p_k^0 \text{ für } k = 1, \dots, m \text{ gegen } H_1 : p_k \neq p_k^0 \text{ für mindestens ein } k.$$

## Approximative Konfidenzintervalle und Hypothesentests

### Beispiel:

Will man überprüfen, ob ein Würfel fair ist, so ist  $m = 6$  und  $p_1^0 = p_2^0 = \dots = p_6^0 = 1/6$ .

Sei für  $1 \leq k \leq m$

$$N_k = |\{j | X_j = k\}|$$

die empirische Häufigkeit des Wertes  $k$  und

$$E_k = np_k^0$$

die unter der Nullhypothese erwartete Häufigkeit des Wertes  $k$ . Die folgende Größe  $\chi^2$  vergleicht die empirischen Häufigkeiten mit den erwarteten Häufigkeiten

$$\chi^2 = \sum_{k=1}^m \frac{(N_k - E_k)^2}{E_k} = \sum_{k=1}^m \frac{(N_k - np_k^0)^2}{np_k^0}.$$

Approximative Konfidenzintervalle und Hypothesentests Man kann zeigen, daß  $\chi^2$  unter der Nullhypothese näherungsweise  $\chi^2$ -verteilt ist mit  $(m-1)$  Freiheitsgraden. Die Approximation wird mit wachsender Anzahl der Beobachtungen immer besser. Dies führt zu folgendem Test für

$H_0 : p_k = p_k^0$  für  $k = 1, \dots, m$  gegen  $H_1 : p_k \neq p_k^0$  für mindestens ein  $k$

$$\phi = \begin{cases} 0 & \text{falls } \chi^2 \leq \chi_{1-\alpha; m-1}^2 \\ 1 & \text{falls } \chi^2 > \chi_{1-\alpha; m-1}^2 \end{cases}.$$

Dieser Test heißt  $\chi^2$ -Test.

Wir betrachten nun den Fall eines beliebigen Wertebereichs und werden diese Situation auf den obigen Fall zurückführen. Sei  $P$  die unbekannte Verteilung der Beobachtungen  $X_1, \dots, X_n$ . Wir suchen einen Test für

$H_0 : P = P_0$  gegen  $H_1 : P \neq P_0$ .

Hierzu teilen wir den Wertebereich der Beobachtungen in  $m$  disjunkte Teilbereiche (Zellen)  $J_1, \dots, J_m$  ein. Meist wählt man Intervalle. Sei nun

$$N_k = |\{j | X_j \in J_k\}|$$

und

$$p_k^0 = P_0(X_1 \in J_k).$$

Sei weiter  $p_k = P(X_1 \in J_k)$ , wobei  $P$  die *wahre* Verteilung der Zufallsgrößen  $X_1, \dots, X_n$  bezeichnet. Wir testen nun

$H_0 : p_k = p_k^0$  für  $k = 1, \dots, m$  gegen  $H_1 : p_k \neq p_k^0$  für mindestens ein  $k$ .

Hierzu berechnen wir wieder

$$\chi^2 = \sum_{k=1}^m \frac{(N_k - E_k)^2}{E_k} = \sum_{k=1}^m \frac{(N_k - np_k^0)^2}{np_k^0}$$

mit  $E_k = np_k^0$ . Die Reduktion auf die Zellenhäufigkeiten ist natürlich eine Vergrößerung, die zu einem gewissen Informationsverlust führt. Bei der Wahl der Zellen und ihrer Anzahl besteht eine gewisse Freiheit. Es gibt allerdings folgende Faustregel, die man beachten sollte:  $np_k^0 \geq 5$  für alle  $1 \leq k \leq m$ .

In vielen Fällen möchte man nicht eine einfache Nullhypothese überprüfen, sondern eine zusammengesetzte Nullhypothese der Form

$$H_0 : P = P_\theta \text{ für ein } \theta \in \Theta.$$

In diesem Fall bestimmt man erst einen Schätzer  $\hat{\theta}$  und bildet dann  $\chi^2$  unter Verwendung von  $E_k = nP_{\hat{\theta}}(J_k)$ . Dann ist die Verteilung von  $\chi^2$  allerdings nicht mehr annähernd  $\chi^2$ -verteilt mit  $m - 1$  Freiheitsgraden. Wählt man für  $\hat{\theta}$  den Maximum-Likelihood Schätzer bezüglich der Verteilung der Beobachtungen  $X_1, \dots, X_n$ , so liegt die Verteilung von  $\chi^2$  in gewissem Sinn zwischen einer  $\chi^2$ -Verteilung mit  $m - 1 - p$  Freiheitsgraden und einer  $\chi^2$ -Verteilung mit  $m - 1$  Freiheitsgraden. Dabei ist  $p$  die Dimension des Parameters  $\theta$ . Dies führt zu folgendem Vorgehen:

- ist  $\chi^2$  kleiner als  $\chi_{1-\alpha; m-1-p}^2$  so wird  $H_0$  nicht verworfen.
- ist  $\chi^2$  größer als  $\chi_{1-\alpha; m-1}^2$  so wird  $H_0$  verworfen.
- liegt der Wert von  $\chi^2$  zwischen beiden Grenzen, so wird keine Entscheidung getroffen.

**Beispiel:**

Als Beispiel betrachten wir den  $\chi^2$ -Anpassungstest für den Datensatz der Hufschlagtote.

$\chi^2$ -Anpassungstest (Hufschlagtote)

$j$	$n_j$	$\frac{n_j}{280}$	$p_j(\hat{\theta})$	$np_j(\hat{\theta})$	$\frac{(n_j - np_j(\hat{\theta}))^2}{np_j(\hat{\theta})}$
0	144	0,5143	0,4966	139,05	0,1762
1	91	0,325	0,3476	97,33	0,4116
2	32	0,1143	0,1217	34,08	0,1269
3	11	0,0393	0,0284	7,95	1,170
$\geq 4$	2	0,00714	0,0057	1,60	0,1

Bei 5 Gruppen ergibt sich als  $\chi^2$ -Wert  $\chi^2 = 1,985$ . Da  $\chi_{0,05;3}^2 = 2,37$  ist, (man beachte auch  $\chi_{0,9;3}^2 = 6,25$ ) wird  $H_0$  nicht verworfen. Daß  $\chi_{0,9;4}^2 = 7,78$  sei noch erwähnt.



Die Zustände sind mit den Ziffern 0 bis 8 bezeichnet, wobei "0" den Zustand benennt, bei dem noch kein Ergebnis vorliegt. Der angegebene Spielverlauf 01100101110011 übersetzt sich nun (in eindeutiger Zuordnung) in 21324621352468. Wie läßt sich nun die Aufgabe lösen? Wir bezeichnen mit  $p_i$  die Wahrscheinlichkeit bei Start im Zustand "i" den Zustand "7" zu erreichen bevor man den Zustand "8" erreicht hat. Dann gilt natürlich sofort  $p_7 = 1$  und  $p_8 = 0$ . Außerdem sind die folgenden Gleichungen intuitiv plausibel.

$$p_1 = \frac{1}{2}p_2 + \frac{1}{2}p_3$$

$$p_2 = \frac{1}{2}p_1 + \frac{1}{2}p_4$$

$$p_3 = \frac{1}{2}p_2 + \frac{1}{2}p_5$$

$$p_4 = \frac{1}{2}p_4 + \frac{1}{2}p_6$$

$$p_5 = \frac{1}{2} + \frac{1}{2}p_2$$

$$p_6 = \frac{1}{2}p_2.$$

Das Schema ist ein lineares Gleichungssystem mit sechs Unbekannten und Gleichungen. Durch Einsetzen erhält man sofort

$$p_4 = p_6 = \frac{1}{2}p_2$$

$$p_1 = \frac{3}{2}p_2$$

$$p_3 = \frac{1}{4} + \frac{3}{4}p_2$$

sowie  $p_1 = \frac{1}{8} + \frac{7}{8}p_2.$

Damit ist  $p_1 = \frac{3}{10}$ ,  $p_2 = \frac{1}{5}$  und folglich  $p_0 = \frac{1}{4}$ . Das heißt Kain gewinnt nur mit Wahrscheinlichkeit  $\frac{1}{4}$ .

## 6.2 Definition von Markov-Ketten und erste Folgerungen

**Definition 6.2.1**  $(\Omega, P)$  sei Wahrscheinlichkeitsraum,  $E$  sei endliche Menge. Seien  $X_i : \Omega \rightarrow E$ ,  $i = 0, 1, 2, 3, \dots$  Zufallsvariablen. Die Menge der Zufallsvariablen  $\mathcal{X} = \{X_0, X_1, X_2, \dots, X_n, \dots\}$  heißt Markov-Kette (der Länge  $n$ ), falls

$$P(X_i = x_i \mid X_0 = x_0, \dots, X_{i-1} = x_{i-1}) = P(X_i = x_i \mid X_{i-1} = x_{i-1})$$

für  $i = 1, 2, \dots, n, \dots$  gilt. Eine Markov-Kette  $\mathcal{X}$  heißt stationär, falls

$$P(X_i = x_1 \mid X_{i-1} = x_0) = P(X_1 = x_1 \mid X_0 = x_0)$$

für  $i = 1, 2, \dots$  gilt. Man nennt diese Größe im folgenden  $q(x_0, x_1)$ .

### Bezeichnungsweisen

1.  $E$  heißt Zustandsraum,
2.  $q(x, y), x, y \in E$  heißt stochastische Matrix, falls  $q(x, y) \geq 0$  und  $\sum_{y \in E} q(x, y) = 1$ ,  
 $\forall x \in E$  gilt,
3.  $\pi(x_0) = P(X_0 = x_0)$  mit  $x_0 \in E$  heißt Startverteilung.

## Konstruktion von stationären Markov-Ketten

Gegeben seien:

1.  $E$  endlich oder abzählbar,
2.  $q(x, y), x, y \in E$  eine stochastische Matrix,
3.  $\pi(x), x \in E$  eine Wahrscheinlichkeitsfunktion auf  $E$ , d.h.  $\pi(x) \geq 0$  für alle  $x \in E$  und  $\sum_{x \in E} \pi(x) = 1$ .

### Zur Existenz:

Sei  $\Omega_{n+1} := \{\omega = (x_0, \dots, x_n) \mid x_i \in E, i = 0, \dots, n\}$  und  $X_i(\omega) := x_i$ , für  $i = 0, 1, \dots, n$ .

**Satz 6.2.2** Sei  $n \in \mathbb{N}$ . Durch  $p(\omega) := \pi(x_0)q(x_0, x_1) \dots q(x_{n-1}, x_n)$  wird eine Wahrscheinlichkeitsfunktion auf  $\Omega_{n+1}$  gegeben, so dass gilt:

1.  $P(X_0 = x_0) = \pi(x_0)$ ,
2.  $P(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = p(\omega)$ ,
3.  $P(X_{i+1} = x_{i+1} \mid X_0 = x_0, \dots, X_i = x_i) = q(x_i, x_{i+1}) = P(X_{i+1} = x_{i+1} \mid X_i = x_i)$   
für  $i = 0, \dots, n - 1$ .

### Beweis:

**Zu (2):** Sei  $\omega = (x_0, \dots, x_n)$ . Mit  $p(\omega) = \pi(x_0)q(x_0, x_1) \dots q(x_{n-1}, x_n)$  gilt:  $p(\omega) \geq 0$  und  $\sum_{\omega \in \Omega_{n+1}} p(\omega) = 1$ . Da  $X_i(\omega) = x_i$  ist, folgt

$$P(\{\omega \mid X_i(\omega) = x_i, i = 0, \dots, n\}) = P(\{\omega\}) = p(\omega).$$

### Zu (1):

$$\begin{aligned} (+) P(\{X_0 = x_0, \dots, X_i = x_i\}) &= \sum_{x_{i+1}, \dots, x_n} \pi(x_0)q(x_0, x_1)q(x_{i-1}, x_i) \dots q(x_{n-1}, x_n) \\ &= \pi(x_0)q(x_0, x_1) \dots q(x_{i-1}, x_i) \end{aligned}$$

für  $i = 0, 1, 2, \dots, n - 1$ .

### Zu (3):

$$\begin{aligned} P(X_{i+1} = x_{i+1} \mid X_0 = x_0, \dots, X_i = x_i) &= \frac{\pi(x_0)q(x_0, x_1) \dots q(x_{i-1}, x_i), q(x_i, x_{i+1})}{\pi(x_0)q(x_0, x_1) \dots q(x_{i-1}, x_i)} \\ &= q(x_i, x_{i+1}) \end{aligned}$$

sowie

$$\begin{aligned} P(X_{i+1} = x_{i+1} \mid X_i = x_i) &= \frac{P(X_{i+1} = x_{i+1}, X_i = x_i)}{P(X_i = x_i)} \\ &= \frac{\sum_{x_0, \dots, x_{i-1}} \pi(x_0)q(x_0, x_1) \dots q(x_{i-1}, x_i)q(x_i, x_{i+1})}{\sum_{x_0, \dots, x_{i-1}} \pi(x_0)q(x_0, x_1) \dots q(x_{i-1}, x_i)} \\ &= q(x_i, x_{i+1}). \end{aligned}$$

**Bemerkungen:**

1.  $P(X_i = x_i) = \sum_{x_0, \dots, x_{i-1}} \pi(x_0)q(x_0, x_1) \dots q(x_{i-1}, x_i) = \sum_{x_0} \pi(x_0)q^i(x_0, x_i)$ .  
Dabei ist  $q^{n+1}(x, y) = \sum_{z \in E} q^n(x, z)q(z, y)$  das  $n + 1$ -fache Matrixprodukt.
2. Mit Satz 6.2.2 ist eine Markov-Kette der Länge  $n$  konstruiert!
3. Wegen (+) im Beweis von Satz 6.2.2 ist die Markov-Kette der Länge  $i$  eingebettet in eine Markov-Kette der Länge  $n$  für  $i \leq n$ .
4. Mit der Bedingung (+) lässt sich eine Markov-Kette beliebiger Länge konstruieren.

**Beispiele für Markov-Ketten:**

1) Kain und Abel-Aufgabe:

Der Zustandsraum ist  $E = \{0, 1, 2, \dots, 8\}$ . Für  $\pi$  gilt  $\pi(0) = 1$ . Weiter ist

$$(q(x, y))_{x, y \in E} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

2) Sei  $E = \{1, 2, 3\}$ ,

$$q(x, y) = \begin{pmatrix} 0 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix} \text{ und } q^2(x, y) = \begin{pmatrix} \frac{5}{8} & 0 & \frac{3}{8} \\ \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \\ 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}.$$

Für jede Startverteilung  $\pi$  mit  $\pi(x) > 0$  für  $x \in E$  gilt  $\lim_{n \rightarrow \infty} \sum_{x \in E} \pi(x)q^n(x, y) = \pi^*(y)$  mit  $\pi^*(1) = \frac{8}{19}$ ,  $\pi^*(2) = \frac{6}{19}$ ,  $\pi^*(3) = \frac{5}{19}$ .

3) Seien  $Y_1, Y_2, \dots$  unabhängig,  $X_i = \sum_{j=1}^i Y_j$  und  $X_0 = x_0$ . Dann ist  $\{X_0, X_1, X_2, \dots\}$  eine Markov-Kette.

**Beweis:**

Sei  $m \in \mathbb{N}$  und sei  $y_i = x_i - x_{i-1}$  für  $i \geq 1$  und  $y_0 = x_0$ . Dann ist:

$$\begin{aligned} &P(X_{m+1} = x | X_0 = x_0, \dots, X_m = x_m) \\ &= P(X_{m+1} = x | X_0 = y_0, X_1 = y_0 + y_1, X_2 = y_0 + y_1 + y_2, \dots, X_m = y_0 + \dots + y_m) \\ &= P(Y_{m+1} = x - (y_0 + y_1 + \dots + y_m) | Y_0 = y_0, Y_1 = y_1, \dots, Y_m = y_m) \\ &= \frac{P(Y_{m+1} = x - (y_0 + y_1 + \dots + y_m))P(Y_0 = y_0)P(Y_1 = y_1) \dots P(Y_m = y_m)}{P(Y_0 = y_0)P(Y_1 = y_1) \dots P(Y_m = y_m)} \end{aligned}$$

$$\begin{aligned}
&= P(Y_{m+1} = x - (y_0 + y_1 + \dots + y_m)) \\
&= P(X_{m+1} = x | X_m = x_m).
\end{aligned}$$

#### 4) Ehrenfest'sches Urnenmodell

$N$  Teilchen befinden sich in den Behältern 1 und 2.

$i$  Teilchen seien in Behälter 1,  $N - i$  Teilchen in Behälter 2.

In jeder Zeiteinheit springt ein Teilchen entweder von  $1 \rightarrow 2$  oder von  $2 \rightarrow 1$ .

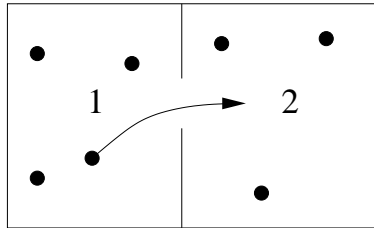


Abbildung 10.2: Ehrenfest'sches Urnenmodell

Die stochastische Matrix wird gegeben durch  $q(i, i + 1) = \frac{N-i}{N}$ ,  $q(i, i - 1) = \frac{i}{N}$  und  $q(i, j) = 0$  für  $j \neq i \pm 1$ .

### 6.3 Absorbierende Zustände\*

Wir können eine Markov-Kette als Bewegung eines Teilchens durch die Zustände auffassen. In manchen Markov-Ketten gibt es Zustände, die nicht verlassen werden können. Solche Zustände heißen absorbierend.

#### Definition 6.3.1 (absorbierender Zustand, absorbierender Rand)

Sei  $\mathcal{X} = \{X_0, X_1, \dots\}$  eine stationäre Markov-Kette mit Zustandsraum  $E$ . Ein Zustand  $x \in E$  heißt absorbierend, wenn gilt  $q(x, x) = 1$ . Die Menge  $A \subset E$  aller absorbierender Zustände heißt absorbierender Rand von  $\mathcal{X}$ .

Bei der "Kain und Abel"-Aufgabe ist  $\{7, 8\}$  der absorbierende Rand. Der folgende Satz liefert eine Begründung für die Kain und Abel-Aufgabe.

**Satz 6.3.2** Sei  $\mathcal{X}$  eine stationäre Markov-Kette und  $A \subset E$  ihr absorbierender Rand. Wir nehmen an, dass für jedes  $x \in E$  die Wahrscheinlichkeit nach  $A$  zu kommen gleich 1 ist.

Sei  $A = A_1 \cup A_2$  und  $A_1 \cap A_2 = \emptyset$ . Sei  $P_{A_1}(x)$  die Wahrscheinlichkeit bei Start in  $x$  beim ersten Eintritt in  $A$  nach  $A_1$  zu gelangen. Dann gilt:

(a)  $P_{A_1}(x) = 1$  für  $x \in A_1$  und  $P_{A_1}(x) = 0$  für  $x \in A_2$ ,

(b) Für  $x \notin A$  ist 
$$P_{A_1}(x) = \sum_{z \in A_1} q(x, z) + \sum_{n \geq 2} \sum_{\substack{x_i \notin A \text{ für} \\ i=1, \dots, n-1, \\ z \in A_1}} q(x, x_1)q(x_1, x_2) \dots q(x_{n-1}, z),$$

(c) Es gilt stets 
$$P_{A_1}(x) = \sum_{y \in E} q(x, y)P_{A_1}(y).$$

**Beweis:**

**Zu (a):** Dies folgt direkt aus der Definition.

**Zu (b):** Sei  $x \notin A$ . Dann ist

$$\begin{aligned} P_{A_1}(x) &= P(X_0 = x, X_1 \in A_1) + P\left(\bigcup_{n \geq 2} \{X_0 = x, X_i \notin A \text{ für } i < n, X_n \in A_1\}\right) \\ &= \sum_{z \in A_1} q(x, z) + \sum_{n \geq 2} P(\{X_0 = x, X_i \notin A \text{ für } i < n, X_n \in A_1\}) \\ &= \sum_{z \in A_1} q(x, z) + \sum_{n \geq 2} \sum_{\substack{x_1, x_2, \dots, x_{n-1} \notin A, \\ z \in A_1}} q(x, x_1) \dots q(x_{n-1}, z). \end{aligned}$$

**Zu (c):** Sei  $x \notin A$ . So gilt

$$\begin{aligned} P_{A_1}(x) &= P(X_0 = x, X_1 \in A_1) + P(X_0 = x, X_1 \notin A, X_2 \in A_1) \\ &\quad + P(X_0 = x, \exists n \geq 3 \text{ mit } X_n \in A_1 \text{ und } X_i \notin A \text{ für } i < n) \\ &= \sum_{y \in A_1} q(x, y) + \sum_{y \notin A, z \in A_1} q(x, y)q(y, z) \\ &\quad + \sum_{n \geq 3} \sum_{y \notin A} \sum_{\substack{x_i \notin A, \\ i=2, \dots, n-1, \\ z \in A_1}} q(x, y)q(y, x_2) \dots q(x_{n-1}, z) \\ &\stackrel{(*)}{=} \sum_{y \in A} q(x, y)P_{A_1}(y) + \sum_{y \notin A} q(x, y) \left( \sum_{z \in A_1} q(y, z) \right) \\ &\quad + \sum_{y \notin A} q(x, y) \left( \sum_{n \geq 3} \sum_{\substack{x_i \notin A, \\ i=2, \dots, n-1, \\ z \in A_1}} q(y, x_2) \dots q(x_{n-1}, z) \right) \\ &= \sum_{y \in A} q(x, y)P_{A_1}(y) + \sum_{y \notin A} q(x, y)P_{A_1}(y) \\ &= \sum_y q(x, y)P_{A_1}(y). \end{aligned}$$

(\*)  $P_{A_1}(x) = 1$  für  $x \in A_1$  und  $P_{A_1}(x) = 0$  für  $x \in A \setminus A_1$ .

**Bemerkung:** Die Gleichung c) aus Satz 6.3.2 lautet in Vektorschreibweise  $P_{A_1} = qP_{A_1}$ . Dies bedeutet  $P_{A_1}$  ist rechter Eigenvektor von  $q$  zum Eigenwert 1. Man sagt auch,  $P_{A_1}$  ist harmonisch oder  $P_{A_1}$  erfüllt die Mittelwerteigenschaft.

## Berechnung von Ruin-Wahrscheinlichkeiten

Hans und Rudolf spielen ein Spiel. In jeder Runde gewinnt Hans mit der Wahrscheinlichkeit  $p$  und Rudolf gewinnt mit der Wahrscheinlichkeit  $q = 1 - p$ . Der Gewinner einer Runde erhält von seinem Gegner einen Euro. Es wird so lange gespielt bis einer der Spieler kein Geld mehr hat. Wie hoch ist die Ruin-Wahrscheinlichkeit  $P(x)$  von Hans, wenn Hans zu Beginn  $x$  Euro hat und Hans und Rudolf zusammen  $b$  Euro haben?

Seien  $x, b \in \mathbb{N}$  mit  $0 < x < b$ . Weiter seien  $X_1, X_2, \dots$  unabhängige Zufallsvariablen mit  $P(X_i = 1) = p = 1 - P(X_i = -1)$  für alle  $i$  und es sei  $S_0 = x$  und  $S_n = x + \sum_{i=1}^n X_i$  für  $n \geq 1$ . Damit ist  $S_0, S_1, S_2, \dots$  eine Markov-Kette und  $S_n$  ist das Kapital, das Hans nach  $n$  Runden besitzt. Wir frage nach der Ruin-Wahrscheinlichkeit:

$$P(x) = P(\exists n \text{ mit } S_n = 0 \text{ und } 0 < S_i < b \text{ für } i < n \mid S_0 = x).$$

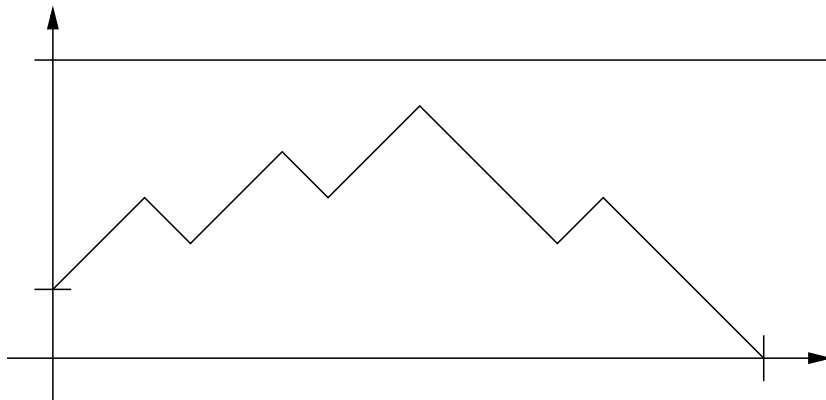


Abbildung 10.3: Ruin-Wahrscheinlichkeiten

**Satz 6.3.3** Sei  $q = 1 - p$ .

Für  $p = \frac{1}{2}$  gilt  $P(x) = 1 - \frac{x}{b}$  für  $0 \leq x \leq b$ .

Für  $p \neq \frac{1}{2}$  gilt  $P(x) = \frac{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^x}{\left(\frac{q}{p}\right)^b - 1}$  für  $0 \leq x \leq b$ .

**Beweis:**

**Fall 1:** Sei  $p = \frac{1}{2}$ .

Satz 6.3.2 liefert:  $P(x) = \frac{1}{2}P(x-1) + \frac{1}{2}P(x+1)$  für  $0 < x < b$ ,  $P(0) = 1$ ,  $P(b) = 0$ .  
Durch Lösen dieses linearen Gleichungssystems ergibt sich:  $P(x) = 1 - \frac{x}{b}$ .

**Fall 2:** Sei  $p \neq \frac{1}{2}$ .

Satz 6.3.2 liefert:  $P(x) = pP(x+1) + qP(x-1)$  für  $0 < x < b$ ,  $P(b) = 0$ ,  $P(0) = 1$ .

Dann ist

$$\begin{aligned} P(x) &= pP(x+1) + qP(x-1) \\ pP(x) + qP(x) &= pP(x+1) + qP(x-1) \\ P(x+1) - P(x) &= \frac{q}{p}(P(x) - P(x-1)). \end{aligned}$$

Wiederholtes Anwenden liefert

$$P(x+1) - P(x) = \left(\frac{q}{p}\right)^x (P(1) - P(0)).$$

Sei  $r := \frac{q}{p}$ .

Aufaddieren liefert  $P(x) - P(0) = \frac{r^x - 1}{r - 1} (P(1) - P(0))$ .

Wegen  $P(0) = 1$  gilt  $P(x) = 1 + \frac{r^x - 1}{r - 1} (P(1) - 1)$ .

Wegen  $P(b) = 0$  gilt  $P(1) - 1 = -\frac{r-1}{r^b-1}$ .

Einsetzen in die vorangegangene Gleichung liefert die Behauptung.

## 6.4 Rekurrente und transiente Zustände

### 6.4.1 Bezeichnungen

Sei  $f^n(x, y) := P(X_n = y, X_i \neq y \text{ für } 1 \leq i < n | X_0 = x)$ . Dann bezeichnet  $f^n(x, y)$  die Wahrscheinlichkeit bei Start in  $x$  nach  $n$  Schritten erstmals den Zustand  $y$  zu erreichen.

$f^*(x, y) := P(X_n = y \text{ für ein } n \geq 1 | X_0 = x) = \sum_{n \geq 1} f^n(x, y)$  ist die Wahrscheinlichkeit bei Start in  $x$  irgendwann nach  $y$  zu gelangen.

$q^*(x, y) := \sum_{n=0}^{\infty} q^n(x, y)$  ist die erwartete Anzahl der Besuche in  $y$  bei Start in  $x$ , denn:

$$\begin{aligned} q^*(x, y) &= \sum_{n=0}^{\infty} q^n(x, y) = \sum_{n=0}^{\infty} P(X_n = y | X_0 = x) \\ &= \sum_{n=0}^{\infty} E(1_{\{X_n=y\}} | X_0 = x) = E\left(\sum_{n=0}^{\infty} 1_{\{X_n=y\}} \middle| X_0 = x\right). \end{aligned}$$

**Definition 6.4.1 (rekurrenter Zustand)**  $x \in E$  heißt *rekurrent*, falls  $f^*(x, x) = 1$ ,  $x \in E$  heißt *transient*, falls  $f^*(x, x) < 1$ .

**Satz 6.4.2** 1.  $x \in E$  ist rekurrent  $\Leftrightarrow q^*(x, x) = \infty$ ,

2.  $x \in E$  ist transient  $\Leftrightarrow q^*(x, x) < \infty$ .

**Beweis:** Dieser folgt aus nachfolgendem Lemma.

**Lemma 6.4.3\*** Es gilt

a)  $q^*(x, y) = f^*(x, y) q^*(y, y) + \delta(x, y)$ ,

b)  $f^*(x, x) = \frac{q^*(x, x) - 1}{q^*(x, x)}$ ,

c)  $q^*(x, x) = \frac{1}{1 - f^*(x, x)}$ .

**Beweis:**

Für alle  $x, y \in E$  gilt:  $q^n(x, y) = \sum_{i=1}^n f^i(x, y) q^{n-i}(y, y)$ . Denn es gilt mit

$$T_y = \min\{m \geq 1 | X_m = y\}$$

$$\begin{aligned} q^n(x, y) &= P(X_n = y | X_0 = x) \\ &= \sum_{i=1}^n P(T_y = i | X_0 = x) P(X_n = y | T_y = i, X_0 = x) \\ &= \sum_{i=1}^n f^i(x, y) P(X_n = y | X_i = y) \\ &= \sum_{i=1}^n f^i(x, y) P(X_{n-i} = y | X_0 = y) \\ &= \sum_{i=1}^n f^i(x, y) q^{n-i}(y, y). \end{aligned}$$

Weiter gilt nach Vertauschung der Summation

$$\begin{aligned}
 q^*(x, y) &= \sum_{i=1}^{\infty} q^n(x, y) + \delta(x, y) \\
 &= \sum_{n=1}^{\infty} \sum_{i=1}^n f^i(x, y) q^{n-i}(y, y) + \delta(x, y) \\
 &= \sum_{i=1}^{\infty} \sum_{n=i}^{\infty} f^i(x, y) q^{n-i}(y, y) + \delta(x, y) \\
 &= \sum_{i=1}^{\infty} f^i(x, y) \sum_{m=0}^{\infty} q^m(y, y) + \delta(x, y) \\
 &= f^*(x, y) \cdot q^*(y, y) + \delta(x, y).
 \end{aligned}$$

Damit ist a) gezeigt. Für  $x = y$  folgt  $q^*(x, x) - f^*(x, x)q^*(x, x) = 1$ .

Auflösen nach  $q^*(x, x)$  liefert  $q^*(x, x)(1 - f^*(x, x)) = 1$  und damit  $q^*(x, x) = \frac{1}{1 - f^*(x, x)}$ .

Auflösen nach  $f^*(x, x)$  liefert:  $f^*(x, x) = \frac{q^*(x, x) - 1}{q^*(x, x)}$ .

### Beispiel auf $\mathbb{Z}$ (Irrfahrt auf $\mathbb{Z}$ ):

Seien  $Y_1, Y_2, \dots$  unabhängige Zufallsvariable mit  $P(Y_i = 1) = p$ ,  $P(Y_i = -1) = q$  für alle  $i$ . Weiter sei  $X_0 = 0$  und  $X_n = Y_1 + Y_2 + \dots + Y_n$  für  $n \geq 1$ .

Behauptung: Für  $p = q = \frac{1}{2}$  ist 0 rekurrent, für  $p \neq q$  ist 0 transient.

#### Beweis:

Wende Satz 6.4.2 an! Sei  $n \in \mathbb{N}$ . Nach  $2n + 1$  Schritten liegt ein ungerader Zustand vor, also kann  $X_{2n+1}$  nicht 0 sein und damit gilt:  $q^{2n+1}(0, 0) = 0$ . Fällt in  $2n$  Experimenten insgesamt  $n$ -mal die 1 und  $n$ -mal die  $-1$ , so ist  $X_{2n} = 0$  und es gilt:  $q^{2n}(0, 0) = \binom{2n}{n} p^n q^n$ . Für große  $n$  können wir die Stirling'schen Formel anwenden:

$$q^{2n}(0, 0) \sim \frac{1}{\sqrt{\pi n}} (4pq)^n, \text{ da } \binom{2n}{n} \sim \frac{1}{\sqrt{\pi n}} 2^{2n}.$$

**Fall 1:**  $p \neq q \Rightarrow (4pq) < 1 \Rightarrow \sum_{n \geq 1} q^{2n}(0, 0) = \sum_{n \geq 1} \frac{1}{\sqrt{\pi n}} (4pq)^n (1 + o(1)) < \infty \Rightarrow$  „0“ ist transient.

**Fall 2:**  $p = q = \frac{1}{2} \Rightarrow (4pq) = 1$

$$\Rightarrow q^{2n}(0, 0) \sim \frac{1}{\sqrt{\pi n}}$$

$$\Rightarrow \sum_n q^{2n}(0, 0) \sim \sum_n \frac{1}{\sqrt{\pi n}} = \infty \Rightarrow \text{„0“ ist rekurrent.}$$

**Definition 6.4.4 (irreduzibel)** Die stochastische Matrix  $q$  heißt irreduzibel, falls für alle  $x, y \in E$  ein  $m \in \mathbb{N}$  existiert mit  $q^m(x, y) > 0$ .

**Satz 6.4.5** Sei  $q$  irreduzibel. Dann gilt: Existiert ein  $z \in E$ , das rekurrent ist, so sind alle  $x \in E$  rekurrent.

**Beweis:**

Sei  $x \in E$  beliebig. Dann existieren  $k, l \in \mathbb{N}$  mit  $q^k(x, z) > 0$  und  $q^l(z, x) > 0$ .

Dann ist

$$q^n(x, x) \geq q^k(x, z)q^m(z, z)q^l(z, x),$$

falls  $n = k + m + l$  ist. Es folgt

$$q^*(x, x) = \sum_{n \geq 0} q^n(x, x) = q^k(x, z) \sum_{m \geq 0} q^m(z, z)q^l(z, x).$$

Da die rechte Seite nach Voraussetzung gleich unendlich ist, ist auch  $q^*(x, x) = \infty$  und damit  $f^*(x, x) = 1$ .

## 6.5 Stationäre Verteilungen

**Definition 6.5.1** Eine Wahrscheinlichkeitsfunktion  $\pi$  auf  $E$  heißt stationär bezüglich  $q$ , falls für alle  $y \in E$  gilt:

$$\sum_x \pi(x)q(x, y) = \pi(y).$$

In Vektorschreibweise:  $\pi^T q = \pi^T$ , d. h.  $\pi$  ist linker Eigenvektor von  $q$  mit Eigenwert 1.

**Bemerkung:**

Ist  $\pi$  stationär und Startverteilung von  $\mathcal{X}$ , d.h.  $P(X_0 = x) = \pi(x)$ , so gilt:

$$P(X_n = y) = \pi(y).$$

**Beweis:** Es gilt

$$\begin{aligned} P(X_n = y) &= \sum_x P(X_n = y \mid X_0 = x)\pi(x) \\ &= \sum_x \pi(x)q^n(x, y) \\ &= \sum_x \pi(x) \sum_z q(x, z)q^{n-1}(z, y) \\ &= \sum_z \sum_x \pi(x)q(x, z)q^{n-1}(z, y) \\ &= \sum_z q^{n-1}(z, y)\pi(z) \\ &\quad \vdots \\ &= \pi(y). \end{aligned}$$

**Beispiel (Ehrenfestsches Urnenmodell)****Behauptung:**

$\pi(i) = \binom{N}{i} 2^{-N}$  ist stationäre Verteilung des Ehrenfestschen Urnenmodells.

**Beweis:** Zu zeigen:  $\sum_i \pi(i)q(i, j) = \pi(j)$ .

$$\begin{aligned}
& \sum_i \pi(i)q(i, j) \\
&= \sum_i \binom{N}{i} 2^{-N} q(i, j) \\
&= \binom{N}{j-1} 2^{-N} \frac{N-(j-1)}{N} + \binom{N}{j+1} 2^{-N} \frac{j+1}{N} \\
&= \left[ \frac{N!}{(j-1)!(N-(j-1))!} \frac{N-(j-1)}{N} + \frac{N!}{(j+1)!(N-(j+1))!} \frac{j+1}{N} \right] 2^{-N} \\
&= \left[ \frac{(N-1)!}{(j-1)!(N-j)!} \frac{j}{j} + \frac{(N-1)!}{j!(N-j-1)!} \frac{N-j}{N-j} \right] 2^{-N} \\
&= \frac{(N-1)!j + (N-1)!(N-j)}{j!(N-j)!} 2^{-N} \\
&= \binom{N}{j} 2^{-N} = \pi(j).
\end{aligned}$$

Es gibt noch einen einfacheren Weg die Stationarität von  $\pi$  zu zeigen. Man verwendet dazu das folgende Lemma.

**Lemma 6.5.2** *Gilt für alle  $x, y \in E$   $\pi(x)q(x, y) = \pi(y)q(y, x)$ , so ist  $\pi$  stationär.*

**Beweis:**

$$\sum_x \pi(x)q(x, y) = \sum_x \pi(y)q(y, x) = \pi(y) \sum_x q(y, x) = \pi(y).$$

**Bemerkung:**

Für das Ehrenfestsche Urnenmodell lautet die in 6.5.2 vorausgesetzte Gleichung  $\binom{N}{i} \frac{N-i}{N} = \binom{N}{i+1} \frac{i+1}{N}$ . Es ist offensichtlich, daß diese gilt.

Der folgende Satz liefert die Existenz einer stationären Verteilung.

**Satz 6.5.3** *Sei  $q$  irreduzibel. Sei  $T_z = \min\{n \geq 1 | X_n = z\}$ . Dann sind folgende Aussagen äquivalent:*

- a) *Es existiert ein  $z \in E$  mit  $E(T_z | X_0 = z) < \infty$ ,*
- b) *Für alle  $x, y \in E$  gilt  $E(T_y | X_0 = x) < \infty$ ,*
- c) *Es existiert eine stationäre Wahrscheinlichkeitsfunktion  $\pi$  bezüglich  $q$ .*

**Bemerkung:**

Bedingung (a) ist stärker als Rekurrenz, denn es gibt Markov-Ketten mit einem rekurrenten Zustand  $x$  für den  $E(T_x|X_0 = x) = \infty$  gilt. Siehe z.B. die Irrfahrt von oben.

**Beweisskizze:**

Zeige (a)  $\Rightarrow$  (c) Setze:  $P_z(\cdot) = P(\cdot|X_0 = z)$ . Dann ist  $P_z$  das Wahrscheinlichkeitsmaß bei Start in  $z$ .

Sei  $H(x) := \sum_{n=1}^{\infty} P_z(X_n = x, T_z \geq n)$ . Dann gilt  $H(z) = 1$  und weiter

$$\begin{aligned} \sum_{x \in E} H(x) &= \sum_{n=1}^{\infty} \sum_{x \in E} P_z(X_n = x, T_z \geq n) \\ &= \sum_{n=1}^{\infty} P_z(T_z \geq n) = E_z T_z. \end{aligned}$$

Wir behaupten (+):  $H(y) = \sum_{x \in E} H(x)q(x, y)$ .

Setzt man nun  $\pi(x) = \frac{H(x)}{E_z T_z}$ , so ist  $\pi(y) = \sum_x \pi(x)q(x, y)$  und  $\sum_y \pi(y) = 1$ , d.h.  $\pi$  ist stationäre Verteilung.

Zeige Behauptung (+):

$$\begin{aligned} H(y) &= \sum_{n=1}^{\infty} P_z(X_n = y, T_z \geq n) \\ &= P_z(X_1 = y) + \sum_{n \geq 2} P_z(X_n = y, X_j \neq z \text{ für } 1 \leq j \leq n-1) \\ &= q(z, y) + \sum_{n \geq 2} \sum_{x \in E \setminus \{z\}} P_z(X_{n-1} = x, X_n = y, X_j \neq z \text{ für } 1 \leq j \leq n-1) \\ &= q(z, y) + \sum_{x \in E \setminus \{z\}} \sum_{n \geq 2} P_z(X_{n-1} = x, X_n = y, X_j \neq z \text{ für } 1 \leq j \leq n-1) \\ &= q(z, y) + \sum_{x \in E \setminus \{z\}} \sum_{n \geq 2} P_z(X_{n-1} = x, X_j \neq z \text{ für } 1 \leq j \leq n-1) \cdot P(X_n = y | X_{n-1} = x) \\ &= q(z, y) + \sum_{x \in E \setminus \{z\}} \sum_{l \geq 1} P_z(X_l = x, X_j \neq z \text{ für } 1 \leq j \leq l) \cdot P(X_n = y | X_{n-1} = x) \\ &= q(z, y) + \sum_{x \in E \setminus \{z\}} H(x)q(x, y) \\ &= \sum_x H(x)q(x, y). \end{aligned}$$

**Korollar 6.5.4** *Es gilt*

$$\pi(x) = \frac{H(x)}{E_z T_z}$$

*und insbesondere*

$$\pi(z) = \frac{1}{E_z T_z},$$

*wobei*

$$H(x) = \sum_{n=1}^{\infty} P_z(X_n = x, T_z \geq n) \text{ ist für } z \in E.$$

**Bemerkung:**

Beim Ehrenfestschen Urnenmodell sei  $\sigma(i) = \min\{n > 0 \mid X_n = i\}$ . Dann ist nach dem Korollar:  $E_0\sigma(0) = 2^{2k}$  und  $E_k\sigma(k) = \binom{2k}{k}^{-1} 2^{2k}$ , wobei  $k = N/2$  ist. Beachte das  $E_k\sigma(k) \approx \sqrt{\pi k}$  wegen der Stirling-Formel ist. Setzt man  $N = 6 \cdot 10^{23}$  und ist ein Zeitschritt 1 Sekunde lang, so ist  $E_k\sigma(k) \cong 10^{12}$  Sekunden, aber  $E_0\sigma(0) \cong 10^{1,8 \cdot 10^{23}}$  Sekunden.

## 6.6 Konvergenz gegen die stationäre Verteilung

**Definition 6.6.1 (Periode, aperiodisch)** *Sei  $q$  eine irreduzible stochastische Matrix. Weiter sei  $N(x, y) := \{n \in \mathbb{N} \mid q^n(x, y) > 0\}$ . Wir definieren die Periode von  $x \in E$  durch  $d(x) := \text{ggT}(N(x, x))$ , wobei  $\text{ggT}(B) := \sup\{k \in \mathbb{N} \mid B \subset k \cdot \mathbb{N}\}$  der größte gemeinsame Teiler der Menge  $B$  ist.  $q$  heißt aperiodisch, falls  $d(x) = 1$  ist für alle  $x \in E$ .  $q$  hat Periode  $k$ , falls  $d(x) = k$  ist für alle  $x \in E$ .*

**Beispiel:**

Für die stochastische Matrix beim Ehrenfestschen Urnenmodell gilt  $d = 2$ .

**Satz 6.6.2** *Sei  $q$  irreduzibel und aperiodisch mit stationärer Verteilung  $\pi$ . Dann gilt für alle  $x \in E$*

$$\lim_{n \rightarrow \infty} \sum_{y \in E} |q^n(x, y) - \pi(x)| = 0.$$

*Sei  $q$  irreduzibel mit Periode  $k$  und mit stationärer Verteilung  $\pi$ , so gilt*

$$\lim_{n \rightarrow \infty} \sum_{y \in E} |q^{nk}(x, y) - \pi(x)| = 0.$$

**Beweis:**

Zum Beweis siehe Dümbgen, S. 129.

**Satz 6.6.3** *Sei  $q$  irreduzibel und aperiodisch mit stationärer Verteilung  $\pi$ . Sei  $f : E \rightarrow \mathbb{R}$ , so daß  $\sum_{x \in E} |f(x)|\pi(x) < \infty$  ist. Dann gilt für jede Startverteilung  $\mu$  mit  $P_\mu$ -Wahrscheinlichkeit 1 :*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N f(x_k) = \sum_{x \in E} f(x)\pi(x).$$

**Bemerkung:**

Dies ist ein grundlegender Satz für Markov-Ketten, der dem Gesetz der Großen Zahlen für unabhängige Beobachtungen entspricht. Es ist ein Basisresultat für die Simulation bei Markov-Ketten, der sogenannten MCMC-Methode.

**Bemerkungen zum Ehrenfestschen Urnenmodell**

Dieses Modell wurde entwickelt, um eine Streitfrage in der statistischen Physik zu klären. Boltzmann behauptete, ein großes Teilchensystem tendiere sehr schnell zu seinem Gleichgewicht, Zermelo entgegnete, auch dann müßten aus physikalischen Gründen sehr unwahrscheinliche Zustände angenommen werden können. Dies erschien paradox. Das Ehrenfestsche Modell gestattet die Details zu berechnen.

Sei  $N = 2k$ . Sei  $\tau(k) = \inf\{n \geq 0 \mid X_n = k\}$ . Dann gilt:

a)  $E_0(\tau(k)) = k \ln k + k + O(1)$

b)  $E_k(\tau(0)) = \frac{1}{2^k} 2^{2k} (1 + o(\frac{1}{k}))$  für  $k \rightarrow \infty$ .

Hier ist noch ein einfaches Beispiel zu Satz 6.6.2.

Sei  $q = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix}$ .

Dann ist die stationäre Verteilung  $\pi = (\frac{1}{2}, \frac{3}{8}, \frac{1}{8})$ . Für sie gilt  $\pi^T q = \pi^T$ .

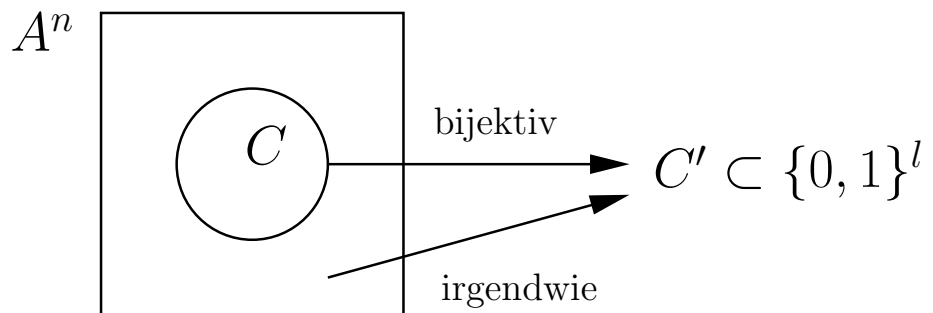
# Anhang A

## 6.7 Der Quellencodierungssatz von Shannon

Eine Nachrichtenquelle sendet zufällige Signale  $X_1, X_2, \dots$  aus einem endlichen Alphabet  $\mathbf{A}$ . Alle  $X_i$  sind identisch verteilt mit  $P(X_i = a) = \varrho(a), a \in \mathbf{A}$  mit  $\varrho(a) \geq 0$  und  $\sum \varrho(a) = 1$ . Wie kann man den Informationsgehalt der Quelle angeben? Wir betrachten den Nachrichtenblock  $\mathbf{X}_n = (X_1, \dots, X_n)$  und fragen, wie viele Ja/Nein Fragen braucht man, um  $\mathbf{X}_n$  bis auf eine Irrtumswahrscheinlichkeit  $\varepsilon > 0$  eindeutig ermitteln zu können? Dazu wähle man eine möglichst kleine Menge  $C \subset A^n$  mit  $P(\mathbf{X}_n \in C) \geq 1 - \varepsilon$  und bestimme das kleinste  $l$  mit  $|C| \leq 2^l$  ( $l$  natürlich  $\leq n$ ). Dann existiert eine Bijektion  $\varphi : C \rightarrow C' \subset \{0, 1\}^l$ .  $\varphi$  codiert jedes Wort  $w \in C$  in  $\varphi(w) \in C'$ , aus dem  $w$  durch  $\varphi^{-1}$  zurückgewonnen werden kann. Den Wörtern in  $A^n \setminus C$  weist man ein beliebiges Codewort in  $C'$  zu. Nach Konstruktion gilt dann

$$P(\mathbf{X}_n \neq \varphi^{-1} \circ \varphi(\mathbf{X}_n)) = P(\mathbf{X}_n \notin C) \leq \varepsilon.$$

Das heißt, das von der Quelle gesendete Wort  $\mathbf{X}_n$  kann mit einer Irrtumswahrscheinlichkeit  $\varepsilon$  aus dem Codewort  $\varphi(\mathbf{X}_n)$  ermittelt werden. Letzteres besteht aus höchstens  $l$  bits, kann also durch  $l$  Ja/Nein-Fragen identifiziert werden. Wir schreiben nun  $\log_2$  für den Logarithmus zur Basis 2.



### Satz 6.7.1 (Quellencodierungssatz, Shannon 1948)

Sei

$$L(n, \varepsilon) = \min\{l \geq 0 \mid \text{es ex. } C \subset A^n \text{ mit } |C| \leq 2^l, P(\mathbf{X}_n \in C) \geq 1 - \varepsilon\}$$

die kleinste Länge eines Binärcodes ... Dann gilt:

$$\lim_{n \rightarrow \infty} \frac{L(n, \varepsilon)}{n} = H(\varrho)$$

mit  $H(\varrho) = -\sum_{a \in A} \varrho(a) \log_2 \varrho(a)$   $H(\varrho)$  heißt die Entropie der Zähl-dichte  $\varrho$ .

**Beweis:**

Sei  $Y_i = -\log_2 \varrho(X_i), i \geq 1$ . Die  $Y_i$  sind unabhängig und nehmen nur endlich viele Werte an. Es ist  $EY_i = H(\varrho)$  und  $V(Y_i) < \infty$ . Nach dem Gesetz der großen Zahlen gilt für jedes  $\delta > 0$  und hinreichend großes  $n$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n Y_i - H(\varrho)\right| > \delta\right) < \delta.$$

Nun ist  $\frac{1}{n}\sum_{i=1}^n Y_i = -\frac{1}{n}\log_2 \varrho^{\otimes n}(\mathbf{X}_n)$  mit  $\varrho^{\otimes n}(w) = \varrho(w_1) \cdot \varrho(w_2) \dots \varrho(w_n), w_i \in A$ . Dies bedeutet, daß  $P(\mathbf{X}_n \in B_n) \geq 1 - \delta$  mit  $B_n = \{w \in \mathbf{A}^n \mid 2^{-n(H(\varrho)+\delta)} \leq \varrho^{\otimes n}(w) \leq 2^{n(H(\varrho)-\delta)}\}$ . Ist  $\delta \leq \varepsilon$  so ist  $B_n = C$  ein Kandidat bei der Infimumsbildung von  $L(n, \varepsilon)$ . Dann ist  $L(n, \varepsilon) \leq \min\{l \mid |B_n| \leq 2^l\}$ .

Da

$$\begin{aligned} 1 &= \sum_{w \in A^n} \varrho^{\otimes n}(w) \\ &\geq \sum_{w \in B_n} 2^{-n(H(\varrho)+\delta)} \\ &= |B_n| 2^{-n(H(\varrho)+\delta)} \end{aligned}$$

$$\Rightarrow |B_n| \leq 2^{n(H(\varrho)+\delta)}$$

$$\Rightarrow L(n, \varepsilon) \leq n(H(\varrho) + \delta) + 1.$$

Da  $\delta > 0$  beliebig ist, folgt

$$\varliminf_n \frac{L(n, \varepsilon)}{n} \leq H(\varrho).$$

Andererseits existiert zu  $l = L(n, \varepsilon)$  ein  $C \subset A^n$  mit  $|C| < 2^l$  und  $P(\mathbf{X}_n \in C) \geq 1 - \varepsilon$ . Da  $\varrho^{\otimes n}(w) 2^{n(H(\varrho)-\delta)} \leq 1$  ist für  $w \in B_n$ , gilt

$$\begin{aligned} 2^{L(n, \varepsilon)} &\geq |C| \geq |C \cap B_n| \geq \sum_{w \in C \cap B_n} \varrho^{\otimes n}(w) 2^{n(H(\varrho)-\delta)} \\ &= P(\mathbf{X}_n \in C \cap B_n) 2^{n(H(\varrho)-\delta)} \\ &\geq (1 - \varepsilon - \delta) 2^{n(H(\varrho)-\delta)} \end{aligned}$$

Es folgt

$$L(n, \varepsilon) \geq \log_2(1 - \varepsilon - \delta) + n(H(\varrho) - \delta)$$

und damit

$$\varliminf_n \frac{L(n, \varepsilon)}{n} \geq H(\varrho),$$

falls  $\delta < 1 - \varepsilon$  gewählt wird.

□

# Literaturverzeichnis

Dümbgen, L.: Stochastik für Informatiker, Springer, 2003

Kersting, G. und Wakolbinger, A.: Elementare Stochastik, Birkhäuser, 2008

Pitman, J.: Probability, Springer, 1993, Corr. 7th printing

Stoyan, D.: Stochastik für Ingenieure und Naturwissenschaftler, Akademie Verlag, 1993