

# Wahrscheinlichkeitsrechnung und Statistik

Prof. Dr. A. Raphaélian  
Fachbereich 1 - Ingenieurwissenschaften I  
Fachhochschule für Technik und Wirtschaft Berlin

# Inhaltsverzeichnis

<b>I</b>	<b>Wahrscheinlichkeitsrechnung</b>	<b>1</b>
<b>1</b>	<b>Zufallsexperimente und empirische Wahrscheinlichkeit</b>	<b>2</b>
1.1	Historische Einführung . . . . .	2
1.2	Deterministische und stochastische Modelle . . . . .	5
1.3	Ereignisraum . . . . .	6
1.3.1	Ereignis . . . . .	6
1.3.2	Verknüpfung von Ereignissen . . . . .	7
<b>2</b>	<b>Kombinatorik</b>	<b>11</b>
2.1	Permutation . . . . .	12
2.1.1	Permutation ohne Wiederholung . . . . .	12
2.1.2	Permutation mit Wiederholung . . . . .	13
2.2	Kombination . . . . .	14
2.2.1	Kombination ohne Wiederholung . . . . .	14
2.2.2	Kombination mit Wiederholung . . . . .	16
2.3	Variation . . . . .	18
2.3.1	Variation ohne Wiederholung . . . . .	18
2.3.2	Variation mit Wiederholung . . . . .	19
2.4	Zusammenfassung . . . . .	21
<b>3</b>	<b>Wahrscheinlichkeit</b>	<b>23</b>
3.1	Laplace-Experimente . . . . .	23
3.2	Wahrscheinlichkeitsaxiome . . . . .	25
3.2.1	Relative Häufigkeit . . . . .	25
3.2.2	Wahrscheinlichkeitsaxiome von Kolmogorov . . . . .	27
3.2.3	Statistische Definition der Wahrscheinlichkeit . . . . .	34
3.3	Bedingte Wahrscheinlichkeit . . . . .	36
3.3.1	Einführung und Beispiele . . . . .	36
3.3.2	Multiplikationssatz, Totale Wahrscheinlichkeit, Formel von Bayes . . . . .	40
3.3.3	Mehrstufige Zufallsexperimente . . . . .	48
3.3.4	Stochastisch unabhängige Ereignisse . . . . .	52
3.3.4.1	Stochastische Unabhängigkeit von mehr als zwei Ereignissen . . . . .	59

<b>4</b>	<b>Zufallsvariable</b>	<b>60</b>
4.1	Einführung und Definition . . . . .	60
4.1.1	Physikalische Interpretation der Dichtefunktion . . . . .	63
4.2	Wahrscheinlichkeitsverteilung einer Zufallsvariable . . . . .	64
4.2.1	Verteilung einer diskreten Zufallsvariable . . . . .	64
4.2.2	Verteilung einer stetigen Zufallsvariable . . . . .	76
<b>5</b>	<b>Kennwerte einer Wahrscheinlichkeitsverteilung</b>	<b>83</b>
5.1	Erwartungswert einer Zufallsvariable . . . . .	83
5.2	Funktionen von Zufallsvariablen . . . . .	88
5.3	Varianz und Standardabweichung . . . . .	92
5.4	Momente einer Verteilung . . . . .	98
5.5	Standardtransformation . . . . .	99
<b>6</b>	<b>Spezielle Wahrscheinlichkeitsverteilungen</b>	<b>102</b>
6.1	Diskrete Wahrscheinlichkeitsverteilungen . . . . .	102
6.1.1	Diskrete Gleichverteilung . . . . .	102
6.1.2	Multinomialverteilung . . . . .	104
6.1.2.1	Bernoulli-Experiment . . . . .	104
6.1.2.2	Binomialverteilung . . . . .	106
6.1.3	Hypergeometrische Verteilung . . . . .	112
6.1.4	Weitere diskrete Verteilungen . . . . .	117
6.1.4.1	Poisson-Verteilung . . . . .	117
6.1.4.2	Geometrische Verteilung . . . . .	120
6.2	Stetige Wahrscheinlichkeitsverteilungen . . . . .	121
6.2.1	Gleichverteilung . . . . .	121
6.2.2	Normalverteilung . . . . .	122
6.2.2.1	Standard-Normalverteilung . . . . .	124
6.2.2.2	Quantile der Standard-Normalverteilung . . . . .	130
6.2.2.3	Logarithmische Normalverteilung . . . . .	133
6.2.3	Approximation der Binomialverteilung durch die Normalverteilung . . . . .	135
6.2.4	Exponentialverteilung . . . . .	140
6.2.5	Weitere stetige Verteilungen . . . . .	141
6.2.5.1	Gamma-Verteilung . . . . .	141
6.2.5.2	Weibull-Verteilung . . . . .	141
<b>7</b>	<b>Mehrdimensionale Wahrscheinlichkeitsverteilungen</b>	<b>143</b>
7.1	Stichprobenvariable und Testverteilungen . . . . .	143
7.1.1	Chi-Quadrat-Verteilung . . . . .	144
7.1.2	$t$ -Verteilung von „Student“ . . . . .	145
<b>II</b>	<b>Statistik</b>	<b>147</b>
<b>8</b>	<b>Inhalt der Statistik</b>	<b>148</b>
8.1	Klassifikation der Statistik . . . . .	148

8.2	Statistik und Wahrscheinlichkeitsrechnung . . . . .	150
<b>9</b>	<b>Deskriptive Statistik</b>	<b>152</b>
9.1	Stichprobe und Grundgesamtheit . . . . .	152
9.2	Statistische Parameter . . . . .	156
9.2.1	Mittelwert . . . . .	156
9.2.2	Empirische Varianz und Standardabweichung . . . . .	158
<b>10</b>	<b>Induktive Statistik</b>	<b>160</b>
10.1	Statistische Schätzverfahren . . . . .	160
10.1.1	Problemstellung . . . . .	160
10.1.2	Schätz- und Stichprobenfunktionen . . . . .	161
10.1.3	Maximum-Likelihood-Methode . . . . .	168
10.1.3.1	Beschreibung des Verfahrens . . . . .	168
10.1.3.2	Anwendung des ML-Verfahrens auf spezielle Wahrscheinlichkeitsverteilungen . . . . .	170
10.1.4	Konfidenzbereiche . . . . .	173
10.1.4.1	Vertrauensintervalle und statistische Sicherheit .	173
10.1.4.2	Erwartungswert einer Normalverteilung bei bekannter Varianz . . . . .	176
10.1.4.3	Erwartungswert einer Normalverteilung bei unbekannter Varianz . . . . .	184
10.1.4.4	Varianz einer Normalverteilung . . . . .	186
10.1.5	Toleranzintervalle . . . . .	189
10.2	Statistische Prüfverfahren . . . . .	191
10.2.1	Statistische Hypothesen und statistische Tests . . . . .	191
10.2.1.1	Parameter-tests . . . . .	193
10.2.1.2	Aufbau eines Parameter-tests . . . . .	197
10.2.1.3	Beispiele von Parameter-tests . . . . .	201
10.2.2	Äquivalenz von statistischen Schätzverfahren und stati- stischen Prüfverfahren . . . . .	206
10.2.3	Fehler bei statistischen Prüfverfahren . . . . .	209
10.2.3.1	Fehler erster Art . . . . .	209
10.2.3.2	Fehler zweiter Art . . . . .	213
10.2.3.3	Trennschärfe (Power) eines Tests . . . . .	216
	<b>Stichwortverzeichnis</b>	<b>219</b>

**Teil I**

# **Wahrscheinlichkeitsrechnung**

# Kapitel 1

## Zufallsexperimente und empirische Wahrscheinlichkeit

### 1.1 Historische Einführung

Der Begriff „Wahrscheinlichkeit“ wird in der Umgangssprache auf zweierlei Arten verwendet:

- (i) Beschreibung der subjektiven Überzeugung einer Person von der Richtigkeit eines bestimmten Sachverhaltes. Beispiel:

Mit einer Chance von 80 % wird Hertha BSC im Jahr 20.. Deutscher Fußballmeister.

Hier handelt es sich um Mutmaßungen über einen nicht wiederholbaren Vorgang, von dem im Prinzip in der Zukunft feststellbar ist, ob er eintritt oder nicht. Eine zusätzliche Zahlenangabe in Prozent ist dann ein Maß für die Stärke der jeweiligen persönlichen Überzeugung; diese wird als *subjektive Wahrscheinlichkeit* bezeichnet.

- (ii) Beschreibung von beobachteten Häufigkeiten bei (mindestens im Prinzip) beliebig oft wiederholbaren Vorgängen, deren Einzelausgänge nicht vorhersehbar sind. Beispiel:

Dieser Würfel ist falsch, denn die Chance, mit ihm eine Sechs zu würfeln, beträgt nur 1/10.

Man hat in einer langen (wie lang?) Serie von Würfeln mit diesem Würfel in etwa 1/10 der Fälle eine Sechs erhalten und erwartet nun, daß der Würfel in Zukunft das gleiche Verhalten zeigen wird.

Empirische Sachverhalte wie in diesem Beispiel heißen *zufällige Experimente*. Die bei der Wiederholung solcher zufälligen Experimente (unter gleichen äußeren Bedingungen) beobachteten Häufigkeiten dienen später dazu, den Begriff *empirische Wahrscheinlichkeit* zu definieren.

(iii) **Aufgabe**

In welche Kategorie fällt das Beispiel

Heute wird es mit einer Wahrscheinlichkeit von 20 % regnen.

Was ist außerdem damit gemeint? 20 % der Menge eines „normalen“ Regengusses; 20 % der Dauer eines „normalen“ Regengusses? Oder noch etwas Drittes?

Wie entstand der Begriff der Wahrscheinlichkeit als ein Maß für das potentielle Eintreten eines Ereignisses?

Im siebzehnten Jahrhundert waren Glücksspiele sehr populär, besonders in Frankreich. Um seine Chancen beim Spiel zu verbessern, konsultierte Antoine Gombaud, der Chevalier de Méré, ein passionierter Glücksspieler mit Interesse an der Mathematik, den berühmten Mathematiker und Theologen Blaise Pascal. Unter den Problemen, über welche er von Pascal Auskunft begehrte, waren die folgenden beiden:

## (1) (Teilungsproblem)

Zwei Spieler (A und B) spielen ein Spiel, in dem jeder die gleiche Gewinnchance hat. Nach jeder Runde erhält der Gewinner einen Punkt, und wer zuerst 5 Punkte erzielt, hat das Spiel gewonnen.

Nun wird das Spiel beim Stand von

A : 4 Punkte  
B : 3 Punkte

vorzeitig abgebrochen. Wie ist der Einsatz auf A und B zu verteilen?

- (i) Vorschlag 1: (Blick in die Vergangenheit)  
Teilung des Einsatzes im Verhältnis 4:3, da bei Abbruch des Spieles A schon 4 Punkte erzielt hat, B dagegen nur 3.
- (ii) Vorschlag 2: (Blick in die Zukunft)  
Teilung des Einsatzes im Verhältnis 2:1, da A in zwei von drei (fiktiven weiteren) Fällen gewinnen würde.
- (iii) Vorschlag 3: (Pascals Vorschlag, Blick in die Zukunft)  
Teilung des Einsatzes im Verhältnis 3:1 mit der im folgenden angeführten Begründung.  
In jedem Fall wäre das Spiel nach zwei weiteren Runden beendet gewesen mit den in Tabelle 1.1 angegebenen vier Möglichkeiten, und Spieler A gewinnt in drei von vier möglichen Fällen.

Ergebnis		Gewinner
A	(A)	A
A	(B)	A
B	A	A
B	B	B

Tabelle 1.1: Gewinnmöglichkeiten beim Teilungsproblem

Dieser (ebenfalls auf die Zukunft gerichtete) Lösungsvorschlag diskreditiert auch Vorschlag 2: hier muß man zuerst die Anzahl der für das Spiel *möglichen* Fälle (4) auflisten, danach die Anzahl der für den Ausgang des Ereignisses *günstigen* Fälle (3) bestimmen.

(2) (Würfelproblem)

De Méré lernte, wahrscheinlich aus Erfahrung, daß man einen regulären Würfel viermal werfen mußte, damit die Chance, eine 6 zu würfeln, größer als 50 % ist. Er schloß daraus, daß es vorteilhaft ist darauf zu wetten, daß für zwei Würfel die Chance, eine Doppel-Sechs zu würfeln, größer als 50 % ist, wenn man  $6 \cdot 4 = 24$  mal würfeln darf, denn für den zweiten Würfel ist die Chance, eine 6 zu würfeln, ja gerade  $1/6$ .

Er wurde bitter enttäuscht und verlor auf lange Sicht die Wette und viel Geld. Daraufhin beklagte er sich bei Pascal, der ihn darüber aufklärte, daß diese Resultate zu erwarten waren:

- (i) Die Chance, mit einem Würfel bei vier Würfeln eine Sechs zu würfeln, ist gegeben durch

$$P(\text{Einmal 6 mit 4 Würfeln}) := 1 - \left(\frac{5}{6}\right)^4 = 0,518 = 51,8\%.$$

- (ii) Die Chance, mit zwei Würfeln bei 24 Würfeln eine Doppel-Sechs zu würfeln, ist gegeben durch

$$P(\text{Zweimal Sechs mit 24 Würfeln}) := 1 - \left(\frac{35}{36}\right)^{24} = 0,491 = 49,1\%.$$

Motiviert durch Fragen dieser Art, begann Pascal eine umfängliche Korrespondenz mit dem französischen Mathematiker Pierre de Fermat. Diese Korrespondenz diente als Grundlage einer einheitlichen Theorie von Zufallsphänomenen, welche heute unter dem Namen *Wahrscheinlichkeitstheorie* bekannt ist.

(3) (Ziegenproblem)<sup>1</sup>

In einer Spielshow im Fernsehen sei zum Schluß noch ein Kandidat übrig, der eine von drei geschlossenen Türen auswählen soll. Hinter einer Tür

---

<sup>1</sup>Gero von Randow: Das Ziegenproblem – Denken in Wahrscheinlichkeiten, rowohlt science 1994, Sachbuch 9337



wartet als Preis ein Auto, hinter den beiden anderen stehen Ziegen. Der Kandidat zeigt auf eine Tür, sagen wir Nummer eins. Sie bleibt vorerst geschlossen. Der Moderator, der weiß, hinter welcher Tür sich das Auto befindet, öffnet eine andere Tür, z.B. Nummer drei, und eine Ziege schaut ins Publikum.

Der Kandidat erhält nun die Möglichkeit, bei der bisher gewählten Tür eins zu bleiben oder zur Tür zwei zu wechseln. Wie soll er sich entscheiden? Anders formuliert: Gibt es jetzt aus stochastischer Sicht berechnete Gründe, bei der gewählten Tür zu bleiben bzw. die Tür zu wechseln, oder sind die Wahrscheinlichkeiten, den Preis hinter Tür eins bzw. Tür zwei zu finden, nach wie vor identisch?

- (4) (Ziegenproblem; Veränderung des subjektiven Wertesystems)<sup>2</sup>

Sie sind Ziegenhirte auf einer autofreien Insel und wollen das auch bleiben. Das Auto interessiert Sie nicht, Sie wollen eine Ziege haben.

Sollten Sie nach Ihrer ersten Wahl die Tür wechseln oder nicht?

## 1.2 Deterministische und stochastische Modelle

Was versteht man unter einem „Experiment“? Worin besteht die Rolle der Wahrscheinlichkeitstheorie, die Ergebnisse eines Experimentes zu interpretieren?

### Beispiel 1.1

- (1) In einem Versuch wird der Siedepunkt des Wassers bestimmt, man erhält einen Wert von 100 °C. Jede Abweichung von diesem Wert ist ein Resultat wechselnder experimenteller Bedingungen, wie Änderung des Luftdrucks, Ablesefehler bei dem Thermometer, etc.

Bleiben die Versuchsbedingungen dieselben, dann wird auch das Ergebnis das gleiche sein. Es handelt sich hier um ein *deterministisches Experiment*.

- (2) Beim Werfen eines Würfels handelt es sich in der Regel (s.u.) um ein Experiment, bei welchem der Ausgang nicht vorhersehbar ist. Hierbei handelt es sich um ein *stochastisches Experiment* oder *Zufallsexperiment*.

### Bemerkung

Es hängt natürlich von den Versuchsbedingungen ab, ob ein Experiment als Zufallsexperiment angesehen werden kann oder nicht. Sind z.B. beim Werfen eines Würfels alle für die Bewegung des Würfels wichtigen Daten bekannt, so kann

---

<sup>2</sup>Dubben/Bornholdt: Mit an Wahrscheinlichkeit grenzender Sicherheit - Logisches Denken und Zufall, rowohlt science 3. Auflage 2006

man das Ergebnis eines jeden Wurfes als Lösung eines Anfangswertproblems einer Differentialgleichung vorausberechnen, und dann handelt es sich um ein deterministisches Experiment.

Andererseits ist es häufig gar nicht wünschenswert, ein zufälliges Experiment durch Festlegen aller Versuchsbedingungen in ein deterministisches Experiment umzuwandeln. Bei der Anwendung der Wahrscheinlichkeitstheorie auf praktische Probleme kann daher die Wahl eines geeigneten mathematischen Modells durchaus mit Schwierigkeiten verbunden sein, und man sollte ein Problem einer genauen prämathematischen Analyse unterziehen, bevor man es mit einem Modell zu lösen versucht.

Im neunzehnten Jahrhundert wurde die Wahrscheinlichkeitstheorie von Laplace weiterentwickelt und angewendet auf Probleme der Fehlerrechnung, statistischen Mechanik und Versicherungsmathematik. Die aus dem Griechischen abgeleitete Bezeichnung *Stochastik*<sup>3</sup> ist ein Sammelbegriff für die Disziplinen Wahrscheinlichkeitsrechnung und Mathematische Statistik.

## 1.3 Ereignisraum

### 1.3.1 Ereignis

#### Definition 1.1

Ein *Zufallsexperiment* ist ein Experiment, bei dem die folgenden Voraussetzungen erfüllt sind:

- (i) Das Experiment läßt sich unter den gleichen (wohldefinierten) äußeren Bedingungen beliebig oft wiederholen.
- (ii) Bei der Durchführung des Experiments sind mehrere sich gegenseitig ausschließende Ergebnisse möglich.
- (iii) Das Ergebnis einer konkreten Durchführung des Experiments läßt sich dabei nicht mit Sicherheit voraussagen, sondern es ist zufallsbedingt.

#### Beispiel 1.2

Mehrmaliges Würfeln mit zwei unterschiedlichen homogenen Würfeln und Notieren der dabei jeweils erzielten Augenpaare. Ein mögliches Ergebnis ist dann beispielsweise das Augenpaar  $\langle 1, 6 \rangle$ .

Anhand dieses Beispiels werden einige Begriffe und Bezeichnungen zusammengestellt:

---

<sup>3</sup>Die früheste Erwähnung des Wortes „Stochastik“ findet man in Platons Werk *Philebos*. Sokrates spricht dort von „... der Fähigkeit des geschickten Vermutens (*στοχαστική*), die sich nur durch stete Handhabung und mühevollen Arbeit herausbilden läßt.“

Ereignis	Beschreibung des Ereignisses
$\mathcal{A} = \{2\}$	Würfeln einer 2 (Elementarereignis)
$\mathcal{B} = \{2, 4, 6\}$	Würfeln einer geraden Zahl
$\mathcal{C} = \{3, 6\}$	Würfeln eines Vielfachen von 3
$\mathcal{D} = \{5, 6\}$	Würfeln einer 5 <i>oder</i> einer 6

Tabelle 1.2: Verschiedene Ereignisse beim Würfeln mit einem homogenen Würfel

**Definition 1.2**

- (i) Es sind 36 verschiedene Ergebnisse möglich, nämlich die  $6^2$  möglichen Ergebnisse

$$\langle 1, 1 \rangle, \langle 1, 2 \rangle, \dots, \langle 5, 6 \rangle, \langle 6, 6 \rangle.$$

Diese Ergebnisse heißen *Elementarereignisse*.

- (ii) Die Menge

$$\Omega := \{\omega : \omega \text{ ist Elementarereignis}\}$$

heißt *Ergebnisraum*, *Ergebnismenge* oder *Merkmalraum*.

- (iii) Jede Teilmenge des Merkmalraums heißt ein *Ereignis*. Beispielsweise ist

$$\mathcal{A} := \{\langle 1, 3 \rangle, \langle 2, 2 \rangle, \langle 3, 1 \rangle\}$$

das Ereignis: Die Augensumme beider Würfel ist gleich vier.

- (iv) Die Menge aller Ereignisse eines Zufallsexperimentes heißt *Ereignisraum*. Der Ereignisraum ist demnach die Potenzmenge der Ergebnismenge  $\Omega$ :

$$\text{Ereignisraum} = \mathcal{P}(\Omega) = \{\mathcal{A} : \mathcal{A} \subseteq \Omega\}.$$

**Beispiel 1.3**

Beim Zufallsexperiment „Wurf eines homogenen Würfels“ ist (Tabelle 1.2)

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

**1.3.2 Verknüpfung von Ereignissen**

Ereignisse sind Teilmengen der Ergebnismenge, also des Merkmalraumes  $\Omega$ . Damit sind sie Elemente der Potenzmenge  $\mathcal{P}(\Omega)$  und lassen sich gemäß den Verknüpfungsregeln für Mengen zu neuen Ereignissen zusammensetzen.

Die wichtigsten Mengenoperationen sind die folgenden. Seien  $\mathcal{M}$  und  $\mathcal{N}$  Mengen, dann sind definiert:

$\mathcal{M} \cup \mathcal{N} := \{x : x \in \mathcal{M} \vee x \in \mathcal{N}\}$	Vereinigung
$\mathcal{M} \cap \mathcal{N} := \{x : x \in \mathcal{M} \wedge x \in \mathcal{N}\}$	Durchschnitt
$\mathcal{M} \setminus \mathcal{N} := \{x : x \in \mathcal{M} \wedge x \notin \mathcal{N}\}$	Differenz
$\overline{\mathcal{M}} := \complement_{\Omega} \mathcal{M} := \Omega \setminus \mathcal{M}$	Komplement von $\mathcal{M}$ bzgl. $\Omega$
$\mathcal{P}(\mathcal{M}) := \{x : x \subseteq \mathcal{M}\}$	Potenzmenge

(1.1)

Durchschnitt  $\mathcal{M} \cap \mathcal{N}$  :

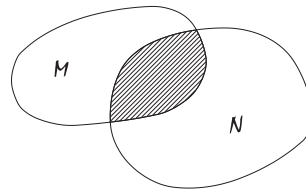


Abbildung 1.1: Durchschnitt

Vereinigung  $\mathcal{M} \cup \mathcal{N}$  :

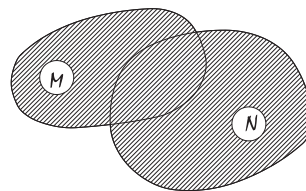


Abbildung 1.2: Vereinigung

Differenzmenge  $\mathcal{M} \setminus \mathcal{N}$  :

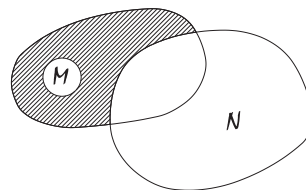


Abbildung 1.3: Differenzmenge

**Bemerkung**

- (i) Wenn klar ist, welche „Obermenge“  $\Omega$  gemeint ist, schreibt man auch  $\complement \mathcal{M}$  anstelle von  $\complement_{\Omega} \mathcal{M}$ .
- (ii) Durch die Anwendung der Mengenlehre sind Ereignisse auf Mengen zurückgeführt, sie sind also Mengen. Da hier jedoch der wahrscheinlichkeitstheoretische Aspekt im Vordergrund steht, werden fortan Ereignisse

nicht mehr kalligraphisch –  $\mathcal{A}, \mathcal{B}$  – geschrieben, sondern in normaler Schrift –  $A, B$ .

Weiterhin wird in der Wahrscheinlichkeitstheorie die Bezeichnung  $\bar{A}$  der Bezeichnung  $\complement_{\Omega} A$  vorgezogen.

(iii) Manchmal sind auch die Bezeichnungen

$$\begin{array}{llll} A + B & \text{für} & A \cup B & \text{und} \\ A \cdot B & \text{für} & A \cap B & \end{array}$$

üblich, ich verwende diese hier jedoch nicht.

(iv) Gilt für zwei Mengen  $\mathcal{A}$  und  $\mathcal{B}$  die Aussage

$$\mathcal{A} \cap \mathcal{B} = \emptyset,$$

so heißen diese beiden Mengen bekanntlich *disjunkt* oder *elementfremd*. Die dadurch definierten Ereignisse  $A$  und  $B$  schließen sich dann gegenseitig aus, sie heißen *unvereinbar* oder *konträr*. In Anlehnung an die Verwendung der Mengenlehre zur Beschreibung von Ereignissen nennt man sich gegenseitig ausschließende Ereignisse daher manchmal auch *disjunkt*.

**Satz 1.1** (Regeln von de Morgan)

Für zwei beliebige Ereignisse (Teilmengen)  $A$  und  $B$  von  $\Omega$  gelten:

$$\begin{array}{l} \overline{A \cup B} = \bar{A} \cap \bar{B} \\ \overline{A \cap B} = \bar{A} \cup \bar{B} \end{array}$$

In Worten:

Das Komplement der Vereinigung ist gleich dem Durchschnitt der Komplemente.

Das Komplement des Durchschnittes ist gleich der Vereinigung der Komplemente.

Beweisen lassen sich diese Regeln mit Hilfe von Inklusionsbeziehungen für Mengen.

**Beispiel 1.4**

Eine kreisrunde Zielscheibe von 50 cm Radius ist in Kreisringe mit den äußeren Radien  $r_i = i \cdot 5$  cm eingeteilt ( $i \in \{1, 2, \dots, 10\}$ ), s. Abb. 1.4.

Mit welcher Wahrscheinlichkeit trifft man in diese Bereiche, wenn vorausgesetzt wird, daß die Zielscheibe auf jeden Fall getroffen wird und die gesuchten Wahrscheinlichkeiten proportional zu den jeweiligen Flächeninhalten sind? Steht dieses Wahrscheinlichkeitsmaß der Gleichverteilung in bezug auf die Flächen der Kreisringe mit der Erfahrung in Einklang?

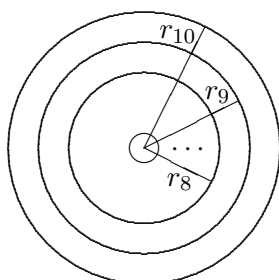


Abbildung 1.4: Zielscheibe mit konzentrischen Kreisen

**Lösung**

Die Größe des  $i$ -ten Kreisringes beträgt ( $i \in \{1, 2, \dots, 10\}$ ,  $r_0 := 0$ )

$$K_i := \pi \cdot (r_i^2 - r_{i-1}^2) = 25\pi(i^2 - (i-1)^2) = 25\pi(2i-1),$$

und die Größe der gesamten Zielscheibe beträgt

$$F_{10} := \pi \cdot r_{10}^2 = \sum_{i=1}^{10} K_i = 25\pi \sum_{i=1}^{10} (2i-1) = 2500\pi.$$

Die Wahrscheinlichkeit, genau den  $i$ -ten Kreisring zu treffen, beträgt nach Voraussetzung (Abb. 1.5)

$$P(K_i) = \frac{25\pi(2i-1)}{2500\pi} = \frac{2i-1}{100} \quad (i \in \{1, \dots, 10\}),$$

und die Erfahrung spricht gegen ein solches Wahrscheinlichkeitsmaß.

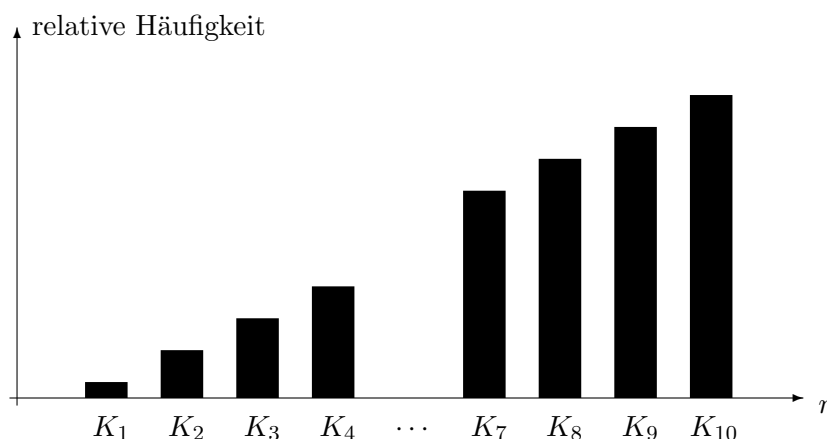


Abbildung 1.5: Häufigkeitsdiagramm für Treffer auf eine runde Zielscheibe

**Bemerkung**

Wie man sieht, läßt sich ein Begriff von „Wahrscheinlichkeit“ vielfältig festlegen, wenn auch nicht immer realitätsnah. Es ist nun Aufgabe der Mathematik dafür zu sorgen, daß diese Festlegung wenigstens konsistent geschieht.

## Kapitel 2

# Kombinatorik

Unter dem Begriff *Kombinatorik* versteht man die Theorie vom systematischen Abzählen von Elementen. Im einzelnen werden behandelt:

(i) **Permutation**

Anordnung von  $n$  Elementen

(ii) **Kombination**

Auswahl von  $k$  aus  $n$  Elementen *ohne* Berücksichtigung der Reihenfolge

(a) ohne Wiederholung: Ziehung ohne Zurücklegen

(b) mit Wiederholung: Ziehung mit Zurücklegen

(iii) **Variation**

Auswahl von  $k$  aus  $n$  Elementen *mit* Berücksichtigung der Reihenfolge

(a) ohne Wiederholung: Ziehung ohne Zurücklegen

(b) mit Wiederholung: Ziehung mit Zurücklegen

Als Modell, um all diese Begriffe zu klären, diene das sog. *Urnenmodell*. In einer Urne befinden sich  $n$  ( $n \in \mathbb{N}$  geeignet) verschiedene gleichgroße Kugeln, die sich in ihrer Farbe voneinander unterscheiden.

(i) Die Anzahl der verschiedenen Arten, diese  $n$  Kugeln anzuordnen, führt zum Begriff *Permutation*.

(ii) Aus der Urne werden nacheinander  $k$  ( $k \in \mathbb{N}$  geeignet) Kugeln gezogen.

(a) Ziehung ohne Zurücklegen

Die jeweils gezogene Kugel wird nicht in die Urne zurückgelegt und scheidet somit für alle weiteren Ziehungen aus. Jede der  $n$  Kugeln kann somit höchstens einmal gezogen werden.

(b) Ziehung mit Zurücklegen

Jede Kugel darf mehrmals verwendet werden, d.h. vor einer nachfolgenden Ziehung wird die gezogene Kugel in die Urne zurückgelegt und kann somit abermals gezogen werden.

In beiden Fällen wird ferner unterschieden, ob die Reihenfolge der Ziehungen berücksichtigt werden soll oder nicht. Wenn ja, liegt eine *Variation* vor, wenn nein, eine *Kombination*.

In der Statistik wird eine solche zufällige Entnahme von  $k$  Kugeln aus einer Gesamtheit von  $n$  Kugeln als *Stichprobe* bezeichnet. Die Stichprobe heißt *geordnet*, wenn die Reihenfolge der Stichprobenelemente (hier: die gezogenen  $k$  Kugeln) berücksichtigt wird. Spielt die Reihenfolge jedoch keine Rolle, so liegt eine *ungeordnete* Stichprobe vor.

### Aufgabe

Welcher Gruppe gehört das Lottospiel „6 aus 49“ an?

## 2.1 Permutation

### Beispiel 2.1

Drei verschiedenfarbige Kugeln lassen sich auf genau sechs verschiedene Arten anordnen, s. Abb. 2.1.

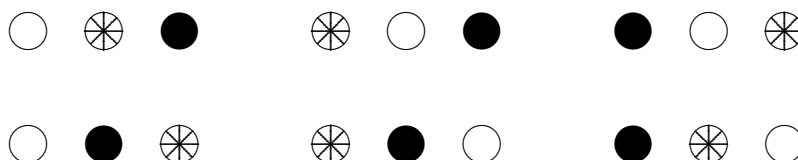


Abbildung 2.1: Anordnungen von drei verschiedenfarbigen Kugeln

### 2.1.1 Permutation ohne Wiederholung

#### Definition 2.1

Eine Anordnung von  $n$  verschiedenen Elementen in einer bestimmten Reihenfolge heißt eine *Permutation* dieser Elemente.

#### Problem

Wie groß ist die Anzahl der Permutationen von  $n$  Elementen?



**Lösung**

Es sind  $n$  Positionen vorhanden, die man sich von 1 bis  $n$  durchnummeriert denkt. Dazu gibt es  $n$  verschiedene Kugeln, mit denen diese Positionen besetzt werden sollen.

Position Nr. 1 läßt sich mit jeder Kugel belegen,  
 Position Nr. 2 läßt sich (dann noch) mit  $n - 1$  Kugeln belegen,  
 Position Nr. 3 läßt sich (dann noch) mit  $n - 2$  Kugeln belegen,  
 $\vdots$   
 Position Nr.  $n$  läßt sich nur noch mit 1 Kugel belegen.

Abb. 2.2 verdeutlicht diese Aussage.

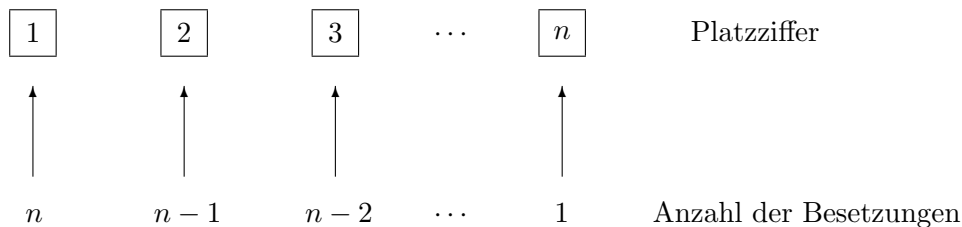


Abbildung 2.2: Besetzungsmöglichkeiten von  $n$  Positionen

Die Anzahl der Permutationen von  $n$  verschiedenen Elementen beträgt daher

$$P(n) := n! := n \cdot (n-1) \cdot \dots \cdot 1.$$

Der Ausdruck  $n!$  heißt  $n$  *Fakultät* und ist eine Abkürzung für das Produkt der natürlichen Zahlen von 1 bis  $n$  einschließlich:

$$n! := \prod_{i=1}^n i = 1 \cdot 2 \cdot \dots \cdot (n-1) \cdot n.$$

Zusätzlich definiert man noch

$$0! := 1.$$

**2.1.2 Permutation mit Wiederholung**

Befinden sich unter den  $n$  Kugeln  $n_1$  gleichfarbige (welche also nicht unterscheidbar sind), so fallen alle jene Anordnungen zusammen, welche durch Vertauschen der gleichfarbigen Kugeln untereinander hervorgehen: diese Anordnungen können nun nicht unterschieden werden und müssen daher als gleich gelten.

Bei  $n_1$  gleichfarbigen Kugeln gibt es  $P(n_1) = n_1!$  verschiedene Möglichkeiten, diese  $n_1$  gleichen Kugeln untereinander zu vertauschen. Daher gibt es genau

$$P(n; n_1) := \frac{P(n)}{P(n_1)} = \frac{n!}{n_1!}$$

verschiedene Anordnungsmöglichkeiten für  $n$  Kugeln, unter denen sich  $n_1$  gleiche befinden. Daher ist die Anzahl der Permutationen von  $n$  Elementen, unter denen sich jeweils  $n_1, n_2, \dots, n_k$  gleiche befinden, gegeben durch

$$P(n; n_1, n_2, \dots, n_k) := \frac{n!}{n_1! n_2! \dots n_k!}$$

mit  $n_1 + n_2 + \dots + n_k = n$ .

### Beispiele 2.2

- (i) Auf einem Regal sollen 6 verschiedene Gegenstände angeordnet werden. Es gibt dann

$$P(6) = 6! = 720$$

verschiedene Anordnungsmöglichkeiten (Permutationen) dafür.

- (ii) Vor einem Wohnhaus gibt es einen Parkhafen mit 6 Stellplätzen. Wieviele Möglichkeiten gibt es, 3 Autos dort zu parken?

Wenn drei von sechs Stellplätzen mit Autos besetzt sind, so lautet die Frage, wieviele Möglichkeiten es gibt, die drei Autos und die drei leeren Plätze zu permutieren. Dabei werden die Autos als unterschiedlich angesehen, die verbleibenden leeren Stellplätze dagegen nicht. Es gibt dann

$$P(6; 3) = \frac{6!}{3!} = \frac{720}{6} = 120$$

verschiedene Möglichkeiten, 3 Autos auf 6 Stellplätzen zu parken.

- (iii) Wieviele Möglichkeiten der Anordnung von  $2n$  Karten gibt es beim Memory-Spiel?

## 2.2 Kombination

Unter einer *Kombination* versteht man die Auswahl von  $k$  aus  $n$  Elementen *ohne* Berücksichtigung der Reihenfolge, in welcher die Elemente ausgewählt werden. Eine *ungeordnete Stichprobe* von  $k$  Kugeln (allgemein: Elementen) heißt eine *Kombination  $k$ -ter Ordnung*.

I.a. ist die natürliche Zahl  $k$  dabei kleiner als  $n$ . Betrachtet man hingegen eine sog. Kombination mit Wiederholung, d.h. legt man ein einmal ausgewähltes Element nach seiner Auswahl wieder zurück, so kann es im nächsten Zug erneut ausgewählt werden; in diesem Fall kann  $k$  auch größer als  $n$  sein.

### 2.2.1 Kombination ohne Wiederholung

Einer Urne mit  $n$  verschiedenen Kugeln  $K_1, K_2, \dots, K_n$  werden nacheinander  $k$  Kugeln *ohne* Zurücklegen entnommen. Die Reihenfolge der gezogenen Kugeln soll dabei ohne Bedeutung sein. Beim Lotto „6 aus 49“ beispielsweise wird

nach dieser Methode verfahren: Aus einer Gesamtheit von 49 Kugeln werden nacheinander 6 Kugeln ohne Zurücklegen gezogen.

Eine solche ungeordnete Stichprobe von  $k$  Elementen heißt eine *Kombination  $k$ -ter Ordnung ohne Wiederholung*.

Auf wieviele verschiedene Arten ist es möglich,  $k$  Kugeln aus einer Urne mit  $n$  verschiedenen Kugeln ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge zu ziehen, d.h. wie groß ist die Anzahl der Kombinationen  $k$ -ter Ordnung ohne Wiederholung bei  $n$  verschiedenen Elementen?

Zur Beantwortung dieser Frage denkt man sich die  $n$  verschiedenen Kugeln  $K_1, K_2, \dots, K_n$  in einer Reihe angeordnet. Dann kennzeichnet man die gezogenen  $k$  Kugeln mit einer 1, die übrigen  $n - k$  Kugeln mit einer 0 (Abb. 2.3).

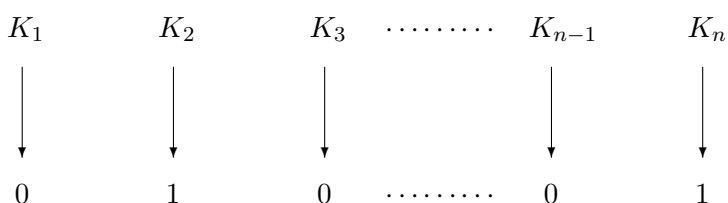


Abbildung 2.3: Kennzeichnung von  $k$  gezogenen Kugeln

In dieser Anordnung tritt somit die Zahl 1 genau  $k$ -mal und die Zahl 0 genau  $(n - k)$ -mal auf. Damit lautet das zu lösende Problem:

Wieviele verschiedene Anordnungsmöglichkeiten gibt es für die  $n$  Zahlen (Elemente), unter denen sich eine Gruppe von  $k$  gleichen und eine andere Gruppe von  $(n - k)$  gleichen Elementen befinden?

Offensichtlich handelt es sich dabei um eine Permutation von  $n$  Elementen mit Wiederholung, unter denen sich sowohl  $k$  gleiche als auch  $n - k$  gleiche befinden. Damit gibt es nach dem letzten Abschnitt genau

$$K(n; k) := P(n; k, n - k) = \frac{n!}{k! (n - k)!} = \frac{n \cdot (n - 1) \cdot \dots \cdot (n - [k + 1])}{k!}$$

verschiedene Möglichkeiten.

Für den letzten Ausdruck ist eine Abkürzung gebräuchlich. Man definiert

$$\binom{n}{k} := \frac{n!}{k! (n - k)!} \quad \left[ = \binom{n}{n - k} \right]$$

und nennt  $\binom{n}{k}$  einen *Binomialkoeffizienten* (lies: „ $n$  über  $k$ “).

### Beispiel 2.3

Sei  $\Omega$  eine Menge mit  $n$  Elementen. Dann hat die Potenzmenge  $\mathcal{P}(\Omega)$  genau  $2^n$  Elemente.

*Beweis*

$\Omega$  hat  $n$  Elemente, und  $\mathcal{P}(\Omega)$  ist die Menge aller Teilmengen von  $\Omega$ . In ihr gibt es jeweils

$$\binom{n}{k} = \frac{n!}{k! (n-k)!} \quad (k \in \{0, \dots, n\})$$

Teilmengen mit je  $k$  Elementen (Auswahl von  $k$  aus  $n$ ), insgesamt also

$$\sum_{k=0}^n \binom{n}{k} = \sum_{k=0}^n \binom{n}{k} \cdot 1^k \cdot 1^{n-k} = (1+1)^n = 2^n$$

Teilmengen. Hierbei wurde der *Binomische Satz* angewendet, und die im Binomischen Satz auftretenden Koeffizienten heißen daher *Binomialkoeffizienten*.

### 2.2.2 Kombination mit Wiederholung

Darf jede der  $n$  verschiedenen Kugeln bei der Ziehung mehrmals verwendet werden, d.h. wird eine gezogene Kugel vor der nächsten Ziehung in die Urne zurückgelegt und hat so die Chance, mehrmals gezogen zu werden, so erhält man eine *Kombination  $k$ -ter Ordnung mit Wiederholung*. Ihre Anzahl ist

$$K_W(n; k) := \binom{k+n-1}{k}.$$

Man beachte, daß  $k$  jetzt auch größer als  $n$  sein kann.

Der Beweis dieser Formel ist etwas knifflig und soll daher anhand eines Beispiels illustriert werden.

#### Beispiel 2.4

Die vier Möglichkeiten, 3 (nicht unterscheidbare) Äpfel auf 2 Kinder zu verteilen, sind in Abbildung 2.4 aufgeführt.

Kind 1	Kind 2
0 0 0	–
0 0	0
0	0 0
–	0 0 0

Abbildung 2.4: Verteilung von 3 Äpfeln auf 2 Kinder

0 0 0 1
0 0 1 0
0 1 0 0
1 0 0 0

Abbildung 2.5: Verteilung von  $k = 3$  Nullen und  $n - 1 = 1$  Einsen

Durch welche Überlegung kommt man auf dieses Ergebnis? Offenbar sind  $4 = 3 + 2 - 1$  Positionen zu besetzen, und zwar mit einer 0 für einen Apfel und mit einer 1 für eine (hier: die einzige) Trennwand zwischen den Kindern. Die Tabelle könnte codiert also wie Abbildung 2.5 aussehen.

Wenn nun  $k$  (nicht unterscheidbare) Äpfel auf  $n$  Kinder zu verteilen sind, so müssen  $k$  Nullen – die Äpfel – und  $n - 1$  Einsen – die Trennwände – berücksichtigt werden. Es liegt demnach eine Permutation mit Wiederholung von  $k + n - 1$  Elementen vor, unter denen  $n_1 = k$  und  $n_2 = n - 1$  gleich sind. Insgesamt erhält man also

$$\begin{aligned} P(k + n - 1; n_1 = k, n_2 = n - 1) &= \frac{(k + n - 1)!}{k! (n - 1)!} \\ &= \frac{(k + n - 1)!}{k! ([k + n - 1] - k)!} \\ &= \binom{k + n - 1}{k} \\ &= K_W(n; k) \end{aligned}$$

Möglichkeiten.

### Beispiele 2.5

- (i) Einer Warenlieferung von 10 Transistoren soll zu Kontrollzwecken eine Stichprobe von 2 Transistoren entnommen werden. Wieviele verschiedene Stichproben sind dabei möglich?

#### Lösung

Da es bei dieser (ungeordneten) Stichprobe auf die Reihenfolge der ausgewählten Transistoren nicht ankommt und die Ziehung (wie in der Praxis allgemein üblich) ohne Zurücklegen erfolgt, gibt es genau

$$K(10; 2) = \binom{10}{2} = \frac{1 \cdot 2 \cdot \dots \cdot 10}{(1 \cdot 2) \cdot (1 \cdot 2 \cdot \dots \cdot 8)} = \frac{9 \cdot 10}{1 \cdot 2} = 45$$

*verschiedene* Stichproben (Kombination 2. Ordnung von 10 Elementen ohne Wiederholung).

- (ii) Für eine Parallelschaltung von drei Widerständen (Abb. 2.6) stehen insgesamt 5 verschiedene Klassen von Ohm'schen Widerständen zur Verfügung. Wieviele verschiedene Schaltungsmöglichkeiten gibt es, wenn jeder Widerstand einer Klasse

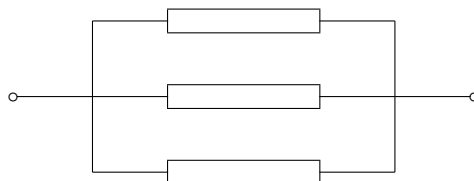


Abbildung 2.6: Parallelschaltung von drei Ohm'schen Widerständen

- (a) höchstens einmal  
 (b) mehrmals, d.h. bis zu dreimal

verwendet werden darf?

### Lösung

- (a) Da jeder Widerstand nur einmal verwendet werden darf, liegt hier eine Kombination 3. Ordnung ohne Wiederholung vor. In diesem Fall gibt es daher

$$K(5; 3) = \binom{5}{3} = \frac{5 \cdot 4 \cdot 3}{1 \cdot 2 \cdot 3} = 10$$

*verschiedene* Möglichkeiten für eine Parallelschaltung aus drei Widerständen.

- (b) Jeder Widerstand kann bis zu dreimal verwendet werden. Dieses Mal handelt es sich um eine Kombination 3. Ordnung mit Wiederholung. Es gibt demnach

$$K_W(5; 3) = \binom{5+3-1}{3} = \binom{7}{3} = \frac{7 \cdot 6 \cdot 5}{1 \cdot 2 \cdot 3} = 35$$

*verschiedene* Möglichkeiten für eine Parallelschaltung aus drei Widerständen.

## 2.3 Variation

Im Gegensatz zu einer Kombination berücksichtigt eine *Variation* nun auch die Reihenfolge, in der die Auswahl der  $k$  Elemente aus den  $n$  Elementen erfolgt. Eine *geordnete Stichprobe* von  $k$  Kugeln (allgemein: Elementen) heißt eine *Variation  $k$ -ter Ordnung*.

### 2.3.1 Variation ohne Wiederholung

Einer Urne mit  $n$  verschiedenen Kugeln werden nacheinander  $k$  Kugeln *ohne* Zurücklegen entnommen. Wie in einer Kombination ohne Wiederholung ist jede der verschiedenen  $n$  Kugeln in der Ziehung damit höchstens einmal vertreten. Dieses Mal wird nun – im Gegensatz zu einer Kombination – allerdings auch die Reihenfolge der Entnahme der Kugeln berücksichtigt.

Eine solche *geordnete Stichprobe* von  $k$  Kugeln heißt eine *Variation  $k$ -ter Ordnung ohne Wiederholung*.

Auf wieviele verschiedene Arten ist es möglich,  $k$  Kugeln aus einer Urne mit  $n$  verschiedenen Kugeln ohne Zurücklegen, aber unter Berücksichtigung der Reihenfolge zu ziehen, d.h. wie groß ist die Anzahl der Variationen  $k$ -ter Ordnung ohne Wiederholung bei  $n$  verschiedenen Elementen?

Zur Lösung dieses Problems geht man zunächst von einer beliebigen Kombination  $k$ -ter Ordnung ohne Wiederholung aus. Diese enthält genau  $k$  verschiedene Kugeln  $K_1, K_2, \dots, K_k$  in einer beliebigen Anordnung. Da sich diese Kugeln auf  $k!$  verschiedene Arten permutieren lassen, erhält man aus dieser Kombination genau  $k!$  Variationen. Dies gilt nun aber für jede Kombination  $k$ -ter Ordnung. Somit ist

$$V(n; k) := P(k) \cdot K(n; k) = k! \binom{n}{k} = k! \frac{n!}{k! (n-k)!} = \frac{n!}{(n-k)!}$$

die Anzahl der möglichen Variationen  $k$ -ter Ordnung ohne Wiederholung.

### 2.3.2 Variation mit Wiederholung

Darf dagegen jede der insgesamt  $n$  verschiedenen Kugeln in der Urne mehrmals verwendet werden, so erhält man eine *Variation  $k$ -ter Ordnung mit Wiederholung*. Jede gezogene Kugel muß vor der nächsten Ziehung in die Urne zurückgelegt werden (Ziehung mit Zurücklegen).

Die Anzahl der Variationen  $k$ -ter Ordnung beträgt in diesem Fall

$$V_W(n; k) := n^k.$$

Denn jede der  $k$  Positionen in einer solchen Anordnung kann mit jeder der verschiedenen Kugeln belegt werden, da Mehrfachziehung jetzt zulässig ist, s. Abb. 2.7.

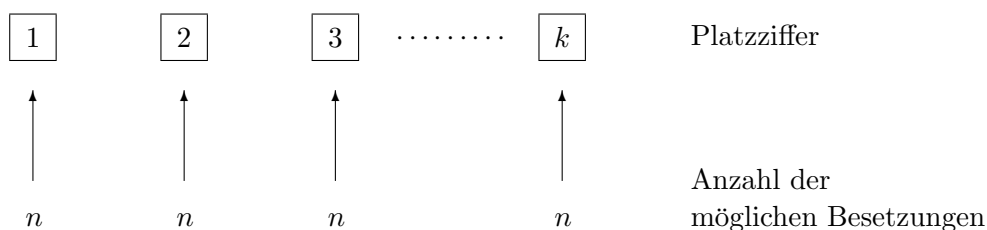


Abbildung 2.7: Besetzungsmöglichkeiten für die  $k$  vorhandenen Positionen

Somit gibt es genau

$$\underbrace{n \cdot n \cdot \dots \cdot n}_{k\text{-mal}} = n^k$$

verschiedene Anordnungen.

#### Bemerkung

- (i) Variationen unterscheiden sich von Kombinationen grundsätzlich dadurch, daß die Reihenfolge oder Anordnung der Elemente berücksichtigt wird. Sie können daher auch als geordnete Stichproben aufgefaßt werden, welche einer sog. Grundgesamtheit von  $n$  Elementen entnommen werden.

- (ii) Variationen  $k$ -ter Ordnung heißen manchmal auch *Kombinationen  $k$ -ter Ordnung unter Berücksichtigung der Anordnung*.

### Beispiele 2.6

- (i) Beim Pferdetoto muß in der sog. *Dreierwette* der Zieleinlauf der ersten drei Pferde in der richtigen Reihenfolge vorausgesagt werden. Wieviele verschiedene Dreier-Wetten sind möglich, wenn zehn Pferde starten?

#### Lösung

Drei der zehn Pferde erreichen als erste das Ziel in einer bestimmten Reihenfolge. Es handelt sich daher um Variationen 3. Ordnung von 10 Elementen ohne Wiederholung. Es gibt daher

$$V(10; 3) = \frac{10!}{(10-3)!} = \frac{10!}{7!} = 8 \cdot 9 \cdot 10 = 720$$

verschiedene Dreier-Wetten.

- (ii) Wieviele verschiedene Augenpaare sind bei einem Wurf zweier unterschiedlich gekennzeichnete homogener Würfel möglich?

#### Lösung

Als Augenzahlen eines jeden der beiden Würfel kommen die Zahlen 1, 2, 3, 4, 5 und 6 in Frage. Bei den Augenpaaren  $\langle i, j \rangle$  handelt es sich um geordnete Zahlenpaare, da die Würfel unterschiedlich gekennzeichnet sind (z.B. ein weißer und ein schwarzer Würfel), und somit um Variationen 2. Ordnung der 6 Zahlen 1 bis 6 mit Wiederholung. Daher gibt es

$$V_W(6; 2) = 6^2 = 36$$

verschiedene Augenpaare, nämlich

$$\begin{aligned} &\langle 1, 1 \rangle, \langle 1, 2 \rangle, \langle 1, 3 \rangle, \langle 1, 4 \rangle, \langle 1, 5 \rangle, \langle 1, 6 \rangle, \\ &\langle 2, 1 \rangle, \langle 2, 2 \rangle, \langle 2, 3 \rangle, \langle 2, 4 \rangle, \langle 2, 5 \rangle, \langle 2, 6 \rangle, \\ &\langle 3, 1 \rangle, \langle 3, 2 \rangle, \langle 3, 3 \rangle, \langle 3, 4 \rangle, \langle 3, 5 \rangle, \langle 3, 6 \rangle, \\ &\langle 4, 1 \rangle, \langle 4, 2 \rangle, \langle 4, 3 \rangle, \langle 4, 4 \rangle, \langle 4, 5 \rangle, \langle 4, 6 \rangle, \\ &\langle 5, 1 \rangle, \langle 5, 2 \rangle, \langle 5, 3 \rangle, \langle 5, 4 \rangle, \langle 5, 5 \rangle, \langle 5, 6 \rangle, \\ &\langle 6, 1 \rangle, \langle 6, 2 \rangle, \langle 6, 3 \rangle, \langle 6, 4 \rangle, \langle 6, 5 \rangle, \langle 6, 6 \rangle. \end{aligned}$$

- (iii) Aus den Ziffern 1 bis 9 sollen dreistellige Zahlen gebildet werden. Wieviele verschiedene Möglichkeiten gibt es dafür, wenn in jeder dreistelligen Zahl keine Ziffer mehrfach auftreten darf?



**Lösung**

Neun Ziffern werden auf drei Plätze verteilt, wobei jede Ziffer nur einmal auftreten darf und die Anordnung der Ziffern von Bedeutung ist. Es handelt sich also um Variationen 3. Ordnung ohne Wiederholung. Ihre Anzahl beträgt

$$V(9; 3) = \frac{9!}{(9-3)!} = \frac{9!}{6!} = 9 \cdot 8 \cdot 7 = 504.$$

**Beispiel 2.7**

Wie groß ist die Chance, im Lotto „6 aus 49“

- (i) 6 richtige Zahlen
- (ii) 3 richtige Zahlen
- (iii) überhaupt keine richtige Zahl

zu tippen?

**Lösung**

Beim Lotto „6 aus 49“ handelt es sich um eine sechsmalige (ohne Berücksichtigung der Zusatzzahl) Ziehung einer Kugel ohne Zurücklegen, bei welcher es auf die Reihenfolge der gezogenen Kugeln nicht ankommt. Somit liegt eine Kombination (Reihenfolge ohne Bedeutung) ohne Wiederholung (die jeweils gezogene Kugel wird vor dem nächsten Zug nicht zurückgelegt) vor.

$$(i) P_1 = \frac{1}{\binom{49}{6}} = \frac{6! \cdot 43!}{49!} = 7.15 \cdot 10^{-8} = 0.00000715 \% \approx \frac{1}{14 \text{ Millionen}}$$

(ii) Die Wahrscheinlichkeit unter (i) ist zu multiplizieren mit der Anzahl

- (a) 3 (richtige) aus 6 (richtigen) zu wählen und
- (b) (die restlichen) 3 (falschen Kugeln) aus den 43 (falschen Kugeln) zu wählen.

Somit ergibt sich hier

$$P_2 = \frac{\binom{6}{3} \cdot \binom{43}{3}}{\binom{49}{6}} = \frac{6!}{3! 3!} \cdot \frac{43!}{3! 40!} \cdot \frac{6! \cdot 43!}{49!} = 0.01765 = 1.765 \%$$

$$(iii) P_0 = 1 - \sum_{i=1}^6 P_i$$

**2.4 Zusammenfassung**

Ungeordnete Stichproben sind Kombinationen, geordnete Stichproben dagegen Variationen. Die Tabellen 2.1 und 2.2 fassen die wichtigsten Formeln übersichtsartig zusammen.

	ohne Wiederholung	mit Wiederholung
Permutation	$P(n) = n!$	$P(n; n_1, n_2, \dots, n_k) = \frac{n!}{n_1! n_2! \dots n_k!}$
Kombination	$K(n; k) = \binom{n}{k}$	$K_W(n; k) = \binom{k+n-1}{k}$
Variation	$V(n; k) = \frac{n!}{(n-k)!}$	$V_W(n; k) = n^k$

Tabelle 2.1: Übersicht über kombinatorische Begriffe und Formeln

	ohne Wiederholung	mit Wiederholung	
Kombination $k$ -ter Ordnung	$K(n; k) = \binom{n}{k}$	$K_W(n; k) = \binom{k+n-1}{k}$	ungeordnete Stichprobe
Variation $k$ -ter Ordnung	$V(n; k) = \frac{n!}{(n-k)!}$	$V_W(n; k) = n^k$	geordnete Stichprobe
	Ziehung ohne Zurücklegen	Ziehung mit Zurücklegen	

Tabelle 2.2: Kombinationen und Variationen

*Die Theorie der Wahrscheinlichkeit ist ein System,  
das uns beim Raten hilft.*

*Richard P. Feynman (1918 - 1988)*

---

## Kapitel 3

# Wahrscheinlichkeit

### 3.1 Laplace-Experimente

#### Definition 3.1 .

Seien  $m \in \mathbb{N}$  und  $\Omega := \{\omega_1, \dots, \omega_m\}$ , also  $\Omega$  eine endliche Menge.  $\Omega$  bestehe aus den Ergebnissen eines Zufallsexperimentes.

- (i) Man nennt das Experiment ein *Laplace-Experiment*, wenn aufgrund der Versuchsbedingungen alle Elementarereignisse  $\omega_i$  die gleiche Chance haben, als Ergebnis des Experimentes aufzutreten. Von keinem Ergebnis kann man also erwarten, daß es bevorzugt oder benachteiligt auftritt.

In diesem Fall ordnet man jedem Elementarereignis  $\omega_i \in \Omega$  die Zahl  $\frac{1}{m}$  zu:

$$P(\{\omega_i\}) := \frac{1}{m} \quad (i \in \{1, \dots, m\})$$

und interpretiert diese Zahl als *Wahrscheinlichkeit*<sup>1</sup>, mit welcher das Elementarereignis  $\omega_i$  bei einem weiteren Versuch als Ergebnis auftritt.

- (ii) Die Wahrscheinlichkeit  $P(A)$  eines Ereignisses  $A \subseteq \Omega$  wird dann wie folgt definiert:

$$P(A) := \frac{\text{Anzahl der } \omega_i \in \Omega \text{ mit } \omega_i \in A}{\text{Anzahl der } \omega_i \in \Omega}.$$

In der Literatur werden die Elementarereignisse  $\omega_i \in \Omega$  auch als *mögliche Ereignisse* und diejenigen aus  $A$  als *günstige Ereignisse* bezeichnet:

$$\begin{aligned} \omega_i \text{ ist ein mögliches Ereignis} & \quad :\iff \quad \omega_i \in \Omega \\ \omega_i \text{ ist ein günstiges Ereignis} & \quad :\iff \quad \omega_i \in A \end{aligned}$$

---

<sup>1</sup>engl.: probability

Aus diesem Grunde bedeutet die obige Definition:

$$\begin{aligned} P(A) &:= \frac{\text{Anzahl der für das Eintreten von } A \text{ günstigen Ereignisse}}{\text{Anzahl der (überhaupt) möglichen Ereignisse}} \\ &= \frac{k}{m} \quad \text{für} \quad A = \bigcup_{j=1}^k \{\omega_{i_j}\} \end{aligned}$$

### Bemerkung

Bei dieser Definition handelt es sich um die sog. *klassische Definition* der Wahrscheinlichkeit. Diese gilt nur für Laplace-Experimente, also für Experimente mit den folgenden Eigenschaften:

- Die Ergebnismenge oder der Merkmalraum  $\Omega$  ist endlich.
- Alle Elementarereignisse besitzen aufgrund der Versuchsbedingungen die gleiche Chance, als Ergebnis des Versuches aufzutreten, d.h. kein Elementarereignis tritt als Ergebnis bevorzugt oder benachteiligt auf.

### Beispiel 3.1

In einer Warenlieferung von 100 Transistoren befinden sich 10 defekte Transistoren. Zu Kontrollzwecken wird der Lieferung wahllos ein Transistor entnommen. Die Wahrscheinlichkeit für das Zufallsereignis

$$A \quad :\iff \quad \text{Ziehung eines defekten Transistors}$$

ist dann gegeben durch

$$P(A) = \frac{d(A)}{m} = \frac{10}{100} = 0.1.$$

Jeder der 100 Transistoren hat die gleiche Chance (Wahrscheinlichkeit), gezogen zu werden, und es gibt  $d(A) := 10$  günstige Fälle<sup>2</sup> für das Eintreten des Ereignisses  $A$ . Damit liegt ein Laplace-Experiment vor.

Das zum Ereignis  $A$  komplementäre Ereignis

$$\bar{A} \quad :\iff \quad \text{Ziehung eines einwandfreien Transistors}$$

hat damit die Wahrscheinlichkeit

$$P(\bar{A}) = \frac{d(\bar{A})}{m} = \frac{90}{100} = 0.9 = 1 - P(A).$$

<sup>2</sup>Die Begriffe „günstig“ und „ungünstig“ sind strikt auf das Experiment und das jeweilige Ereignis  $A$  zu beziehen. Das Auftreten eines defekten Transistors ist in diesem Sinne also „günstig“.

## 3.2 Wahrscheinlichkeitsaxiome

Der klassische Wahrscheinlichkeitsbegriff aus Definition 3.1 läßt sich nur auf Laplace-Experimente anwenden. Mehr noch, er scheint nicht frei von logischen Kreisschlüssen:

- (i) Die Definition stützt sich auf die Voraussetzung von *gleichmöglichen* – also gleichwahrscheinlichen – Fällen. Damit enthält diese Definition des Wahrscheinlichkeitsbegriffes aber den zu erklärenden Begriff selbst.

L. Vietoris (geb. 1891) konnte dieses Argument jedoch entkräften. Er argumentierte dahingehend, daß die Feststellung der Gleichwahrscheinlichkeit nicht schon aufgrund des zu definierenden Wahrscheinlichkeitsbegriffes geschieht sondern aufgrund geometrischer oder physikalischer Symmetrien (Würfel, Urne).

- (ii) Wie verhält man sich aber, wenn die auftretenden Elementarereignisse nicht gleichwahrscheinlich sind?

### Beispiele

- Wurf mit einem nicht homogenen Würfel
- Werfen eines Reißnagels auf eine ebene harte Platte mit zwei möglichen Ausgängen

In der modernen Wahrscheinlichkeitstheorie leitet man den Begriff der Wahrscheinlichkeit daher nicht aus anderen bekannten Größen ab, sondern *definiert* ihn als ein Maß auf der Menge der Ereignisse (dem Ereignisraum), welches die intuitive Vorstellung von der relativen Häufigkeit eines Ereignisses erfüllt und gewissen Gesetzen (Axiomen) genügt.

### 3.2.1 Relative Häufigkeit

Wird ein Zufallsexperiment  $n$ -mal durchgeführt und tritt das Ereignis  $A$  genau  $n_A$ -mal ein, so definiert man die

$$\begin{aligned} \text{absolute Häufigkeit} &:= h_n^{(a)}(A) := n_A, \\ \text{relative Häufigkeit} &:= h_n^{(r)}(A) := \frac{h_n^{(a)}(A)}{n}. \end{aligned}$$

Die im folgenden Abschnitt eingeführten Wahrscheinlichkeitsaxiome sollen durch die Eigenschaften von relativen Häufigkeiten von Ereignissen motiviert werden. Das Ganze geschieht anhand eines Beispiels.

#### Beispiel 3.2

In einer Urne befinden sich 6 Kugeln, und zwar

eine weiße Kugel,  
zwei graue Kugeln und  
drei schwarze Kugeln.

Das Experiment bestehe im zufälligen Entnehmen einer Kugel. Die Ergebnismenge dazu ist dann

$$\Omega := \{ W, G, S \}.$$

Nun wird die  $n$ -malige Ausführung des Zufallsexperimentes „Ziehung einer Kugel“ betrachtet, d.h. ein  $n$ -maliges Ziehen *mit* Zurücklegen.

### Aufgabe

Welches ist die Ergebnismenge dazu und wieviele Elemente hat sie? (Kombination mit Wiederholung)

Für die entsprechenden absoluten Häufigkeiten gilt

$$n_W + n_G + n_S = n.$$

(1) Für die relativen Häufigkeiten

$$h_n^{(r)}(W) = \frac{n_W}{n}, \quad h_n^{(r)}(G) = \frac{n_G}{n}, \quad h_n^{(r)}(S) = \frac{n_S}{n}$$

erhält man daher

$$0 \leq h_n^{(r)}(W), h_n^{(r)}(G), h_n^{(r)}(S) \leq 1. \quad (3.1)$$

(2) Das sichere Ereignis – bei jedem Zug wird eine weiße *oder* eine graue *oder* eine schwarze Kugel gezogen – tritt immer ein. Daher gilt

$$h_n(\Omega) = h_n^{(r)}(W) + h_n^{(r)}(G) + h_n^{(r)}(S) = \frac{n_W + n_G + n_S}{n} = 1. \quad (3.2)$$

(3) Nun betrachte man das Ereignis „Ziehen einer weißen *oder* einer grauen Kugel“, also das Ereignis  $W \cup G$ . Die beiden (Elementar-) ereignisse  $W$  und  $G$  schließen sich gegenseitig aus, treten also nicht gleichzeitig ein.

Bei  $n$  Ausführungen des Experimentes erhält man somit in genau  $n_W + n_G$  Fällen eine weiße oder eine graue Kugel. Daher erhält man für die entsprechenden relativen Häufigkeiten

$$h_n^{(r)}(W \cup G) = \frac{n_W + n_G}{n} = h_n^{(r)}(W) + h_n^{(r)}(G). \quad (3.3)$$

(4) Erfahrungsregel:

Mit zunehmender Anzahl  $n$  der Versuche „stabilisiert“ sich im allgemeinen die relative Häufigkeit  $h_n^{(r)}(A)$  eines zufälligen Ereignisses  $A$  um einen bestimmten Wert, d.h. bei umfangreichen Versuchsreihen ist sie „nahezu“ konstant.

Es liegt daher nahe, den Begriff der Wahrscheinlichkeit  $P(A)$  eines Ereignisses  $A$  als Grenzwert der relativen Häufigkeiten  $h_n^{(r)}(A)$  für  $n \rightarrow \infty$  wie folgt zu definieren:

$$P(A) := \lim_{n \rightarrow \infty} h_n^{(r)}(A).$$

In der Tat ist dies auch versucht worden, führt jedoch zu unüberbrückbaren Schwierigkeiten. Es läßt sich nicht beweisen, daß

- der entsprechende Grenzwert für *jede* Serie von Versuchen existiert und
- der Grenzwert für verschiedene Serien von „gleichen“ Versuchen stets derselbe ist.

Nach diesen Vorbereitungen sollte der nächste Abschnitt hinreichend motiviert sein.

### 3.2.2 Wahrscheinlichkeitsaxiome von Kolmogorov

**Definition 3.2** (Wahrscheinlichkeitsaxiome von Kolmogorov)

Ein *diskreter Wahrscheinlichkeitsraum* ist ein Paar  $\langle \Omega, P \rangle$ , wobei gilt:

- (i)  $\Omega$  ist eine nicht-leere abzählbare Menge.
- (ii)  $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$  ist eine Funktion mit den folgenden Eigenschaften:

$$(1) P(A) \geq 0 \quad (A \subseteq \Omega)$$

$$(2) P(\Omega) = 1$$

- (3) Für jede abzählbare Familie<sup>3</sup>  $A_n$  von unvereinbaren Ereignissen ist

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

die Summe kann dabei endlich oder unendlich (unendliche Reihe) sein.

---

<sup>3</sup>In der Mathematik ist eine Familie eine „Menge“, die im Unterschied zu einer gewöhnlichen Menge *Wiederholungen* enthalten kann; d.h. sie kann auch Elemente enthalten, die ursprünglich – d.h. in der gewöhnlichen Menge, der sie entstammen – identisch sind, hier aber durch Indizierung mehrfach als verschiedene Exemplare des ursprünglichen Elementes aufgeführt werden können. So ist beispielsweise eine Zahlenfolge eine Familie von Zahlen.

Die in der letzten Definition aufgeführten Gesetze nennt man auch *Wahrscheinlichkeitsaxiome von Kolmogorov*. Eine Abbildung  $P$  mit diesen Eigenschaften heißt ein *Wahrscheinlichkeitsmaß* oder ein *Verteilungsgesetz*.

**Beispiel 3.3** (Laplace-Experiment)

Eine Münze wird zweimal geworfen. Der Merkmalraum (Ergebnismenge) ist dann ( $W$ : Wappen,  $Z$ : Zahl)

$$\Omega := \{W, Z\} \times \{W, Z\} = \{\langle W, W \rangle, \langle W, Z \rangle, \langle Z, W \rangle, \langle Z, Z \rangle\}.$$

Wenn es sich um eine reguläre Münze handelt, findet das Auftreten keines der Elementarereignisse bevorzugt statt, d.h. alle Elementarereignisse sind gleichwahrscheinlich; das ist jedoch eine empirische Annahme! Nun ist

$$\Omega = \{\langle W, W \rangle\} \cup \{\langle W, Z \rangle\} \cup \{\langle Z, W \rangle\} \cup \{\langle Z, Z \rangle\}$$

die disjunkte Vereinigung aller aus Elementarereignissen gebildeten Mengen. Aus

$$1 = P(\Omega) = \sum_{i=1}^4 P(\langle W, Z \rangle_i)$$

folgt wegen der Gleichwahrscheinlichkeit

$$P(\langle W, W \rangle) = P(\langle W, Z \rangle) = P(\langle Z, W \rangle) = P(\langle Z, Z \rangle) = \frac{1}{4}.$$

Für spezielle Ereignisse gilt dann beispielsweise:

$$\begin{aligned} (A : \iff \text{wenigstens einmal Wappen}) & \implies P(A) = \frac{3}{4} \\ (B : \iff \text{beide Male dasselbe Ereignis}) & \implies P(B) = \frac{1}{2} \end{aligned}$$

**Beispiel 3.4** (kein Laplace-Experiment; Auftrittschancen nicht gleich)

Für einen homogenen Würfel werde ein Verteilungsgesetz wie folgt definiert:

$$\begin{aligned} P(1) & := \frac{1}{2} \\ P(2) & := \frac{1}{2} \\ P(k) & := 0 \quad (k \in \{3, 4, 5, 6\}) \end{aligned}$$

Dann liegt im Sinne von Definition 3.2 ein Wahrscheinlichkeitsmaß vor. Ob dieses Maß der Wirklichkeit entspricht oder nicht, kann nicht von der Mathematik entschieden werden.



**Beispiel 3.5** (kein Laplace-Experiment; Auftrittschancen nicht gleich)

Ein Würfel sei so beschaffen, daß das Auftreten einer geraden Zahl doppelt so häufig erfolge wie das Auftreten einer ungeraden Zahl.

Wie groß ist die Wahrscheinlichkeit, daß bei einem einzigen Wurf eine Zahl kleiner als 4 erscheint?

**Lösung**

Die Ergebnismenge ist hier  $\Omega := \{1, 2, 3, 4, 5, 6\}$ . Jede ungerade Zahl habe die Auftrittswahrscheinlichkeit  $w$ , dann hat jede gerade Zahl die Auftrittswahrscheinlichkeit  $2w$ . Da die Summe aller Wahrscheinlichkeiten 1 sein muß (Definition 3.2 (ii) (2)), folgt

$$9w = 1 \quad \longleftrightarrow \quad w = \frac{1}{9}.$$

Gesucht ist nun die Wahrscheinlichkeit des Auftretens des Ereignisses  $A := \{1, 2, 3\}$ .  $A$  besteht aus disjunkten Elementarereignissen, damit folgt aus Definition 3.2 (ii) (3):

$$\begin{aligned} P(A) &= P(\{1\}) + P(\{2\}) + P(\{3\}) \\ &= \frac{1}{9} + \frac{2}{9} + \frac{1}{9} \\ &= \frac{4}{9}. \end{aligned}$$

**Beispiel 3.6** (kein Laplace-Experiment; nicht-endliche Ergebnismenge  $\Omega$ )

Eine kreisförmige Zielscheibe mit 1 m Radius sei in Segmente  $\mathcal{S}_n$  gemäß Abb. 3.1 eingeteilt. Wie groß ist die Wahrscheinlichkeit, daß ein bestimmtes Segment der Zielscheibe getroffen wird, wenn vorausgesetzt wird, daß die Scheibe in jedem Fall getroffen wird und die gesuchte Wahrscheinlichkeit proportional zur Größe des jeweiligen Segments ist?

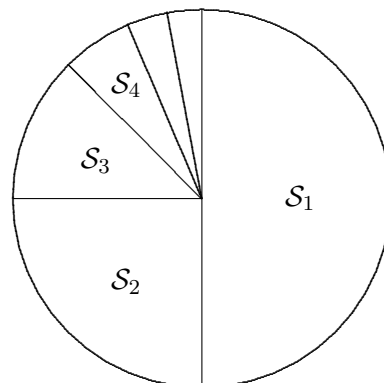


Abbildung 3.1: In Segmenten angeordnete kreisförmige Zielscheibe

**Lösung**

Die Ergebnismenge  $\Omega$  besteht aus allen solchen Elementen aus  $\{0, 1\}^{\mathbb{N}}$  (Folgen), welche an genau einer Stelle eine 1 aufweisen, ansonsten Nullen. Sei die Menge  $\mathcal{S}$  definiert durch  $\mathcal{S} := \bigcup_{n \in \mathbb{N}} S_n$ . Dann ist das Wahrscheinlichkeitsmaß gegeben durch

$$P(S_n) := \frac{1}{2^n} \quad (n \in \mathbb{N}).$$

Bei  $P$  handelt es sich tatsächlich um ein Wahrscheinlichkeitsmaß:

$$(1) \quad P(S_n) = \frac{1}{2^n} \geq 0 \quad (n \in \mathbb{N})$$

$$(2) \quad P(\mathcal{S}) = \sum_{n=1}^{\infty} \frac{1}{2^n} = \sum_{n=0}^{\infty} \frac{1}{2^n} - 1 = \frac{1}{1 - \frac{1}{2}} - 1 = 1$$

$$(3) \quad P\left(\bigcup_{n \in \mathcal{I}} S_n\right) = \sum_{n \in \mathcal{I}} P(S_n) \quad (\mathcal{I} \subseteq \mathbb{N}: \text{Indexmenge})$$

**Bemerkung**

Bedingung (ii) (3) von Definition 3.2 gilt nur für sog. diskrete Wahrscheinlichkeitsräume ( $\Omega$  abzählbar), denn für eine überabzählbare Menge  $\Omega$  ist der Ausdruck

$$\sum_{\omega \in \Omega} P(\omega) = 1$$

sinnlos:

Seien  $\mathcal{I}$  eine überabzählbare Menge (z.B. ein Intervall) und  $\sum_{i \in \mathcal{I}} a_i$  eine konvergente Reihe nichtnegativer Zahlen. Dann sind höchstens abzählbar viele Reihenglieder  $a_i > 0$ .

*Beweis*

Sei

$$A_n := \left\{ a_i : a_i \geq \frac{1}{n} \right\} \quad (n \in \mathbb{N}).$$

Da die Reihe  $\sum_{i \in \mathcal{I}} a_i$  nach Voraussetzung konvergiert, ist  $A_n$  für jede Zahl  $n \in \mathbb{N}$  eine endliche Menge. Nun ist

$$A := \{a_i : a_i > 0\} = \bigcup_{n=1}^{\infty} A_n$$

als eine abzählbare Vereinigung endlicher Mengen selbst eine abzählbare Menge.

**Satz 3.1** (Folgerung aus den Wahrscheinlichkeitsaxiomen)

$\langle \Omega, P \rangle$  sei ein diskreter Wahrscheinlichkeitsraum, und  $A, B, A_n \subseteq \Omega$  ( $n \in \mathbb{N}$ ) seien Ereignisse. Dann gelten die folgenden Aussagen:

- (i)  $P(\bar{A}) = 1 - P(A)$
- (ii)  $P(\emptyset) = 0$
- (iii)  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- (iv)  $P(A \cup B) \leq P(A) + P(B)$
- (v)  $P\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} P(A_n)$
- (vi) Alle Aussagen gelten auch, wenn man  $A$  bzw.  $B$  bzw.  $A_n$  durch ihre jeweiligen Komplemente  $\bar{A}$  bzw.  $\bar{B}$  bzw.  $\bar{A}_n$  ersetzt.

*Beweis*

Der Beweis ist einfach und eine direkte Folgerung aus den Wahrscheinlichkeitsaxiomen von Definition 3.2.

- (i) Da  $A \subseteq \Omega$  ein Ereignis ist, beschreibt  $P(\bar{A}) = P(\Omega \setminus A)$  dasjenige Ereignis, daß  $A$  nicht eintritt. Wegen der Disjunktheit von  $A$  und  $\bar{A}$  folgt aus (ii) (2) und (3) der Axiome

$$1 = P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A}),$$

und damit also

$$P(\bar{A}) = 1 - P(A).$$

- (ii) Diese Aussage ist ein Spezialfall der letzten, da auch  $\Omega$  und  $\emptyset$  disjunkte Mengen sind:

$$\begin{aligned} 1 = P(\Omega) &= P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset), \\ \longrightarrow & P(\emptyset) = 0. \end{aligned}$$

- (iii) Diese Aussage heißt *Additionssatz*. Vor dem analytischen Beweis soll dazu eine graphische Illustration gegeben werden, s. Abb. 3.2.

Aus Abb. 3.2 liest man ab:

$P(A \cup B)$  : Summe der Wahrscheinlichkeiten  
der Elementarereignisse in  $A \cup B$

$P(A) + P(B)$  : Summe der Wahrscheinlichkeiten der  
Elementarereignisse von  $A$  + derjenigen von  $B$

$\longrightarrow$  die Elementarereignisse in  $A \cap B$  sind doppelt gezählt worden, ihre Auftretswahrscheinlichkeit muß also in  $P(A) + P(B)$  einmal subtrahiert werden, um das richtige Ergebnis  $P(A \cup B)$  zu erhalten.

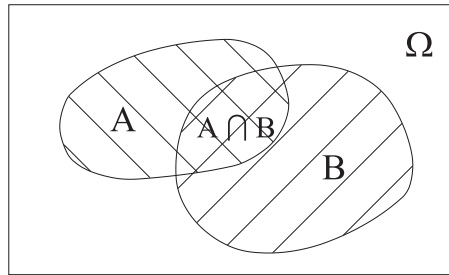


Abbildung 3.2: Graphische Illustration des Additionssatzes

Nun zum korrekten Beweis von (iii) selbst. Man setze  $C := B \setminus (A \cap B)$ , dann ist

$$\begin{aligned} A \cap C &= \emptyset, & A \cup C &= A \cup B, \\ \xrightarrow{(3)} P(A \cup B) &= P(A \cup C) = P(A) + P(C). \end{aligned}$$

Weiterhin ist  $B$  die disjunkte Vereinigung von  $A \cap B$  und  $C$ , also gilt wieder nach (3):

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

Aus beiden Zwischenergebnissen folgt dann die Behauptung:

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

(iv) Klar wegen  $P(A \cap B) \geq 0$ .

(v) Folgt aus (iv) mit vollständiger Induktion.

(vi) Inhalt einer Hausaufgabe.

### Bemerkung

Für mehr als zwei Ereignisse ist der Additionssatz komplizierter. Beispielsweise gilt für drei Ereignisse  $A, B, C$ :

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

**Beispiel 3.7**

Ein homogener Würfel wird zweimal geworfen. Wie groß ist die Wahrscheinlichkeit, mindestens einmal die Augenzahl „6“ zu erzielen?

**Lösung**

$$\begin{aligned} A & : \iff \text{Augenzahl „6“ beim ersten Wurf} \\ B & : \iff \text{Augenzahl „6“ beim zweiten Wurf} \\ A \cap B & : \iff \text{Augenzahl „6“ beim ersten und beim zweiten Wurf} \end{aligned}$$

Es interessiert die Wahrscheinlichkeit des Ereignisses  $A \cup B$ . Da sich die beiden Ereignisse  $A$  und  $B$  *nicht* gegenseitig ausschließen ( $A \cap B \neq \emptyset$ ), muß mit dem Additionssatz (Satz 3.1 (iii)) gearbeitet werden:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= \frac{1}{6} + \frac{1}{6} - \frac{1}{36} \\ &= \frac{11}{36} \end{aligned}$$

**Beispiel 3.8**

Die Wahrscheinlichkeit, die Elektrotechnik-Klausur zu bestehen, sei  $\frac{4}{9}$ , und die Wahrscheinlichkeit, die Mathematik-Klausur zu bestehen, sei doppelt so groß. Die Wahrscheinlichkeit, (wenigstens) eine der beiden Klausuren zu bestehen, sei  $\frac{4}{5}$ .

Wie groß ist dann die Wahrscheinlichkeit, beide Klausuren zu bestehen?

**Lösung**

Die Lösung wird wieder mit dem Additionssatz ermittelt:

$$\begin{aligned} P(E \cap M) &= P(E) + P(M) - P(E \cup M) \\ &= \frac{4}{9} + \frac{8}{9} - \frac{4}{5} \\ &= \frac{60}{45} - \frac{36}{45} \\ &= \frac{8}{15} \end{aligned}$$

**Beispiel 3.9**

Zwei Sportschützen  $A$  und  $B$  schießen auf eine Zielscheibe. Die Wahrscheinlichkeit, daß  $B$  die Scheibe verfehlt, sei um 50% größer als die, daß  $A$  sie trifft;  $A$  treffe sie mit einer Wahrscheinlichkeit von  $\frac{1}{3}$ .

Mit welcher Wahrscheinlichkeit treffen beide Schützen die Scheibe, wenn die Wahrscheinlichkeit eines Treffers überhaupt gegeben ist durch  $P(\text{Treffer}) = \frac{2}{3}$ ?

**Lösung**

Der Additionssatz gilt ebenso für die zu  $A$  und  $B$  komplementären Ereignisse  $\overline{A}$  und  $\overline{B}$  (Satz 3.1 (vi)). Damit erhält man

$$\begin{aligned} P(A) &= \frac{1}{3} \\ P(\overline{B}) &= \frac{3}{2} \cdot \frac{1}{3} = \frac{1}{2} \quad \longrightarrow \quad P(B) = \frac{1}{2}, \end{aligned}$$

und es folgt wieder mit dem Additionssatz:

$$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ &= \frac{1}{3} + \frac{1}{2} - \frac{2}{3} \\ &= \frac{5}{6} - \frac{4}{6} \\ &= \frac{1}{6}. \end{aligned}$$

**3.2.3 Statistische Definition der Wahrscheinlichkeit**

Die bisher entwickelte Theorie diskreter Wahrscheinlichkeitsräume leistet folgendes:

- (a) Jedem Elementarereignis einer Ergebnismenge bzw. eines Merkmalraumes wird eine Zahl zwischen 0 und 1 zugeordnet, ein *Maß* oder ein *Gewicht*.
- (b) Die Summe aller dieser Gewichte ist Eins.
- (c) Das Gewicht einer Menge  $A \subseteq \Omega$  ist gegeben als die Summe der Gewichte aller Elemente von  $A$ .

Diese Zuordnung von Gewichtszahlen wird dann interpretiert als die Zuordnung von Wahrscheinlichkeiten, mit denen man das Auftreten eines Ereignisses in einem statistischen Experiment versieht.

**Problem**

Wie läßt sich in einem konkreten Fall der Praxis diese Zuordnung der Gewichtszahlen so vornehmen, daß hinterher ein Ergebnis herauskommt, welches man sinnvollerweise als Auftrittswahrscheinlichkeit eines zukünftigen Experimentes interpretieren kann? Bis jetzt wurde ja noch nicht festgelegt, *wie* solch eine Zuordnung denn nun vorzunehmen ist.

**Lösung**

Bei diesem sog. *axiomatischen* Zugang zur Wahrscheinlichkeitstheorie bleibt der Begriff der „Wahrscheinlichkeit eines Ereignisses“ undefiniert, es wird nur

sein Gebrauch durch Axiome geregelt. Letztendlich läßt sich daher *jede* Zuordnung vornehmen, die den Axiomen eines diskreten Wahrscheinlichkeitsraumes genügt.

Für praktische Untersuchungen sind natürlich nur solche Zuordnungen sinnvoll, welche in irgendeiner Weise die Realität widerspiegeln und dann sinnvoll als *Wahrscheinlichkeitsmaß* interpretiert werden können. Hier ist also – unabhängig von der Mathematik – erst einmal eine Übereinkunft über ein Modell zu erzielen, mit welchem man ein gegebenes Problem beschreiben und lösen will. Man betrachte in diesem Zusammenhang Beispiel 1.4.

### Laplace-Experimente

Im einfachsten Fall liegt ein Laplace-Experiment vor. Hat der Merkmalraum  $\Omega$   $m$  Elemente und ein Ereignis  $A \subseteq \Omega$   $g(A)$  Elemente, so läßt sich definieren:

$$P(A) := \frac{g(A)}{m} \quad (A \subseteq \Omega).$$

Diese Definition erfüllt die Gesetze eines Wahrscheinlichkeitsraumes nach Kolmogorov.

### Beispiele

- Werfen einer Münze
- Würfeln mit einem homogenen Würfel
- Ziehen einer Karte aus einem Stapel „gleicher“ Karten

### Andere Experimente

Als Grundlage für die Festlegung von Wahrscheinlichkeiten dient die Erfahrung, daß sich die relativen Häufigkeiten zufälliger Ereignisse bei umfangreichen Versuchsreihen i.a. stabilisieren, d.h. gegen einen Grenzwert konvergieren. Diesen vermuteten Grenzwert der relativen Häufigkeiten  $h_n^{(r)}(A)$  eines Ereignisses  $A \subseteq \Omega$  nimmt man als das Wahrscheinlichkeitsmaßes dieses Ereignisses:

$$P(A) := \lim_{n \rightarrow \infty} h_n(A) \quad (A \subseteq \Omega).$$

Man beachte aber, daß hier die Existenz des Grenzwertes eine Erfahrungstat-  
sache ist und nicht bewiesen werden kann!

### Beispiel 3.10

Man betrachte noch einmal das Urnenbeispiel 3.2 einer Urne mit einer weißen -, zwei grauen - und drei schwarzen Kugeln. Bei umfangreichen Versuchsreihen erhält man für die drei Ereignisse

- $W$  : $\iff$  Ziehen einer weißen Kugel
- $G$  : $\iff$  Ziehen einer grauen Kugel
- $S$  : $\iff$  Ziehen einer schwarzen Kugel

die relativen Häufigkeiten

$$\begin{aligned} h_n^{(r)}(W) &\approx \frac{1}{6} \\ h_n^{(r)}(G) &\approx \frac{2}{6} = \frac{1}{3} \\ h_n^{(r)}(S) &\approx \frac{3}{6} = \frac{1}{2} \end{aligned}$$

Diese Zahlen werden dann als Wahrscheinlichkeiten für das Auftreten der Ereignisse  $A, B, C$  festgelegt. Es handelt sich hierbei um keine Definition im mathematischen Sinne, dennoch spricht man oft von einer

*statistischen Definition der unbekanntenen Wahrscheinlichkeitswerte*

oder von einer

*Definition der Wahrscheinlichkeitswerte gemäß relativer Häufigkeiten.*

### Problem

Wie soll man Wahrscheinlichkeiten festlegen für folgende Experimente?

- Wurf mit einem nicht homogenen Würfel
- Werfen eines Reißnagels auf eine ebene harte Platte mit zwei möglichen Ausgängen

## 3.3 Bedingte Wahrscheinlichkeit

### 3.3.1 Einführung und Beispiele

Häufig ist man an Wahrscheinlichkeiten von Ereignissen interessiert, welche nur in einem speziellen Teil des Ereignisraumes enthalten sind.

#### Beispiel 3.11

Aufgrund genetischer Forschung ist bekannt, daß zwischen Augenfarbe und Haarfarbe ein Zusammenhang besteht. Damit sind die folgenden Wahrscheinlichkeiten verschieden:

- die Wahrscheinlichkeit, daß eine zufällig ausgewählte Person blaue Augen hat;
- die Wahrscheinlichkeit, daß eine Person blaue Augen hat, wenn man sie zufällig aus einer Menge von Personen auswählt, welche alle blonde Haare haben.



Hier interessieren also Zusatzbedingungen, welche nur durch einen Teil der Bevölkerung definiert werden. Mit solchen Zusatzbedingungen zusammenhängende Wahrscheinlichkeiten bezeichnet man als *bedingte Wahrscheinlichkeiten*.

Der Begriff der bedingten Wahrscheinlichkeit ist auch dann nützlich, wenn im Verlauf eines Experimentes die Bedingungen geändert werden, so daß über den möglichen Ausgang schon Teilinformationen vorliegen.

### Beispiel 3.12

Wird ein homogener Würfel einmal geworfen, so ist die Wahrscheinlichkeit des Auftretens einer „6“ gleich  $\frac{1}{6}$ .

Hat man jedoch die Zusatzinformation, daß das Ergebnis eine gerade Zahl ist, so wird die Wahrscheinlichkeit einer „6“ gleich  $\frac{1}{3}$ . In dieser Situation ist der Merkmalraum  $\Omega$  verkleinert worden, wodurch die Wahrscheinlichkeit eines möglichen Ereignisses steigt, da weniger Ergebnisse als vorher zur Konkurrenz zugelassen sind.

An einem weiteren Beispiel soll untersucht werden, *wie* Zusatzinformationen über den Ausgang eines Experimentes die Ergebnismenge verkleinern und damit die interessierende Auftrittswahrscheinlichkeit vergrößern.

### Beispiel 3.13

Man betrachte eine Menge von Familien mit genau zwei Kindern unterschiedlichen Alters – also keine eineiigen Zwillinge – und wähle davon eine Familie zufällig aus. Wie groß ist die Wahrscheinlichkeit, daß beide Kinder Jungen sind?

Die Ergebnismenge ist

$$\Omega := \{\langle J, J \rangle, \langle J, M \rangle, \langle M, J \rangle, \langle M, M \rangle\},$$

und es gilt  $P(\langle J, J \rangle) = \frac{1}{4}$ .

Nun erfährt man, daß die Daten von Familien mit mindestens einem Jungen stammen. Wie wirkt sich das auf unsere Wahrscheinlichkeit aus?

Dazu muß ein verkleinerter Merkmalraum – der sog. *reduzierte Merkmalraum*

$$\Omega^* := \{\langle J, J \rangle, \langle J, M \rangle, \langle M, J \rangle\}$$

untersucht werden. Damit wird die Wahrscheinlichkeit untersucht, daß eine Familie zwei Jungen hat *unter der Voraussetzung*, daß sie wenigstens einen Jungen hat. Seien dazu die folgenden beiden Ereignisse definiert:

$A : \iff$  Zwei Jungen

$B : \iff$  Wenigstens ein Junge

Dann wird die Wahrscheinlichkeit von  $A$  unter der Bedingung  $B$  mit

$$P(A | B)$$

bezeichnet und *bedingte Wahrscheinlichkeit von  $A$  unter der Bedingung  $B$*  genannt.

Welche Wahrscheinlichkeiten sollen nun den Ergebnissen in  $\Omega^*$  zugeordnet werden? Zur Erinnerung:<sup>4</sup>

$$\begin{aligned} P(X) &= \frac{\text{Anzahl der für } X \text{ günstigen Fälle}}{\text{Anzahl der für } X \text{ möglichen Fälle}} \\ \longrightarrow P(A | B) &= \frac{|A \cap B|}{|B|} \quad \left[ \text{hier: } = \frac{|A|}{|B|} \text{ wegen } A \subseteq B \right] \\ &= \frac{|A \cap B|}{|\Omega|} \cdot \frac{|\Omega|}{|B|} = \frac{\frac{|A \cap B|}{|\Omega|}}{\frac{|B|}{|\Omega|}} \\ &= \frac{P(A \cap B)}{P(B)}. \end{aligned}$$

Die Wahrscheinlichkeiten auf der rechten Seite dieser Formel gehören nun zum ursprünglichen Versuch. In dem Beispiel ist

$$P(A | B) = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

Ein und dasselbe Ereignis kann also durchaus unterschiedliche Auftretenswahrscheinlichkeiten haben, wenn es auf verschiedene Ergebnismengen bezogen wird.

### Definition 3.3

Die Wahrscheinlichkeit für das Eintreten des Ereignisses  $A$  unter der Voraussetzung, daß das Ereignis  $B$  bereits eingetreten ist, heißt *bedingte Wahrscheinlichkeit von  $A$  unter der Bedingung  $B$*  oder *konditionale Wahrscheinlichkeit* und wird definiert durch

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

falls  $P(B) \neq 0$  ist.

### Bemerkung

- (i) Ist  $P(B) = 0$ , so ist wegen  $A \cap B \subseteq B$  auch  $P(A \cap B) \leq P(B) = 0$ .
- (ii) Die bedingte Wahrscheinlichkeit eines Ereignisses bezieht sich auf einen verkleinerten Merkmalraum  $\Omega^*$  von  $\Omega$ . Insofern ist *jede* Wahrscheinlichkeit eines Ereignisses  $A$  eine bedingte Wahrscheinlichkeit bzgl.  $\Omega$ :

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

<sup>4</sup>Für endliche Mengen  $X$  bezeichne  $|X|$  die Anzahl der Elemente von  $X$ .

(iii) Seien  $B \subseteq \Omega$  und

$$\begin{aligned}\Omega^* &:= \Omega_B := B \\ \mathcal{M}_B &:= \{A \cap B : A \subseteq \Omega\}\end{aligned}$$

$\mathcal{M}_B$  heißt die *Spur* von  $P(\Omega)$  auf  $B$ . Die durch das Ereignis  $B$  bedingten Wahrscheinlichkeiten  $P(A | B)$  bilden ein Wahrscheinlichkeitsmaß auf der Menge  $\mathcal{M}_B$ :

$$P(\cdot | B) : \mathcal{M}_B \longrightarrow [0, 1].$$

Dadurch wird  $(\Omega_B, P(\cdot | B))$  zu einem Wahrscheinlichkeitsraum. Er ist das mathematische Modell des neuen, aus einem Versuch  $V$  abgeleiteten Versuchs  $V_B$ .

Damit gelten für bedingte Wahrscheinlichkeiten alle bisher abgeleiteten Rechenregeln, z.B. der Additionssatz, s. Satz 3.1 (iii):

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

### Aufgabe

Man betrachte noch einmal Beispiel 3.13. Anstatt zu erfahren, daß die Daten von Familien mit mindestens einem Jungen stammen, erfährt man nun, daß das älteste von zwei Kindern jeweils ein Junge ist. Wirkt sich das auf die Wahrscheinlichkeit aus? Wenn ja, wie?

### Beispiel 3.14

Gegeben sei ein Würfel, bei dem die Wahrscheinlichkeit des Auftretens einer geraden Zahl doppelt so groß ist wie die des Auftretens einer ungeraden Zahl. Sei  $A$  das Ereignis

$$A : \iff \text{Eine 4 wird gewürfelt.}$$

Unter Zugrundelegung des Merkmalraumes

$$\Omega := \{1, 2, 3, 4, 5, 6\}$$

mit den Gewichtsfaktoren  $\frac{1}{9}$  bzw.  $\frac{2}{9}$  (cf. Beispiel 3.5) ist  $P(A) = \frac{2}{9}$ .

Nun sei zusätzlich bekannt, daß das Resultat des Wurfes größer als 3 ist, d.h.  $B$  ist das Ereignis

$$B : \iff \text{Das Ergebnis des Wurfes ist größer als 4,}$$

und  $A$  ist dann enthalten in einem reduzierten Merkmalraum:

$$A \subseteq \Omega^* := \{4, 5, 6\}.$$

Welches ist dann die Wahrscheinlichkeit des Ereignisses  $A$  relativ zum neuen Merkmalraum  $\Omega^*$ ? Um das herauszufinden, gibt es zwei Möglichkeiten:

- (i) Den Elementarereignissen von  $\Omega^*$  müssen neue Gewichte gegeben werden, deren Summe wieder gleich 1 ist:

$$2w + w + 2w = 5w = 1 \quad \longrightarrow \quad w = \frac{1}{5}.$$

Bezogen auf den reduzierten Merkmalraum  $\Omega^*$  gilt also

$$P(A | B) = \frac{2}{5}.$$

- (ii) Dieses Ergebnis hätte man mit Hilfe von Definition 3.3 auch direkt ausrechnen können:

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{9}}{\frac{5}{9}} = \frac{2}{5}.$$

Man beachte, daß es sich bei den Größen  $P(A \cap B)$  und  $P(B)$  um Größen handelt, für welche der ursprüngliche Merkmalraum  $\Omega$  zugrundegelegt worden ist.

### 3.3.2 Multiplikationssatz, Totale Wahrscheinlichkeit, Formel von Bayes

Die Definitionsgleichung für die bedingte Wahrscheinlichkeit lautet

$$P(A | B) := \frac{P(A \cap B)}{P(B)} \quad (P(B) \neq 0),$$

bzw. unter Vertauschung der Rollen von  $A$  und  $B$ :

$$P(B | A) := \frac{P(B \cap A)}{P(A)} \quad (P(A) \neq 0).$$

Daraus ergibt sich wegen  $A \cap B = B \cap A$  der *Multiplikationssatz für Wahrscheinlichkeiten* oder die *Gleichung von Bayes*:

$$\boxed{P(A \cap B) = P(A) \cdot P(B | A) = P(B) \cdot P(A | B)} \quad (3.4)$$

Die Wahrscheinlichkeit für das gleichzeitige Eintreten der Ereignisse  $A$  und  $B$  ist gleich dem Produkt von

- der Wahrscheinlichkeit des Eintretens von  $A$  und
- der Wahrscheinlichkeit des Eintretens von  $B$  unter der Voraussetzung, daß  $A$  eintritt bzw. eingetreten ist.
- Oder umgekehrt

Man beachte, daß bei dem Multiplikationssatz die Voraussetzung  $P(A) = 0$  bzw.  $P(B) = 0$  mit eingeschlossen ist.

### Beispiel 3.15

Gegeben sei ein Sicherungskasten mit zwanzig Sicherungen, von denen fünf defekt sind. Dem Kasten werden zwei Sicherungen zufällig entnommen.

Wie groß ist die Wahrscheinlichkeit, daß die beiden entnommenen Sicherungen defekt sind?

### Lösung

$S_1$  : $\iff$  die erste Sicherung ist defekt

$S_2$  : $\iff$  die zweite Sicherung ist defekt

Dann ist man an der Wahrscheinlichkeit des Ereignisses  $S_1 \cap S_2$  interessiert, und es folgt mit dem Multiplikationssatz (3.4)

$$P(S_1) = \frac{5}{20} = \frac{1}{4} \quad , \quad P(S_2 | S_1) = \frac{4}{19}$$

$$\longrightarrow P(S_1 \cap S_2) = P(S_1) \cdot P(S_2 | S_1) = \frac{1}{4} \cdot \frac{4}{19} = \frac{1}{19}.$$

### Bemerkung

Der Multiplikationssatz oder die Gleichung von Bayes läßt sich auch auf mehr als zwei Ereignisse verallgemeinern. Beispielsweise lautet er für drei gleichzeitig eintretende Ereignisse  $A, B$  und  $C$ :

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

### Beweis

O.B.d.A. seien  $P(A)$  und  $P(A \cap B) \neq 0$ . Dann folgt

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

Im folgenden soll eine Verallgemeinerung des Multiplikationssatzes besprochen werden. Dazu braucht man einen neuen Begriff.

**Definition 3.4**

Eine Menge  $\mathcal{B}$  von endlich vielen oder abzählbar unendlich vielen Teilmengen  $B_n \subseteq \Omega$  heißt eine *Partition* oder eine *Ausschöpfung* von  $\Omega$ , wenn gilt:

$$\begin{aligned} B_n &\neq \emptyset && (n \in \mathbb{N}) \\ \bigcup_{n=1}^{\infty} B_n &= \Omega \\ B_m \cap B_n &= \emptyset && (m \neq n) \end{aligned}$$

Für jedes Ereignis  $A \subseteq \Omega$  gilt dann (Abb. 3.3)

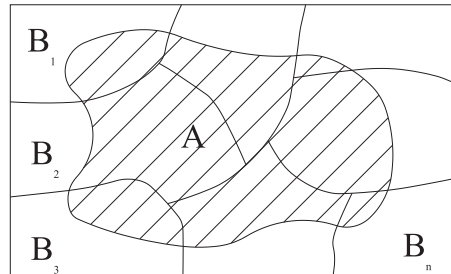


Abbildung 3.3: Spur von  $A$  in einer Partition von  $\Omega$

$$A = A \cap \Omega = A \cap \left( \bigcup_{n=1}^{\infty} B_n \right) = \bigcup_{n=1}^{\infty} (A \cap B_n).$$

Nun ist für  $m \neq n$

$$(A \cap B_m) \cap (A \cap B_n) = A \cap B_m \cap B_n = \emptyset,$$

folglich handelt es sich bei dem letzten Ausdruck um eine disjunkte Vereinigung, für welche nach den Kolmogorov-Axiomen (Definition 3.2 (ii) (3)) gilt:

$$P(A) = \sum_{n=1}^{\infty} P(A \cap B_n) = \sum_{n=1}^{\infty} P(A | B_n) \cdot P(B_n) \quad (3.5)$$

Diese Aussage heißt *Satz von der totalen Wahrscheinlichkeit für unvereinbare Ereignisse*.

**Beispiel 3.16**

In einem medizinischen Diagnose-Test (etwa einem Hauttest zur möglichen Erkennung von Tuberkulose oder einem HIV-Test) können zwei Arten von Fehlern auftreten:

(i) **Falsch-Positiver Fehler**

Der Test zeigt eine Krankheit  $K$  an, welche die untersuchte Person tatsächlich nicht hat.

(ii) **Falsch-Negativer Fehler**

Der Test zeigt die Krankheit  $K$  nicht an, obwohl die untersuchte Person sie hat.

Aus der Vergangenheit seien folgende Daten bekannt:

- 1 % der Bevölkerung habe die Krankheit  $K$  (die sog. *Prävalenz*);
- die Wahrscheinlichkeiten für einen Falsch-Positiven Test bzw. für einen Falsch-Negativen Test seien gegeben durch

$$P(+ | \bar{K}) = 0.1 \quad , \quad P(- | K) = 0.02.$$

Wie groß ist die Wahrscheinlichkeit dafür, daß der Test bei einer beliebigen Person positiv ausfällt?

**Lösung**

Die gesamte Bevölkerung läßt sich disjunkt zerlegen in Kranke und Nicht-Kranke:

$$\Omega = K \cup \bar{K}.$$

Dann gilt mit Formel (3.5) von der totalen Wahrscheinlichkeit

$$\begin{aligned} P(+) &= P[(+ \cap K) \cup (+ \cap \bar{K})] \\ &= P(+ \cap K) + P(+ \cap \bar{K}) \\ &= P(K \cap +) + P(\bar{K} \cap +) \\ &= P(K) \cdot P(+ | K) + P(\bar{K}) \cdot P(+ | \bar{K}) \\ &= 0,01 \cdot (1 - 0,02) + (1 - 0,01) \cdot 0,1 \\ &= 0,1088 = 10,88\% \end{aligned}$$

Bei ca. 11 % der untersuchten Personen fällt der Test positiv aus; diese müssen eingehender untersucht werden.

Nach dem Multiplikationssatz (Gleichung von Bayes (3.4)) und dem Satz von der totalen Wahrscheinlichkeit (3.5) soll als drittes Ergebnis in diesem Abschnitt die *Formel von Bayes* besprochen werden.

**Satz 3.2** (Bayes'sche Formel)

Sei  $\{B_n : n \in \mathbb{N} \text{ geeignet}\}$  eine (abzählbare) Menge paarweise disjunkter Ereignisse, welche die Ergebnismenge  $\Omega$  ausschöpfen, d.h. eine Partition von  $\Omega$ , s. Definition 3.4.

Dann gilt:

$$P(B_k | A) = \frac{P(A | B_k) \cdot P(B_k)}{\sum_{n=1}^{\infty} P(A | B_n) \cdot P(B_n)} \quad (k \in \mathbb{N}, A \subseteq \Omega) \quad (3.6)$$

Diese Formel dient zur Bestimmung der „umgekehrten“ bedingten Wahrscheinlichkeiten  $P(B_k | A)$  bei gegebenen  $P(A | B_k)$  und  $P(B_k)$ .

*Beweis*

Der Satz von der totalen Wahrscheinlichkeit (3.5)

$$P(A) = \sum_{n=1}^{\infty} P(A | B_n) \cdot P(B_n)$$

und die sich aus der Gleichung von Bayes (3.4) ergebende Formel

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

zusammen liefern für  $B := B_k$ :

$$P(B_k | A) = \frac{P(A | B_k) \cdot P(B_k)}{P(A)} = \frac{P(A | B_k) \cdot P(B_k)}{\sum_{n=1}^{\infty} P(A | B_n) \cdot P(B_n)}.$$

### Beispiel 3.17

Es wird noch einmal zu Beispiel 3.16 mit dem medizinischen Diagnose-Test zurückgekehrt und gefragt, mit welcher Wahrscheinlichkeit eine untersuchte Person tatsächlich die Krankheit  $K$  hat, wenn der Test positiv ausfällt.

### Lösung

Die Lösung erhält man mit der Bayes'schen Formel (3.6):

$$\begin{aligned} P(K | +) &= \frac{P(+ | K) \cdot P(K)}{P(+ | K) \cdot P(K) + P(+ | \bar{K}) \cdot P(\bar{K})} \\ &= \frac{P(+ | K) \cdot P(K)}{P(+)} \\ &= \frac{0,98 \cdot 0,01}{0,1088} \\ &= 0,09 = 9\% \end{aligned}$$



In diesem Fall haben also 91 % der Personen, welche auf den Test positiv reagieren, die Krankheit *nicht*.

### Begründung

Die Krankheit ist mit 1 % recht selten, aber die Rate  $P(+ | \bar{K})$  der Falsch-Positiv-Reaktionen beträgt immerhin 10%. Daher muß der Großteil der Positiv-Reaktionen an

$$+ = (+ \cap K) \cup (+ \cap \bar{K})$$

von der gesunden Bevölkerung herrühren, also

$$P(\bar{K} | +) = 0,91 = 91\%.$$

### Bemerkung

Medizinische Tests sind immer etwas ungenau. Bei dem quantitativen Diagnose-Test oben etwa möchte man die Fehler klein halten, welche von Falsch-Positiven bzw. von Falsch-Negativen Reaktionen herrühren. Wie man an diesem Beispiel erkennt, läßt sich einer der Fehler nur auf Kosten des anderen verkleinern.

Man muß sich einen solchen Test als eine Art Netz vorstellen, mit dem man Fische einer bestimmten Größe fangen will. Sind die Maschen des Netzes zu eng, fängt man zwar alle Fische, die man fangen will, aber auch viele kleine Fische, die man nicht fangen will. Macht man hingegen die Maschen des Netzes zu weit, so sind zwar keine kleinen Fische dabei, aber es schlüpft auch ein Teil der größeren Fische, die man fangen will, durch die Maschen. Bei engen Maschen ist die Ausbeute hoch, aber nicht notwendig richtig, bei weiten Maschen verhält es sich genau umgekehrt.

Im medizinischen Sprachgebrauch hat sich für die beiden Fehlerarten der folgende Sprachgebrauch etabliert:

- Die (*diagnostische*) *Spezifität* eines Tests ist der Anteil der korrekt klassifizierten Gesunden, gemessen an der Anzahl aller Gesunden:

$$\text{Spezifität} = \frac{P(\text{negativ} | \text{gesund})}{P(\text{gesund}) = P(\text{gesund} \cap \text{negativ}) + P(\text{gesund} \cap \text{positiv})}$$

- Die (*diagnostische*) *Sensitivität* eines Tests ist der Anteil der korrekt klassifizierten Kranken, gemessen an der Anzahl aller Kranken:

$$\text{Sensitivität} = \frac{P(\text{positiv} | \text{krank})}{P(\text{krank}) = P(\text{krank} \cap \text{positiv}) + P(\text{krank} \cap \text{negativ})}$$

- Die *Effizienz* eines Tests ist eine aus Spezifität und Sensitivität kombinierte Kennzahl dafür, wie viele richtig positive und richtig negative Befunde ein diagnostisches Verfahren liefert, gemessen an der Gesamtzahl aller untersuchten Patienten:

$$\text{Effizienz} = \frac{P(\text{negativ} | \text{gesund}) + P(\text{positiv} | \text{krank})}{n = \text{Anzahl aller Patienten}}$$

Übertragen auf einen Test, bedeuten weite Maschen eine hohe Spezifität: man möchte nicht unbedingt alle Erkrankten finden, aber diejenigen, die man herausfiltert, sollen auf jeden Fall krank sein. Umgekehrt bedeuten enge Maschen eine hohe Sensitivität: man möchte möglichst alle kranken Personen aus einer zu testenden Personengruppe im wahrsten Sinne des Wortes herausfischen, selbst unter Inkaufnahme, daß man einige Gesunde dabei mitherausfischt.

Damit gilt für einen Test:

- Hohe Spezifität = weite Maschen = kleiner Falsch-Positiver Fehler  
Die diagnostische Spezifität eines Testes drückt aus, bei wieviel Prozent der Gesunden mit diesem Test eine bestimmte Erkrankung ausgeschlossen werden kann. Ein Test mit einer hohen Spezifität ist daher nützlich, um eine (vermutete) Krankheit zu bestätigen.
- Hohe Sensitivität = enge Maschen = kleiner Falsch-Negativer Fehler  
Die diagnostische Sensitivität eines Testes drückt aus, bei wieviel Prozent der Erkrankten mit diesem Test eine bestimmte Erkrankung erkannt werden kann. Ein Test mit einer hohen Sensitivität ist daher nützlich, um alle Kranken zu erfassen.
- Eine zu geringe diagnostische Spezifität eines Testes stellt u.a. ein wirtschaftliches Problem dar, während eine zu geringe diagnostische Sensitivität Patienten gefährdet.

Ein erster Diagnose-Test sollte also ein solcher mit einer hohen Sensitivität sein, während ein nachfolgender bei einer zweiten Untersuchung eine hohe Spezifität aufweisen sollte. Es muß dabei allerdings betont werden, daß es sich bei allen Tests um Wahrscheinlichkeitsaussagen handelt, welche gute Vorhersagen für Gruppen von Menschen liefern können, nicht aber unbedingt für Einzelpersonen.

### Übersicht

Multiplikationssatz für Wahrscheinlichkeiten = Gleichung von Bayes

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

Satz von der totalen Wahrscheinlichkeit für unvereinbare Ereignisse

$$P(A) = \sum_{n=1}^{\infty} P(A \cap B_n) = \sum_{n=1}^{\infty} P(A | B_n) \cdot P(B_n)$$

Formel von Bayes

$$P(B_k | A) = \frac{P(A | B_k) \cdot P(B_k)}{\sum_{n=1}^{\infty} P(A | B_n) \cdot P(B_n)} \quad (k \in \mathbb{N}, A \subseteq \Omega)$$

**Bemerkung**<sup>5</sup>

Häufig wird der Fehler gemacht, das bedingte Ereignis mit der Bedingung zu verwechseln und auf diese Weise Wahrscheinlichkeiten auszurechnen, welche gar nicht weiter interessant sind.

**Der Tod fährt mit! Vier von zehn tödlich verunglückten  
Autofahrern trugen keinen Sicherheitsgurt.**

Was *will* uns diese Schlagzeile sagen? Wahrscheinlich das, daß es sicherer ist, angeschnallt zu fahren.

Was *sagt* uns die Schlagzeile tatsächlich? Es ist sicherer, nicht angeschnallt zu fahren, denn von den zehn tödlich verunglückten Unfallopfern waren immerhin sechs angeschnallt.

Hier lautet die Bedingung:  $U \iff$  Man ist Unfallopfer,  
und das Ereignis:  $A \iff$  Man ist angeschnallt.

Somit wird das bedingte Ereignis  $A | U$  untersucht, und dessen Wahrscheinlichkeit ist tatsächlich  $0,6 = 60\%$ :

$$P(A | U) = \frac{P(A \cap U)}{P(U)} = \frac{6/10}{10/10} = 0,6.$$

Interessant wäre hier die umgekehrte bedingte Wahrscheinlichkeit  $P(U | A)$ , also die Wahrscheinlichkeit, als Unfallopfer zu enden, wenn man angeschnallt ist, und das ist etwas ganz Anderes:

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

Im zweiten Fall wird durch eine viel größere Zahl dividiert, nämlich durch die Wahrscheinlichkeit  $P(A)$ , angeschnallt zu fahren. Diese Zahl sollte sehr viel größer sein als die Wahrscheinlichkeit  $P(U)$ , als Unfallopfer zu enden.

Um eine Kausalbeziehung empirisch abzustützen, sollte daher das bedingende Ereignis die Ursache und das bedingte Ereignis die Wirkung sein, nicht umgekehrt:

Ursache : Man ist angeschnallt  
Wirkung : Man ist Unfallopfer

---

<sup>5</sup>W. Krämer: Denkste! Trugschlüsse aus der Welt der Zahlen und des Zufalls, Piper Verlag, 2. Auflage 1999, p. 137

### Weitere Beispiele

- (i) Die Wahrscheinlichkeit, daß jemand, der deutsch spricht, ein Deutscher ist, ist weder gleich groß der Wahrscheinlichkeit, daß jemand, der ein Deutscher ist, auch deutsch spricht, noch ergänzen sich beide Wahrscheinlichkeiten zu 100 %.
- (ii) In einem Strafprozeß in Kalifornien aus dem Jahr 1968 wurde ein männlicher Bankräuber unter anderem deshalb verurteilt, weil der Täter gemäß Zeugenaussagen einen Bart und einen Schnurrbart trug. Wer einen Bart hat, hat sehr oft auch einen Schnurrbart – das Gericht ging in seinem Fehlurteil aber nicht von bedingten Wahrscheinlichkeiten aus.

### 3.3.3 Mehrstufige Zufallsexperimente

Viele Zufallsprozesse bestehen nicht nur aus einem, sondern aus einer endlichen Folge von Zufallsexperimenten. Es handelt sich dann um ein sog. *mehrstufiges Zufallsexperiment*.

#### Beispiel 3.18

In einer Urne befinden sich acht Kugeln, drei weiße und fünf schwarze. Der Urne werden nacheinander zwei Kugeln entnommen, die entnommenen Kugeln jedoch nicht zurückgelegt.

Mit welcher Wahrscheinlichkeit erhält man

- (i) zwei gleichfarbige Kugeln;
- (ii) zwei verschiedenfarbige Kugeln?

#### Lösung

Dieser Zufallsprozeß besteht aus einem zweistufigen Zufallsexperiment. Seien

$$\begin{aligned}
 G & : \iff \text{zwei gleichfarbige Kugeln} \\
 V & : \iff \text{zwei verschiedenfarbige Kugeln} \\
 W_1 & : \iff \text{weiße Kugel im ersten Zug} \\
 & \vdots \\
 S_2 & : \iff \text{schwarze Kugel im zweiten Zug}
 \end{aligned}$$

Damit erhält man

$$\begin{aligned}
 G & = (W_1 \cap W_2) \cup (S_1 \cap S_2) \\
 V & = (S_1 \cap W_2) \cup (W_1 \cap S_2),
 \end{aligned}$$

also wegen der Disjunktheit aller Ereignisse

$$\begin{aligned}
 P(G) &= P(W_1 \cap W_2) + P(S_1 \cap S_2) \\
 &= P(W_1) \cdot P(W_2 | W_1) + P(S_1) \cdot P(S_2 | S_1) \\
 &= \frac{3}{8} \cdot \frac{2}{7} + \frac{5}{8} \cdot \frac{4}{7} \\
 &= \frac{13}{28},
 \end{aligned}$$

$$\begin{aligned}
 P(V) &= P(W_1 \cap S_2) + P(S_1 \cap W_2) \\
 &= P(W_1) \cdot P(S_2 | W_1) + P(S_1) \cdot P(W_2 | S_1) \\
 &= \frac{3}{8} \cdot \frac{5}{7} + \frac{5}{8} \cdot \frac{3}{7} \\
 &= \frac{15}{28} \\
 &= 1 - P(G).
 \end{aligned}$$

Ein anschauliches Hilfsmittel bei der Berechnung von Wahrscheinlichkeiten mehrstufiger Zufallsexperimente ist der *Ereignisbaum* oder das *Baumdiagramm*. Ein Ereignisbaum besteht aus

- einer *Wurzel* (Ausgangspunkt),
- mehreren *Verzweigungspunkten* und
- mehreren *Zweigen*.

Die Verzweigungspunkte charakterisieren dabei die Zwischenergebnisse nach der jeweiligen Stufe, und jeder Zweig wird mit der entsprechenden Wahrscheinlichkeit versehen. Abb. 3.4 zeigt den Ereignisbaum für Beispiel 3.18.

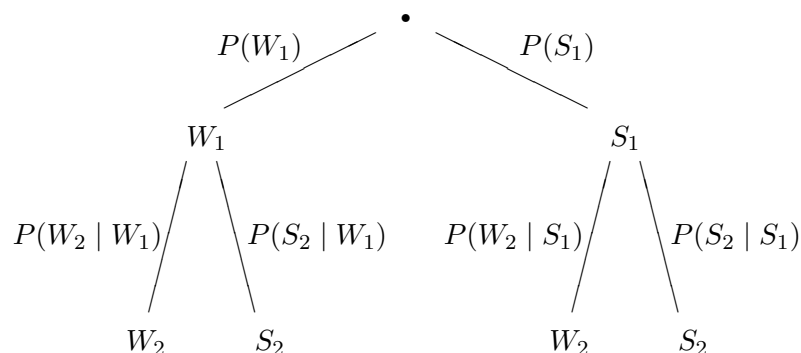


Abbildung 3.4: Zweimalige Ziehung zweier Kugeln ohne Zurücklegen

Da die verschiedenen Stufen miteinander in Beziehung stehen, hängen die Wahrscheinlichkeiten der zweiten Stufe vom Ausgang des Experimentes auf der ersten Stufe ab: sie sind bedingte Wahrscheinlichkeiten.

Bei einem in mehreren Stufen ablaufenden Zufallsexperiment läßt sich jedes mögliche Endergebnis durch einen bestimmten Pfad im Ereignisbaum darstellen. Die Berechnung der dafür gültigen Wahrscheinlichkeiten erfolgt dabei anhand der folgenden beiden Regeln:

- Die Wahrscheinlichkeiten eines Pfades werden miteinander *multipliziert*.
- Führen mehrere Pfade zum gleichen Endergebnis, so werden ihre Wahrscheinlichkeiten *addiert*.

### Beispiel 3.19

Eine Kiste enthalte drei defekte und sieben einwandfreie Glühbirnen. Nacheinander werden drei Glühbirnen zufällig ausgewählt.

- Wie sieht der entsprechende Ereignisbaum aus?
- Welches ist die Wahrscheinlichkeit dafür, daß
  - wenigstens eine einwandfreie Glühbirne ausgewählt wird;
  - alle Glühbirnen einwandfrei sind unter der Bedingung, daß die erste Birne einwandfrei ist?

### Lösung

Dieser Zufallsprozeß besteht aus einem dreistufigen Zufallsexperiment mit einem Ereignisbaum wie in Abb. 3.5.

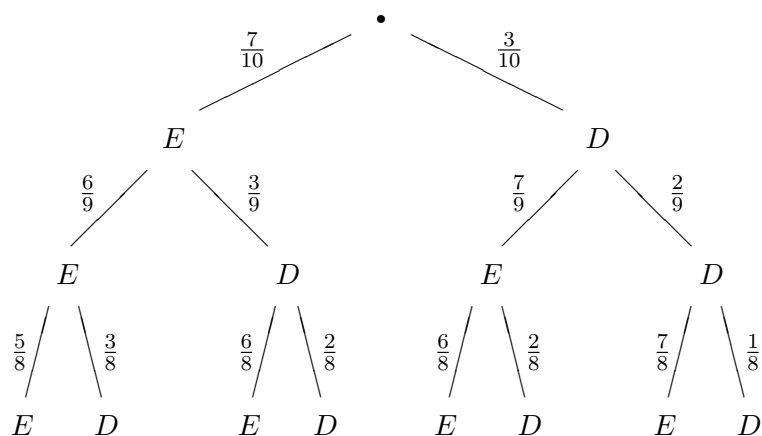


Abbildung 3.5: Dreimalige Ziehung einer Glühbirne ohne Zurücklegen

Alle bis auf den letzten Pfad führen zu mindestens einer einwandfreien Glühbirne, somit ist

$$P((a)) = 1 - \frac{3}{10} \cdot \frac{2}{9} \cdot \frac{1}{8} = 1 - \frac{1}{120} = \frac{119}{120}$$

$$P((b)) = 1 \cdot \frac{6}{9} \cdot \frac{5}{8} = \frac{5}{12}$$

**Beispiel 3.20**

Auf einem binären Kanal werde eine Nachricht als Sequenz der beiden binären Signale **O** und **L** übertragen. Bei der Übertragung treten Fehler auf derart, daß im Mittel

$$40\% \text{ der Information } \mathbf{O} \quad \text{und} \quad 33,3\% \text{ der Information } \mathbf{L}$$

fehlerhaft übertragen werden. Das Verhältnis der übertragenen Signale betrage  $\mathbf{O} : \mathbf{L} = 5 : 3$ .

Wie groß ist die Wahrscheinlichkeit, daß das empfangene Signal mit dem gesendeten übereinstimmt, wenn

- (i) das empfangene Signal ein **O** ist;
- (ii) das empfangene Signal ein **L** ist?

**Lösung**

$OG$  : $\iff$  **O** gesendet

$OE$  : $\iff$  **O** empfangen

$LG$  : $\iff$  **L** gesendet

$LE$  : $\iff$  **L** empfangen

**Beh. 1**      $P(OG) = \frac{5}{8}$  ,    $P(LG) = \frac{3}{8}$  .

*Bew.*      $P(OG) : P(LG) = 5 : 3$       $\wedge$       $P(LG) + P(OG) = 1$ .

Folgende Wahrscheinlichkeiten sind bekannt:

$$P(OE | OG) = \frac{3}{5} \quad , \quad P(LE | OG) = \frac{2}{5}$$

$$P(LE | LG) = \frac{2}{3} \quad , \quad P(OE | LG) = \frac{1}{3}$$

**Beh. 2**      $P(OE) = \frac{1}{2}$  ,    $P(LE) = \frac{1}{2}$  .

*Bew.*     Mit dem Satz von der totalen Wahrscheinlichkeit (3.5) folgt

$$\begin{aligned} P(OE) &= P(OE \cap OG) && + && P(OE \cap LG) \\ &= P(OE | OG) \cdot P(OG) && + && P(OE | LG) \cdot P(LG) \\ &= \frac{3}{5} && \cdot && \frac{5}{8} && + && \frac{1}{3} && \cdot && \frac{3}{8} \\ &= && \frac{1}{2} \end{aligned}$$

$$\begin{aligned} P(LE) &= P(LE | LG) \cdot P(LG) && + && P(LE | OG) \cdot P(OG) \\ &= \frac{2}{3} && \cdot && \frac{3}{8} && + && \frac{2}{5} && \cdot && \frac{5}{8} \\ &= && \frac{1}{2} \end{aligned}$$

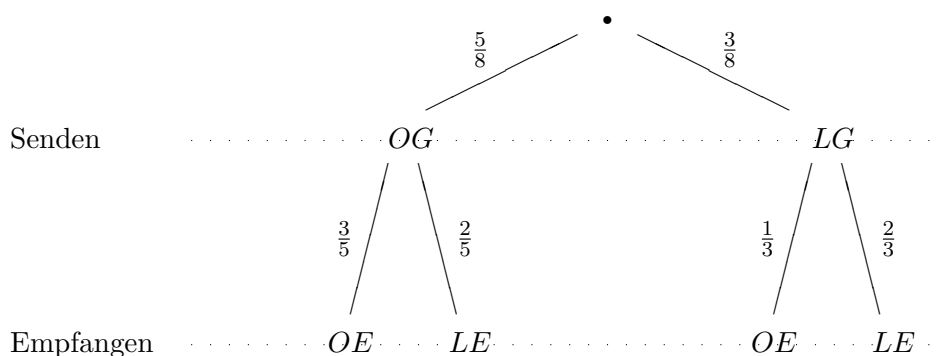


Abbildung 3.6: Senden und Empfangen von binären Informationen

Zur Berechnung der Wahrscheinlichkeiten  $P(OE)$  und  $P(LE)$  kann auch ein Baumdiagramm verwendet werden, s. Abb. 3.6

Damit erhält man die zu berechnenden Wahrscheinlichkeiten mit Hilfe der Formel von Bayes (3.6) zu

$$P(OG | OE) = \frac{P(OE | OG) \cdot P(OG)}{P(OE)} = \frac{\frac{3}{5} \cdot \frac{5}{8}}{\frac{1}{2}} = \frac{3}{4}$$

$$P(LG | LE) = \frac{P(LE | LG) \cdot P(LG)}{P(LE)} = \frac{\frac{2}{3} \cdot \frac{3}{8}}{\frac{1}{2}} = \frac{1}{2}$$

### 3.3.4 Stochastisch unabhängige Ereignisse

Seien  $(\Omega, P)$  ein diskreter Wahrscheinlichkeitsraum und  $A, B \in \mathcal{P}(\Omega)$  zwei Ereignisse mit  $P(A), P(B) > 0$ .

Im letzten Abschnitt wurde deutlich, daß die Wahrscheinlichkeit des Ereignisses  $A$  noch vom Eintreten des Ereignisses  $B$  abhängen kann; das führte zum Begriff der bedingten Wahrscheinlichkeit

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad (\text{falls } P(B) > 0).$$

In vielen Fällen jedoch ist die Wahrscheinlichkeit für das Eintreten des Ereignisses  $A$  völlig unabhängig davon, ob das Ereignis  $B$  eintritt oder nicht.

#### Definition 3.5

Zwei Ereignisse  $A$  und  $B$  eines Ereignisraumes heißen *stochastisch unabhängig*, wenn gilt:

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



**Bemerkung**

- (i) Nach dem Multiplikationssatz (3.4) gilt

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

also hätte man anstelle der Definitionsgleichung der (stochastischen) Unabhängigkeit auch  $P(B | A) = P(B)$  fordern können: Das Erfülltsein einer dieser beiden Beziehungen liefert auch die jeweils andere.

- (ii) Zwei Ereignisse sind genau dann unabhängig, wenn gilt:

$$P(A \cap B) = P(A) \cdot P(B). \quad (3.7)$$

Das folgt wieder direkt aus dem Multiplikationssatz und der Definition der Unabhängigkeit:

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

Umgekehrt folgt aus dieser Produktformel auch die Unabhängigkeit:

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

falls  $P(B) \neq 0$  ist. Daher wird die stochastische Unabhängigkeit häufig über diese Produktformel definiert.

**Aufgabe**

Man zeige, daß zwei unvereinbare Ereignisse  $A$  und  $B$  mit  $P(A), P(B) > 0$  nicht stochastisch unabhängig sind.

**Beispiel 3.21**

Von 100 Modulen weisen 30 einen Defekt  $a$  und 20 einen Defekt  $b$  auf; 6 weisen beide Defekte auf.

- (i) Mit welcher Wahrscheinlichkeit hat ein zufällig entnommenes Modul, das den Defekt  $a$  aufweist, auch den Defekt  $b$ ?
- (ii) Mit welcher Wahrscheinlichkeit hat ein zufällig entnommenes Modul, das den Defekt  $b$  aufweist, auch den Defekt  $a$ ?

**Lösung**

$$P(b | a) = \frac{P(b \cap a)}{P(a)} = \frac{6/100}{30/100} = \frac{1}{5} = \frac{20}{100} = P(b),$$

$$P(a | b) = \frac{P(a \cap b)}{P(b)} = \frac{6/100}{20/100} = \frac{3}{10} = \frac{30}{100} = P(a),$$

somit ist das Vorkommen beider Defekte unabhängig. Man beachte auch das Erfülltsein der Produktformel (3.7) für die Wahrscheinlichkeiten.

### Beispiel 3.22

Im letzten Semester bearbeiteten 10 % der Umwelttechnik-Studenten einer Berliner Hochschule regelmäßig ihre Mathematik-Hausaufgaben. Nach der Abschlußklausur wurde festgestellt, daß von 25 Teilnehmern, die durchgefallen sind, nur eine(r) regelmäßig Hausaufgaben angefertigt hat.

- (i) Hatte das Anfertigen der Hausaufgaben einen wesentlichen Einfluß auf das Bestehen der Klausur?
- (ii) Wenn man zusätzlich weiß, daß von 15 Studenten, die regelmäßig ihre Hausaufgaben gemacht haben, nur eine(r) durchgefallen ist, mit welcher Wahrscheinlichkeit fällt dann jemand durch, der nicht regelmäßig Hausaufgaben angefertigt hat?

### Lösung

$H$  : $\iff$  StudentIn hat regelmäßig Hausaufgaben angefertigt

$\bar{K}$  : $\iff$  StudentIn hat die Klausur nicht bestanden

- (i) Als Wahrscheinlichkeiten werden die relativen Häufigkeiten der Ereignisse genommen, d.h.  $P(H) = \frac{1}{10}$  und  $P(H | \bar{K}) = \frac{1}{25}$ . Wie man sieht, sind die Ereignisse  $H$  und  $\bar{K}$  *nicht* unabhängig:

$$\begin{aligned} P(H \cap \bar{K}) &= P(H) \cdot P(\bar{K} | H) = \frac{1}{10} \cdot P(\bar{K} | H) \\ P(\bar{K} \cap H) &= P(\bar{K}) \cdot P(H | \bar{K}) = \frac{1}{25} \cdot P(\bar{K}) \\ \longrightarrow P(\bar{K} | H) &\neq P(\bar{K}) \end{aligned}$$

Das heißt, daß das Anfertigen von Hausaufgaben wenigstens einen Einfluß auf das Bestehen der Klausur hatte. Welchen?

- (ii) Zusätzlich sei nun bekannt:  $P(\bar{K} | H) = \frac{1}{15} = 6, \bar{6} \%$ .

$$P(\bar{K}) = \frac{P(\bar{K} | H) \cdot P(H)}{P(H | \bar{K})} = \frac{\frac{1}{15} \cdot \frac{1}{10}}{\frac{1}{25}} = \frac{1}{6}$$

ist die Wahrscheinlichkeit, die Klausur nicht zu bestehen. Die Wahrscheinlichkeit, die Klausur nicht zu bestehen, wenn man nicht regelmäßig Hausaufgaben angefertigt hat, ist dann gegeben durch

$$P(\bar{K} | \bar{H}) = \frac{P(\bar{K} \cap \bar{H})}{P(\bar{H})} = \frac{P(\bar{H} | \bar{K}) \cdot P(\bar{K})}{P(\bar{H})} = \frac{(1 - \frac{1}{25}) \cdot \frac{1}{6}}{1 - \frac{1}{10}} = \frac{8}{45} \approx 17, \bar{7} \%$$

Die Wahrscheinlichkeit, die Klausur nicht zu bestehen, ist also für einen „Ungeübten“ mit fast 18 % zwischen zweieinhalb und dreimal so groß wie für einen „Geübten“ mit nur 6,6 %.

**Satz 3.3**

Seien  $A$  und  $B$  zwei unabhängige Ereignisse in  $\mathcal{P}(\Omega)$ . Dann sind die Ereignisse  $A, B, \bar{A}, \bar{B}$  unabhängig.

*Beweis*

Hausaufgabe

**Beispiel 3.23** (Zuverlässigkeit von Systemen)

Technische Systeme sind nie perfekt, und die Wahrscheinlichkeit für einen Ausfall ist nie gleich Null. Seien  $S_1, \dots, S_n$  die Komponenten eines Systems  $S$ . Der Ausfall der Komponenten  $S_i$  geschehe unabhängig voneinander mit der Wahrscheinlichkeit  $p_i$ . Dabei werden unterschieden:

- **Parallele Systeme**

Hier bewirkt erst der Ausfall aller Komponenten den Ausfall des Gesamtsystems, s. Abb. 3.7.

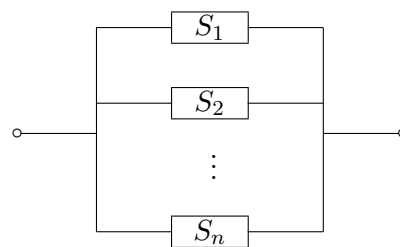


Abbildung 3.7: Paralleles System

- **Serielle Systeme**

Hier bewirkt schon der Ausfall mindestens einer Komponente den Ausfall des Gesamtsystems, s. Abb. 3.8.

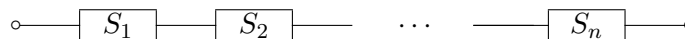


Abbildung 3.8: Seriell System

Seien

$A_i$  : $\iff$  Die  $i$ -te Komponente fällt aus

$A_S$  : $\iff$  Das Gesamtsystem fällt aus

Wegen der Unabhängigkeit des Ausfalles der Komponenten gilt für parallele Systeme die Produktformel, und man erhält:

$$P(A_S) = P(A_1 \cap A_2 \cap \dots \cap A_n) = \prod_{i=1}^n P(A_i) = \prod_{i=1}^n p_i.$$

Für serielle Systeme dagegen gilt

$$P(A_S) = P(A_1 \cup A_2 \cup \dots \cup A_n).$$

Die Ereignisse  $A_i$  sind nicht disjunkt ( $\sim$  nicht unvereinbar), denn sie können ja auch gleichzeitig eintreten. Daher darf man *nicht*

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i)$$

schreiben, sondern muß dazu den komplizierten Additionssatz (Satz 3.1 (iii)) für  $n$  Elemente anwenden; vgl. auch die Bemerkung auf p. 32.

Folgende einfache Überlegung reduziert die Rechnung allerdings erheblich: Da die  $A_1, \dots, A_n$  unabhängig sind, sind auch die komplementären Ereignisse  $\bar{A}_1, \dots, \bar{A}_n$  unabhängig. Daher gilt

$$\begin{aligned} P(A_S) &= 1 - P(\bar{A}_S) \\ &= 1 - P(\overline{A_1 \cup A_2 \cup \dots \cup A_n}) \\ &= 1 - P(\bar{A}_1 \cap \dots \cap \bar{A}_n) \\ &= 1 - \prod_{i=1}^n P(\bar{A}_i) \\ &= 1 - \prod_{i=1}^n (1 - p_i) \end{aligned}$$

Durch Kombination von Serien- und Parallelschaltung lassen sich kompliziertere Gesamtsysteme mit anderen Ausfallwahrscheinlichkeiten erzeugen.

### Aufgabe

Man bestimme die Ausfallwahrscheinlichkeit des Systems aus Abb. 3.9.

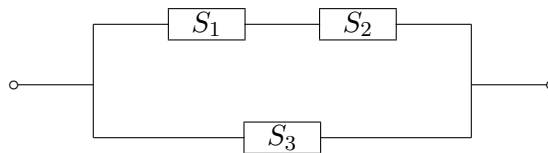


Abbildung 3.9: System aus drei Bauteilen

### Beispiel 3.24

- (i) Sechs Bäume stehen nebeneinander in einer geraden Linie, zwei von ihnen haben Schädlinge. Jeder Baum kann dabei mit der gleichen Wahrscheinlichkeit von Schädlingen befallen sein. Wie groß ist die Wahrscheinlichkeit, daß die beiden kranken Bäume nebeneinander stehen?

**Lösung**

Der Merkmalraum  $\Omega$  besteht aus allen Paaren unterschiedlicher Bäume, also

$$\Omega = \{\langle i, j \rangle : i, j \in \{1, \dots, 6\}, i < j\}.$$

Von solchen Paaren gibt es  $\binom{6}{2} = 15$  Stück (Anzahl der Möglichkeiten, aus sechs Elementen zwei auszuwählen). Fünf dieser Paare repräsentieren benachbarte Bäume:

$$S := \{\langle 1, 2 \rangle, \langle 2, 3 \rangle, \langle 3, 4 \rangle, \langle 4, 5 \rangle, \langle 5, 6 \rangle\}.$$

Sei  $KB : \iff$  Kranke Bäume sind benachbart.

Dann folgt

$$P(KB) = \frac{|S|}{|\Omega|} = \frac{5}{15} = 33,3\%.$$

Da zwei Bäume krank sind, geht es hier nur darum, die relative Anzahl benachbarter Bäume in bezug auf die Gesamtzahl festzustellen.

- (ii) Wie groß ist die Wahrscheinlichkeit, daß die beiden kranken Bäume nebeneinander stehen unter der Voraussetzung, daß Baum 3 einer der beiden kranken Bäume ist?

**Lösung**

Sei  $B_3 : \iff$  Baum 3 ist krank.

Dann ist  $KB \cap B_3 = \{\langle 2, 3 \rangle, \langle 3, 4 \rangle\}$ , und es folgt

$$P(KB | B_3) = \frac{P(KB \cap B_3)}{P(B_3)} = \frac{2/15}{5/15} = 40\%.$$

Natürlich läßt sich diese Aufgabe auch durch Konstruktion des reduzierten Ergebnisraumes lösen, und man gelangt zum selben Ergebnis:

$$\begin{aligned} \Omega^* &:= \{\langle 1, 3 \rangle, \langle 2, 3 \rangle, \langle 3, 4 \rangle, \langle 3, 5 \rangle, \langle 3, 6 \rangle\} \\ S^* &:= \{\langle 2, 3 \rangle, \langle 3, 4 \rangle\} \\ \longrightarrow P^*(B_3) &= \frac{|S^*|}{|\Omega^*|} = \frac{2}{5} = 40\%. \end{aligned}$$

- (iii) Wie groß ist die Wahrscheinlichkeit in (ii), falls die Bäume nicht in einer geraden Linie stehen, sondern im Kreis gepflanzt sind?

**Lösung**

Jetzt hat das Ereignis  $S$  ein Element mehr, nämlich das Element  $\langle 1, 6 \rangle$ : Baum 1 und Baum 6 sind ebenfalls benachbart. Dann ergibt sich beide Male

$$\begin{aligned} P(KB) &= \frac{6}{15} = 40\%, \\ P(KB | B_3) &= \frac{2/15}{5/15} = 40\%. \end{aligned}$$

**Diskussion**

- (a) Stehen die Bäume in einer geraden Linie, so wird die Wahrscheinlichkeit dafür, daß kranke Bäume benachbart sind, beeinflusst durch die Zusatzinformation, daß Baum 3 krank ist:

$$P(KB) = \frac{1}{3} \neq \frac{2}{5} = P(KB | B_3).$$

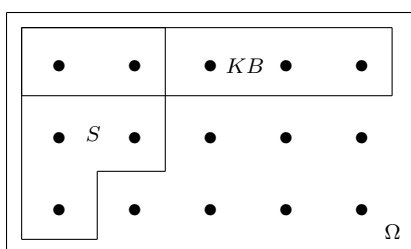
Die Ereignisse  $KB$  und  $B_3$  sind abhängig.

- (b) Stehen die Bäume dagegen im Kreis, so liefert die Aussage, daß Baum 3 krank ist, keine die Wahrscheinlichkeiten beeinflussende Zusatzinformation:

$$P(KB) = \frac{2}{5} = P(KB | B_3),$$

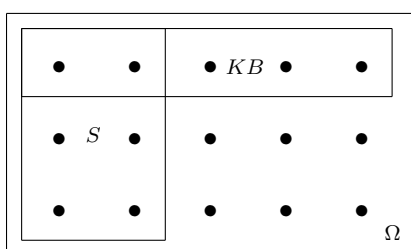
die Ereignisse  $KB$  und  $B$  sind hier unabhängig.

- (c) Man studiere in diesem Zusammenhang Abb. 3.10 und 3.11.



$$\begin{aligned} |\Omega| &= 15 \\ |KB| &= 5 \\ |S| &= 5 \\ |KB \cap S| &= 2 \\ \frac{|KB \cap S|}{|S|} &= \frac{2}{5} \neq \frac{5}{15} \end{aligned}$$

Abbildung 3.10: Die Ereignisse sind nicht unabhängig



$$\begin{aligned} |\Omega| &= 15 \\ |KB| &= 5 \\ |S| &= 6 \\ |KB \cap S| &= 2 \\ \frac{|KB \cap S|}{|S|} &= \frac{2}{6} = \frac{5}{15} \end{aligned}$$

Abbildung 3.11: Die Ereignisse sind jetzt unabhängig

- (d) Demzufolge liefert in (ii) – also bei Bäumen in einer geraden Linie – die Zusatzinformation, daß ein Baum am Rand, etwa Baum 1, einer der beiden kranken Bäume ist, dann eine andere Wahrscheinlichkeit:

Sei  $B_1 : \iff$  Baum 1 ist krank.

Dann ist  $KB \cap B_1 = \{(1, 2)\}$ , und es folgt

$$P(KB \cap B_1) = \frac{P(KB \cap B_1)}{P(B_1)} = \frac{1/15}{5/15} = 20\%.$$

Man beachte:

$$\sum_{i=1}^6 P(B_i) = \frac{1}{6} (0, 2 + 0, 4 + 0, 4 + 0, 4 + 0, 4 + 0, 2) = 33, \bar{3} \%$$

### 3.3.4.1 Stochastische Unabhängigkeit von mehr als zwei Ereignissen

Der Begriff der stochastischen Unabhängigkeit läßt sich auch auf mehr als zwei Ereignisse übertragen.

#### Definition 3.6

- (i) Seien  $n \in \mathbb{N}$  und  $A_1, \dots, A_n \subseteq \Omega$  Ereignisse. Diese Ereignisse heißen *stochastisch unabhängig*, wenn für *jede* Teilmenge

$$A_{i_1}, \dots, A_{i_k} \quad (2 \leq k \leq n)$$

die Produktregel gilt:

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot \dots \cdot P(A_{i_k}).$$

- (ii) Eine (unendliche) Familie von Ereignissen  $A_1, A_2, \dots \subseteq \Omega$  heißt *stochastisch unabhängig*, wenn je endlich viele der Ereignisse stochastisch unabhängig sind.

Es genügt also nicht, lediglich die paarweise Unabhängigkeit aller Ereignisse zu fordern! Siehe dazu das Gegenbeispiel im Anschluß an diese Bemerkung sowie eine Hausaufgabe.

#### Beispiel 3.25

In einer Urne befinden sich vier Kugeln mit den Beschriftungen 110, 101, 011, 000. Es wird eine Kugel zufällig entnommen.  $A_i$  sei das Ereignis, daß die gezogene Kugel an der  $i$ -ten Stelle eine 1 trägt ( $i \in \{1, 2, 3\}$ ). Dann ist

$$P(A_1 \cap A_2 \cap A_3) = 0 \neq \frac{1}{8} = \underbrace{P(A_1)}_{1/2} \cdot \underbrace{P(A_2)}_{1/2} \cdot \underbrace{P(A_3)}_{1/2},$$

jedoch gilt

$$\begin{aligned} P(A_1 \cap A_2) &= \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1) \cdot P(A_2), \\ P(A_2 \cap A_3) &= \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_2) \cdot P(A_3), \\ P(A_1 \cap A_3) &= \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1) \cdot P(A_3). \end{aligned}$$

# Kapitel 4

## Zufallsvariable

### 4.1 Einführung und Definition

Häufig ist man nicht an den Details eines jeden Elementarereignisses interessiert sondern nur an einer numerischen Beschreibung seines jeweiligen Auftretens.

#### Beispiel 4.1

Beim dreimaligen Wurf einer Münze besteht der Merkmalraum  $\Omega$  (die Ergebnismenge) aus acht Elementarereignissen ( $K$ : Kopf,  $W$ : Wappen):

$$\Omega := \{ \langle K, K, K \rangle, \langle K, K, W \rangle, \langle K, W, K \rangle, \langle W, K, K \rangle, \\ \langle K, W, W \rangle, \langle W, K, W \rangle, \langle W, W, K \rangle, \langle W, W, W \rangle \}.$$

Nun sei man lediglich an der *Anzahl* der auftretenden Wappen interessiert und nicht etwa daran, wann oder in welcher Reihenfolge diese auftreten. Man ordnet also jedem Elementarereignis  $\omega_i$  einen der Zahlenwerte 0, 1, 2, 3 zu, s. Tab. 4.1.

Anstatt mit den Ereignissen  $\omega_i \in \Omega$  selbst zu rechnen, benutzt man die Zahlen  $x_i \in \{0, 1, 2, 3\}$ , also die Werte, welche die Funktion

$$X : \omega \longmapsto \text{Anzahl der auftretenden Wappen}$$

auf der Definitionsmenge  $\Omega$  annimmt.

#### Definition 4.1

Eine Abbildung

$$X : \Omega \longrightarrow \mathbb{R} \\ \omega \longmapsto X(\omega)$$

heißt eine (eindimensionale) reelle *Zufallsvariable*.



Ereignis $\omega_i$	$P(\omega_i)$	$x_i$	$P(„X = x_i“)$
$\langle K, K, K \rangle$	$\frac{1}{8}$	0	$\frac{1}{8}$
$\langle K, K, W \rangle$ $\langle K, W, K \rangle$ $\langle W, K, K \rangle$	$\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$	1	$\frac{3}{8}$
$\langle K, W, W \rangle$ $\langle W, K, W \rangle$ $\langle W, W, K \rangle$	$\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$	2	$\frac{3}{8}$
$\langle W, W, W \rangle$	$\frac{1}{8}$	3	$\frac{1}{8}$

Tabelle 4.1: Anzahl der auftretenden Wappen bei dreimaligem Münzwurf

**Bemerkung**

- (i) Die Abbildung  $X$  ist also eine Funktion, die jedem Elementarereignis  $\omega$  genau eine reelle Zahl zuordnet. *Welchen* der Werte  $x = X(\omega)$  von  $X$  – diese heißen auch *Merkmalswerte* oder *Realisierungen* von  $X$  – man in der weiteren Betrachtung benutzt, hängt vom jeweiligen Versuch ab und ist vom Zufall bestimmt; daher der Name „Zufallsvariable“.
- (ii) Strenggenommen ist nicht jede auf  $\Omega$  definierte reellwertige Funktion eine Zufallsvariable. Eine Zufallsvariable muß noch einer Bedingung genügen, die in allen praktisch interessierenden Fällen zwar automatisch erfüllt ist, die aber für den strengen Aufbau der Theorie zusätzlich gefordert werden muß.

Für jede Zufallsvariable  $X$  muß die Menge

$$A_x := \{\omega : X(\omega) \leq x\} \quad (x \in \mathbb{R})$$

definiert sein. M.a.W., die Menge  $A_x$  muß für jedes  $x \in \mathbb{R}$  ein Ereignis sein:

$$x \in \mathbb{R} \implies A_x \in \mathcal{P}(\Omega).$$

Eine Funktion  $X$ , für die alle Mengen der Form  $A_x = \{\omega : X(\omega) \leq x\}$  zum Definitionsbereich des Wahrscheinlichkeitsmaßes  $P$  gehören, nennt man eine *meßbare Funktion*. Demnach sind Zufallsvariablen genau die meßbaren Funktionen auf  $\Omega$ .

(iii) In dem Einführungsbeispiel 4.1

$$X \quad : \Longleftrightarrow \quad \text{Anzahl der auftretenden Wappen}$$

ist beispielsweise  $W(X) = \{0, 1, 2, 3\}$ , und welchen der vier möglichen Merkmalwerte die Abbildung  $X$  annimmt, ist vom Zufall bestimmt.

### Definition 4.2

- (i) Die Zufallsvariable heißt *diskret*, wenn sie nur endlich viele oder höchstens abzählbar unendlich viele Werte  $x_i$  hat.
- (ii)  $X$  sei eine diskrete Zufallsvariable. Die Funktion

$$\begin{aligned} f_X : \quad \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto f_X(x), \end{aligned}$$

wobei

$$f_X(x) := \begin{cases} p_i & , \quad x = x_i \in W(X) \\ 0 & , \quad x \notin W(X) \end{cases} ,$$

heißt *Wahrscheinlichkeitsfunktion*, (*diskrete*) *Dichtefunktion* oder *Zähl-dichte* der Zufallsvariable  $X$ .

### Wiederaufgriff von Beispiel 4.1

Man betrachte die Zufallsvariable

$$\begin{aligned} X : \quad \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto X(\omega), \end{aligned}$$

definiert durch

$$\begin{aligned} X(\langle K, K, K \rangle) &= 0, \\ X(\langle K, K, W \rangle) &= X(\langle K, W, K \rangle) = X(\langle W, K, K \rangle) = 1, \\ X(\langle K, W, W \rangle) &= X(\langle W, K, W \rangle) = X(\langle W, W, K \rangle) = 2, \\ X(\langle W, W, W \rangle) &= 3. \end{aligned}$$

Die zugehörige Dichtefunktion  $f_X$  gibt die relative Häufigkeit an, mit der eine Zahl des Wertebereiches der Zufallsvariablen auftritt, s. Abb. 4.1.

Gewöhnlich wird die folgende laxe Schreibweise benutzt: für  $x = x_i \in W(X)$  ist

$$f_X(x_i) = p_i = P(\{\omega : \omega \in \Omega, X(\omega) = x_i\}) =: P(X = x_i).$$

Diese Schreibweise ist insofern nicht korrekt, als die „Gleichung“  $X = x_i$  formal keinen Sinn ergibt. Sie wird aber weithin verwandt, und daher soll das auch hier geschehen. Sei

$$A_i := \{\omega : \omega \in \Omega, X(\omega) = x_i\},$$

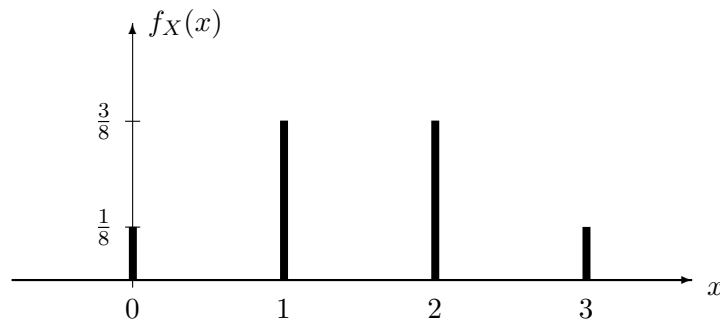
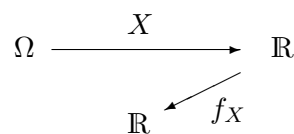
Abbildung 4.1: Dichtefunktion der Zufallsvariable  $X$  aus Beispiel 4.1

Abbildung 4.2: Zufallsvariable und Dichtefunktion

dann ist  $P(X = x_i) = P(A_i)$ , und die Zufallsvariable  $X$  wird vollständig beschrieben durch alle Ereignisse  $A_i$  für  $x_i \in W(X)$ .

In dem Beispiel sind etwa

$$\begin{aligned}
 A_0 &= \{\langle K, K, K \rangle\}, \\
 A_1 &= \{\langle K, K, W \rangle, \langle K, W, K \rangle, \langle W, K, K \rangle\}, \\
 A_2 &= \{\langle K, W, W \rangle, \langle W, K, W \rangle, \langle W, W, K \rangle\}, \\
 A_3 &= \{\langle W, W, W \rangle\}.
 \end{aligned}$$

Die acht verschiedenen Elementarereignisse werden anhand des untersuchten Merkmals (Anzahl der Wappen) zu vier verschiedenen Gruppen zusammengefaßt. Man rechnet dann mit einem neuen Wahrscheinlichkeitsraum, und zwar mit

$$\langle \Omega', P' \rangle := \langle W(X), P' \rangle,$$

wobei  $P'$  durch die Festlegung

$$P'(\{0\}) := \frac{1}{8}, \quad P'(\{1\}) := \frac{3}{8}, \quad P'(\{2\}) := \frac{3}{8}, \quad P'(\{3\}) := \frac{1}{8}$$

auf den neuen Elementarereignissen, also den Elementen aus  $W(X)$ , definiert ist. Im Gegensatz zu  $\langle \Omega, P \rangle$  ist  $\langle W(X), P' \rangle$  ein Wahrscheinlichkeitsraum, in welchem nur mit Zahlen gearbeitet wird.

#### 4.1.1 Physikalische Interpretation der Dichtefunktion

In jedem Elementarereignis  $\omega \in \Omega$  sei eine physikalische Masse angebracht, so daß die Gesamtmasse 1 ergibt. In Beispiel 4.1 etwa ist

$$m_i = \frac{1}{8} \quad (i \in \{1, \dots, 8\}) \quad , \quad M := \sum_{i=1}^8 m_i = 1.$$

Bei dieser Veranschaulichung eines Wahrscheinlichkeitsmaßes als der einer Aufteilung einer physikalischen Masseneinheit in Teilmassen bedeutet  $P(A)$  dann die Masse des „Bereichs“  $A$ .

Auf diese physikalische Veranschaulichung wird später noch eingehender eingegangen werden, und zwar bei der Einführung der Begriffe „Erwartungswert“ und „Varianz“ als Momente einer Wahrscheinlichkeitsverteilung.

## 4.2 Wahrscheinlichkeitsverteilung einer Zufallsvariable

### 4.2.1 Verteilung einer diskreten Zufallsvariable

Neben der Dichtefunktion  $f$  sollen jetzt zwei neue Funktionen eingeführt werden, und zwar die *Verteilung*  $P_X$  und die *Verteilungsfunktion*  $F_X$  einer Zufallsvariable.

$X$  sei eine diskrete Zufallsvariable mit abzählbarem Wertebereich

$$W(X) = \{x_1, x_2, \dots\}.$$

Die *Anfangswahrscheinlichkeiten*

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

seien bekannt. Die Abbildung

$$\begin{array}{lll} P_X : & \mathcal{P}[W(X)] & \longrightarrow [0, 1] \\ & A & \longmapsto P_X(A) \end{array}$$

sei definiert durch

$$P_X(A) := \sum_{\{i: x_i \in A\}} p_i.$$

#### Satz 4.1

$P_X$  ist ein Wahrscheinlichkeitsmaß auf der Potenzmenge  $\mathcal{P}[W(X)]$ .

*Beweis*

- (i) Die Potenzmenge einer jeden Menge ist nicht leer, also gilt  $\mathcal{P}[W(X)] \neq \emptyset$ . Man beachte, daß  $\mathcal{P}[W(X)]$  jedoch nicht notwendig mehr abzählbar sein muß (abzählbar viele Folgen).<sup>1</sup>
- (ii) (1)  $P_X(A) \geq 0$  für alle  $A \subseteq W(X)$ .

<sup>1</sup>Die abzählbare Vereinigung abzählbar unendlich vieler Mengen ist abzählbar, nicht jedoch die Potenzmenge einer abzählbar unendlichen Menge.

(2) Das Maß der Gesamtmenge ist 1:

$$\begin{aligned}
 P_X[W(X)] &= \sum_{\{i: x_i \in W(X)\}} p_i \\
 &= \sum_{\{i: x_i \in W(X)\}} P(\{\omega : X(\omega) = x_i\}) \\
 &= \sum_{\{i: x_i \in W(X)\}} \sum_{\{j: X(\omega_j) = x_i\}} P(\omega_j) \\
 &= \sum_{\{\omega_j: \omega_j \in \Omega\}} P(\{\omega_j\}) \\
 &= 1
 \end{aligned}$$

(3) Additivität:

$$\begin{aligned}
 P_X\left(\bigcup_i A_i\right) &= \sum_{\{j: x_j \in \bigcup_i A_i\}} p_j \\
 &= \sum_i \sum_{\{j: x_j \in A_i\}} p_j \\
 &= \sum_i P_X(A_i),
 \end{aligned}$$

falls die  $A_i$  paarweise disjunkt sind. Man beachte, daß alle auftretenden Reihen absolut konvergent sind, da sie aus nichtnegativen Gliedern bestehen.

### Definition 4.3

(i) Das durch die Folge  $\{p_i\}_{i \in \mathbb{N}}$  definierte Wahrscheinlichkeitsmaß

$$\begin{aligned}
 P_X : \quad \mathcal{P}[W(X)] &\longrightarrow [0, 1] \\
 A &\longmapsto \sum_{\{i: x_i \in A\}} f_X(x_i)
 \end{aligned}$$

$$\left[ \begin{array}{l} f_X \quad : \quad \text{Dichtefunktion} \\ f_X(x) \quad := \quad \begin{cases} p_i & , \quad x = x_i \in W(X) \\ 0 & , \quad x \notin W(X) \end{cases} \end{array} \right]$$

heißt die *Verteilung* der Zufallsvariable  $X$ .

(ii) Die *Verteilungsfunktion*  $F_X$  von  $X$  wird definiert durch

$$\begin{aligned}
 F_X : \quad \mathbb{R} &\longrightarrow \mathbb{R} \\
 x &\longmapsto P(X \leq x),
 \end{aligned}$$

wobei

$$P(X \leq x) := P_X(\{x_i : x_i \in W(X), x_i \leq x\}) = \sum_{\{i: x_i \leq x\}} f_X(x_i)$$

**Beispiel 4.2**

Ich kehre noch einmal zu Beispiel 4.1 des dreimaligen Münzwurfes zurück. Die einzelnen Elementarereignisse  $\omega_i$  sind zwar gleichwahrscheinlich, die Zufallsvariable

$$X \quad : \Longleftrightarrow \quad \text{Anzahl der auftretenden Wappen}$$

nimmt jedoch eine Gruppierung dieser Elementarereignisse derart vor, daß die auftretenden Gruppen eine unterschiedliche Anzahl von Elementen haben, also keine „Gleichverteilung“ der Gruppen vorliegt.

(i) Verteilung  $P_X$  (s. dazu auch das Liniendiagramm in Abb. 4.1)

$$A_0 = \{\langle K, K, K \rangle\}$$

$$A_1 = \{\langle K, K, W \rangle, \langle K, W, K \rangle, \langle W, K, K \rangle\}$$

$$A_2 = \{\langle K, W, W \rangle, \langle W, K, W \rangle, \langle W, W, K \rangle\}$$

$$A_3 = \{\langle W, W, W \rangle\}$$

$$P(A_0) = P_X(\{0\}) = f_X(0) = \frac{1}{8}$$

$$P(A_1) = P_X(\{1\}) = f_X(1) = \frac{3}{8}$$

$$P(A_2) = P_X(\{2\}) = f_X(2) = \frac{3}{8}$$

$$P(A_3) = P_X(\{3\}) = f_X(3) = \frac{1}{8}$$

(ii) Verteilungsfunktion  $F_X$  (s. dazu auch die Treppenfunktion in Abb. 4.3)

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & , \quad x < 0 \\ \frac{1}{8} & , \quad 0 \leq x < 1 \\ \frac{4}{8} & , \quad 1 \leq x < 2 \\ \frac{7}{8} & , \quad 2 \leq x < 3 \\ 1 & , \quad 3 \leq x \end{cases}$$

Beispielsweise ist

$$\begin{aligned} F_X(2,5) &= P(X \leq 2,5) = P_X(\{0,1,2\}) \\ &= \sum_{\{i: x_i \in \{0,1,2\}\}} p_i = p_0 + p_1 + p_2 \\ &= \frac{1}{8} + \frac{3}{8} + \frac{3}{8} \\ &= \frac{7}{8} \end{aligned}$$

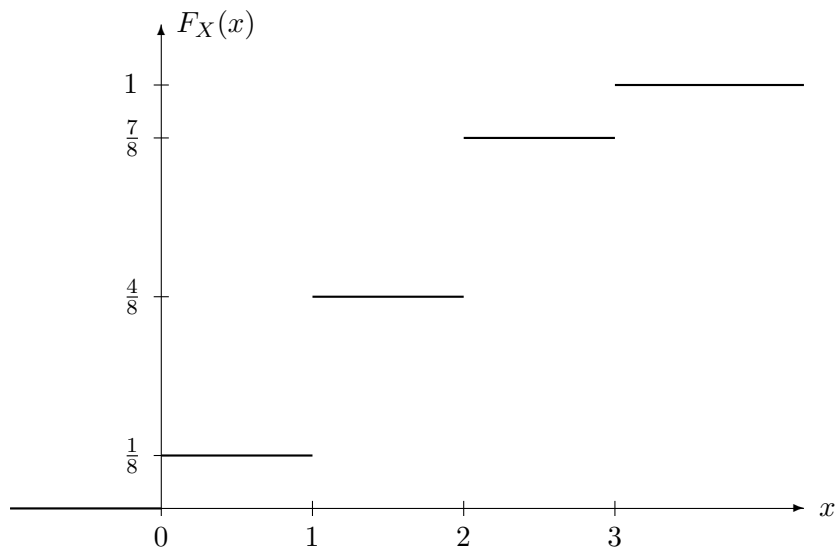


Abbildung 4.3: Verteilungsfunktion der Zufallsvariable  $X$  aus Beispiel 4.1

### Bemerkung

Die Verteilungsfunktion  $F_X$  ist eine monoton steigende Treppenfunktion. Sie sammelt gewissermaßen diejenigen Wahrscheinlichkeiten auf, welche sich beim Durchlauf von  $-\infty$  bis zu  $x$  einschließlich „ansammeln“, und summiert diese. Dabei wird angenommen, daß die Wahrscheinlichkeiten  $p_i$  bekannt sind. Es geht also jetzt nicht darum, wie man diese findet, sondern wie man mit ihnen rechnet, wie man die dadurch gegebene Verteilung darstellt und was man aus der Kenntnis dieser Verteilung schließen kann.

### Beispiel 4.3

Beim Zufallsexperiment „Wurf mit zwei unterscheidbaren homogenen Würfeln“ besteht der Merkmalraum  $\Omega$  aus den insgesamt 36 gleichwahrscheinlichen Elementarereignissen (geordneten Augenpaaren)

$$\langle 1, 1 \rangle, \langle 1, 2 \rangle, \langle 1, 3 \rangle, \dots, \langle 5, 6 \rangle, \langle 6, 6 \rangle.$$

Betrachtet wird die diskrete Zufallsvariable

$$X \quad :\Leftrightarrow \quad \text{Erreichte Augensumme,}$$

welche die möglichen Zahlenwerte  $2, 3, \dots, 11, 12$  mit unterschiedlichen Wahrscheinlichkeiten annimmt.

Im folgenden werden dazu bestimmt:

- (i) eine Häufigkeitstabelle (Tab. 4.2),
- (ii) eine Verteilungstabelle (Tab. 4.3),
- (iii) das Liniendiagramm (Abb. 4.4) und

$x_i$	2	3	4	5	6	7	8	9	10	11	12
	(1,1)	(1,2) (2,1)	(1,3) (2,2) (3,1)	(1,4) (2,3) (3,2) (4,1)	(1,5) (2,4) (3,3) (4,2) (5,1)	(1,6) (2,5) (3,4) (4,3) (5,2) (6,1)	(2,6) (3,5) (4,4) (5,3) (6,2)	(3,6) (4,5) (5,4) (6,3)	(4,6) (5,5) (6,4)	(5,6) (6,5)	(6,6)
$n_i$	1	2	3	4	5	6	5	4	3	2	1

Tabelle 4.2: Häufigkeitstabelle bei einem Wurf mit zwei unterschiedlichen Würfeln

$x_i$	2	3	4	5	6	7	8	9	10	11	12
$f(x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Tabelle 4.3: Verteilungstabelle bei einem Wurf mit zwei unterschiedlichen Würfeln

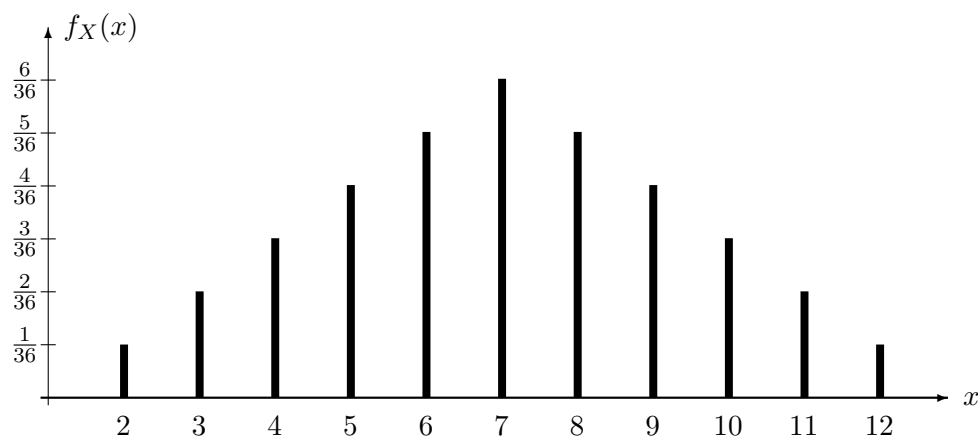


Abbildung 4.4: Dichtefunktion der Zufallsvariable „Augensumme von zwei Würfeln“



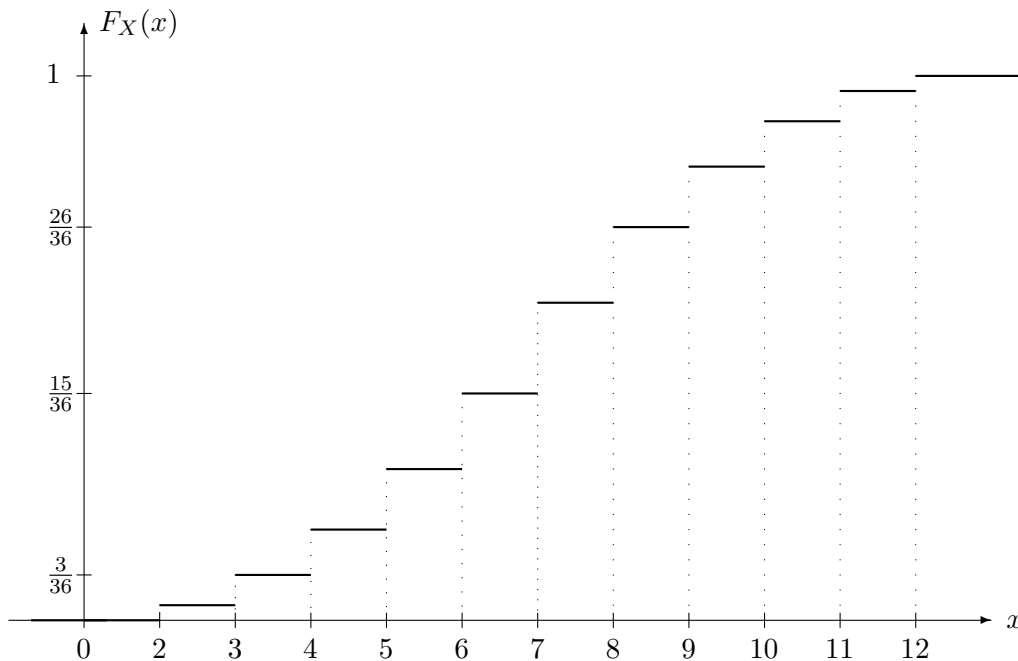


Abbildung 4.5: Verteilungsfunktion der Zufallsvariable „Augensumme von zwei Würfeln“

(iv) die Verteilungsfunktion (Abb. 4.5).

#### Beispiel 4.4

Eine Urne enthalte fünf Kugeln, davon drei weiße und zwei schwarze. Aus ihr werde dreimal zufällig eine Kugel entnommen und die gezogene Kugel vor einem nächsten Zug in die Urne zurückgelegt. Daher sind alle Ziehungen voneinander unabhängig. Die Zufallsvariable

$X \iff$  Anzahl der erhaltenen schwarzen Kugeln  
bei drei Ziehungen mit Zurücklegen

kann die möglichen Zahlenwerte 0, 1, 2 und 3 annehmen. Die acht möglichen Elementarereignisse sind gegeben durch

$$\Omega := \{ \langle W, W, W \rangle, \langle S, W, W \rangle, \langle W, S, W \rangle, \langle W, W, S \rangle, \langle S, S, W \rangle, \langle S, W, S \rangle, \langle W, S, S \rangle, \langle S, S, S \rangle \},$$

und sie treten mit unterschiedlichen Wahrscheinlichkeiten auf.

Im folgenden werden dazu bestimmt:

- (i) eine Wahrscheinlichkeitstabelle (Tab. 4.4),
- (ii) eine Verteilungstabelle (Tab. 4.5),

Anzahl der gezogenen schwarzen Kugeln	Elementarereignis	Wahrscheinlichkeit
0	$\langle W, W, W \rangle$	$\frac{3}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} = \frac{27}{125}$
1	$\langle S, W, W \rangle$ $\langle W, S, W \rangle$ $\langle W, W, S \rangle$	$\left. \vphantom{\begin{matrix} \langle S, W, W \rangle \\ \langle W, S, W \rangle \\ \langle W, W, S \rangle \end{matrix}} \right\} \text{jeweils } \frac{2}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} = \frac{18}{125}$
2	$\langle S, S, W \rangle$ $\langle S, W, S \rangle$ $\langle W, S, S \rangle$	$\left. \vphantom{\begin{matrix} \langle S, S, W \rangle \\ \langle S, W, S \rangle \\ \langle W, S, S \rangle \end{matrix}} \right\} \text{jeweils } \frac{2}{5} \cdot \frac{2}{5} \cdot \frac{3}{5} = \frac{12}{125}$
3	$\langle S, S, S \rangle$	$\frac{2}{5} \cdot \frac{2}{5} \cdot \frac{2}{5} = \frac{8}{125}$

Tabelle 4.4: Wahrscheinlichkeitstabelle der Zufallsvariable „Anzahl der erhaltenen schwarzen Kugeln bei drei Ziehungen mit Zurücklegen“

$x_i$	0	1	2	3
$f(x_i)$	$\frac{27}{125}$	$3 \cdot \frac{18}{125} = \frac{54}{125}$	$3 \cdot \frac{12}{125} = \frac{36}{125}$	$\frac{8}{125}$

Tabelle 4.5: Verteilungstabelle der Zufallsvariable „Anzahl der erhaltenen schwarzen Kugeln bei drei Ziehungen mit Zurücklegen“

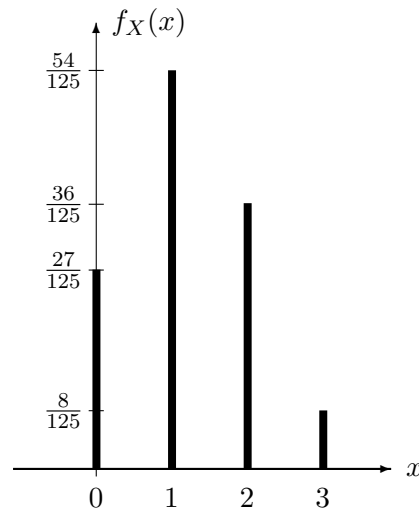


Abbildung 4.6: Dichtefunktion der Zufallsvariable „Anzahl der erhaltenen schwarzen Kugeln bei drei Ziehungen mit Zurücklegen“

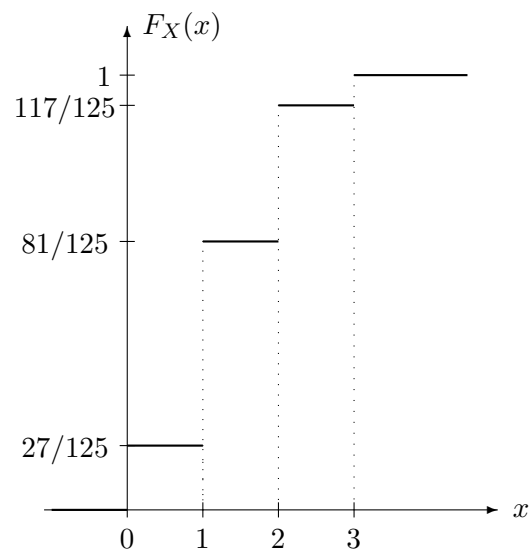


Abbildung 4.7: Verteilungsfunktion der Zufallsvariable „Anzahl der erhaltenen schwarzen Kugeln bei drei Ziehungen mit Zurücklegen“

- (iii) das Liniendiagramm (Abb. 4.6) und
- (iv) die Verteilungsfunktion (Abb. 4.7).

### Bemerkung

Bisher wurde stets vorausgesetzt, daß die Anfangswahrscheinlichkeiten  $p_i$  vorliegen bzw. man sie anhand leichter Überlegungen selbst bestimmen kann, so wie etwa bei Laplace-Experimenten:

- Münzwurf,
- Würfeln mit zwei Würfeln,
- Ziehen von Kugeln aus Urnen.

Es ging also nicht darum, wie man diese *findet*, sondern wie man mit ihnen *rechnet*.

Häufig gelangt man jedoch erst durch die Untersuchung einer Zufallsvariable zu einer Möglichkeit, diese Anfangswahrscheinlichkeiten selbst festzulegen. Erhält man beispielsweise eine Formel für die  $p_i$ , in der nur ein oder mehrere von  $i$  unabhängige Parameter unbekannt bleiben, so ist das Problem der Zuweisung von Wahrscheinlichkeiten im wesentlichen auf eine Schätzung dieser Parameter zurückgeführt.

### Beispiel 4.5<sup>2</sup>

Unabhängig vom Wochentag und der genauen Uhrzeit seien am späten Vormittag alle Arbeitsplätze in einem Rechnerraum mit der Wahrscheinlichkeit  $p$  ( $p \in (0, 1)$  geeignet) belegt. Ein Student  $S$  versucht maximal  $n$ -mal, einen freien Platz zu bekommen. Scheitern alle Versuche, so gibt er frustriert auf und schimpft auf die Rechnerausstattung.

### Gesucht

- (i) die Wahrscheinlichkeit, erstmals im  $k$ -ten Versuch einen freien Platz zu finden ( $1 \leq k \leq n$ );
- (ii) die Wahrscheinlichkeit, bei  $n$  Versuchen keinen freien Platz zu finden.

### Lösung

$$\Omega := \{\omega_0, \omega_1, \dots, \omega_n\}, \quad \text{wobei}$$

$\omega_k$  : $\iff$   $S$  findet erstmals im  $k$ -ten Versuch einen freien Platz ( $1 \leq k \leq n$ )

$\omega_0$  : $\iff$   $S$  findet bei  $n$  Versuchen keinen freien Platz

---

<sup>2</sup>Greiner/Tinhofer: Stochastik für Studienanfänger der Informatik, Hanser-Verlag, 1. Auflage 1996, p. 64

$\omega_k$  kann durch die Zahl  $k$  charakterisiert werden, d.h. es läßt sich eine Zufallsvariable  $X$  wählen, welche die Anzahl der Versuche bis zum Erfolg mißt ( $\omega_0$ : kein Erfolg):

$$\begin{array}{ll} X : \Omega & \longrightarrow \{0, 1, \dots, n\} & \text{Anzahl der Versuche bis zum Erfolg} \\ \omega_k & \longmapsto X(\omega_k) := k & (k \in \{0, \dots, n\}). \end{array}$$

Sei  $A_k$  das Ereignis

$$A_k \quad :\Longleftrightarrow \quad S \text{ findet beim } k\text{-ten Versuch keinen Platz,}$$

dann gelten

$$\begin{aligned} \{\omega_1\} &= \bar{A}_1, \\ \{\omega_k\} &= A_1 \cap A_2 \cap \dots \cap \bar{A}_k, & (2 \leq k \leq n) \\ \{\omega_0\} &= A_1 \cap A_2 \cap \dots \cap A_n. \end{aligned}$$

Damit findet man für die gesuchten Wahrscheinlichkeiten wegen der Unabhängigkeit der einzelnen Ereignisse  $\omega_k$  die folgenden Beziehungen:

$$\begin{aligned} P(X = k) &= P(\{\omega : X(\omega) = k\}) \\ &= P(\omega_k) \\ &= P(A_1 \cap A_2 \cap \dots \cap A_{k-1} \cap \bar{A}_k) \\ &= \prod_{i=1}^{k-1} P(A_i) \cdot P(\bar{A}_k) \\ &= p^{k-1} \cdot (1 - p) \end{aligned}$$

$$\begin{aligned} P(X = 0) &= P(\{\omega : X(\omega) = 0\}) \\ &= P(\omega_0) \\ &= P(A_1 \cap A_2 \cap \dots \cap A_n) \\ &= \prod_{i=1}^n P(A_i) \\ &= p^n \end{aligned}$$

Damit kennt man die Werte der Zufallsvariable  $X$  (ihre „Verteilung“), sobald man den Wert des einzelnen Parameters  $p$  kennt.

**Probe**

Die Wahrscheinlichkeit, in höchstens  $k$  Versuchen einen freien Platz zu bekommen, ist wegen der angenommenen Unabhängigkeit der Versuche gegeben durch

$$\begin{aligned}
 P(X \leq k) &= \sum_{i=1}^k P(X = i) \\
 &= \sum_{i=1}^k p^{i-1}(1-p) \\
 &= (1-p) \sum_{i=0}^{k-1} p^i \\
 &= (1-p) \frac{1-p^k}{1-p} \\
 &= 1-p^k,
 \end{aligned}$$

wie nicht anders zu erwarten.

**Aufgabe**

- (i) Wieviele Versuche muß ein Student mindestens durchführen, um mit einer Wahrscheinlichkeit von mindestens 50 % einen freien Rechnerplatz zu bekommen, falls  $p = 0.9$  gilt?
- (ii) Wie groß darf  $p$  höchstens sein, damit bei höchstens 5 Versuchen maximal 1 % der Studenten keinen freien Platz finden?

**Lösung**

- (i) Er/sie muß mindestens 7 Versuche durchführen.
- (ii) Die Wahrscheinlichkeit  $p$ , daß alle Plätze des Rechnerraumes belegt sind, darf den Wert 40 % nicht überschreiten.

**Bemerkung**

Im Hinblick auf die im nächsten Abschnitt zu besprechende stetige oder kontinuierliche Zufallsvariable sollen die Eigenschaften einer diskreten Zufallsvariable  $X$  noch einmal übersichtsartig zusammengefaßt werden:

- (i) Eigenschaften der Dichtefunktion  $f_X$ :

$$(a) \quad f_X(x) \geq 0 \quad (x \in D(f_X))$$

$$(b) \quad \sum_{x_i \in W(X)} f_X(x_i) = 1$$

(c)  $f_X(x) = P(\{\omega : X(\omega) = x\}) =: P(X = x)$

(ii) Eigenschaften der Verteilungsfunktion  $F_X$ :

$F_X$  ist eine monoton wachsende Treppenfunktion mit

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad , \quad \lim_{x \rightarrow +\infty} F_X(x) = 1.$$

(iii) Die Wahrscheinlichkeit dafür, daß die (diskrete) Zufallsvariable  $X$  einen Wert zwischen  $a$  (ausschließlich) und  $b$  (einschließlich) annimmt, berechnet sich zu

$$\begin{aligned} P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F_X(b) - F_X(a). \end{aligned}$$

(iv) Die Bedingung (i) (b)

$$1 = P(\Omega) = \sum_{x_i \in W(X)} f_X(x_i) = \sum_{x_i \in W(X)} f_X(x_i) \cdot \Delta x_i = 1$$

mit  $\Delta x_i = 1$  für alle  $i$  läßt sich anhand eines *Histogrammes* erläutern, s. Abb. 4.8.

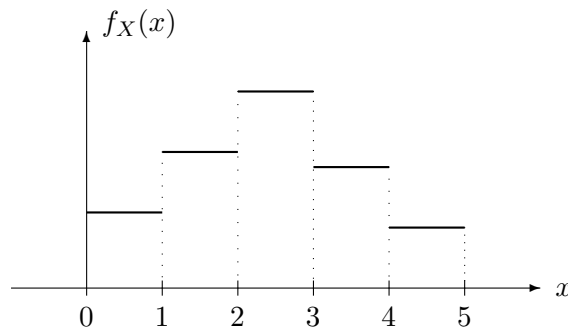


Abbildung 4.8: Histogramm einer diskreten Zufallsvariable  $X$

In einem zu Abb. 4.8 gehörigen Beispiel ist  $W(X) = \{0, 1, 2, 3, 4\}$ , und

$$F_X(x) = P(X \leq x) = \sum_{\substack{x_i \in W(X) \\ x_i \leq x}} f_X(x_i) \cdot \Delta x_i$$

liefert den Wert der Fläche zwischen der Dichtefunktion und der  $x$ -Achse im Intervall  $(-\infty, x]$ . Im stetigen Fall wird daraus ein Integral:

$$F(x) = \sum_{\substack{x_i \in W(X) \\ x_i \leq x}} f_X(x_i) \cdot \Delta x_i \quad \xrightarrow{(|W(X)| \rightarrow \infty)} \quad \int_{-\infty}^x f_X(u) du.$$

### 4.2.2 Verteilung einer stetigen Zufallsvariable

#### Definition 4.4

Eine Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  heißt *stetig* oder *kontinuierlich*, wenn gilt:

- (i)  $W(X)$  ist ein Kontinuum, also z.B. ein beschränktes oder unbeschränktes Intervall.
- (ii) Es existiert eine nicht-negative integrierbare Funktion  $f_X$ , so daß für die (unter diesen Bedingungen dann stetige) Verteilungsfunktion  $F_X$  die folgende Darstellung gilt:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(u) du \quad (x \in (-\infty, \infty)).$$

Die Funktion  $f_X$  heißt *Dichtefunktion* oder kurz *Dichte* der Verteilung von  $X$ ; s. Abb. 4.9 und 4.10.

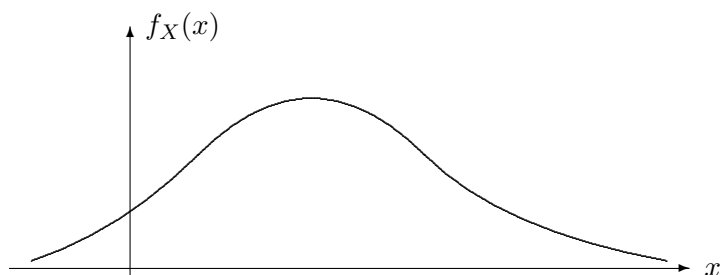


Abbildung 4.9: Dichtefunktion einer stetigen Zufallsvariable  $X$

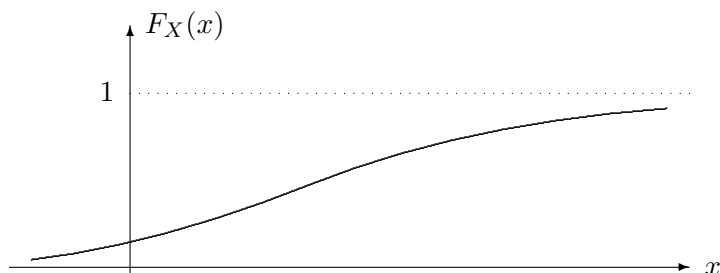


Abbildung 4.10: Verteilungsfunktion einer stetigen Zufallsvariable  $X$

#### Beispiel 4.6

Die Zufallsvariable  $X$  beschreibe die Wahrscheinlichkeit des Reißens eines Gummibandes bei dessen Streckung um  $x$  cm. Diese Zufallsvariable ist dann etwa



durch die folgende lineare Dichtefunktion beschreibbar, s. Abb. 4.11:

$$f_X(x) := \begin{cases} \frac{1}{8}x & , \quad 0 \leq x \leq 4 \\ 0 & , \quad \text{sonst} \end{cases}$$

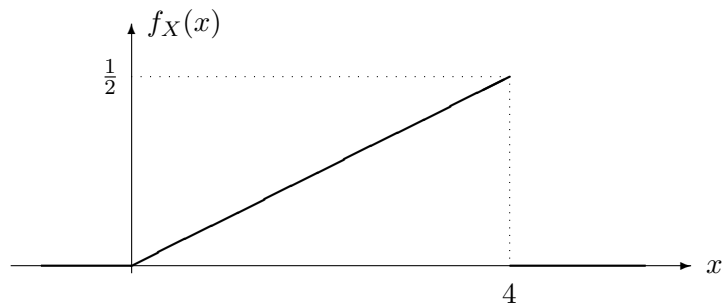


Abbildung 4.11: Lineare Dichtefunktion einer stetigen Zufallsvariable  $X$

Wegen

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_0^4 \frac{1}{8}x dx = \left[ \frac{x^2}{16} \right]_0^4 = 1$$

handelt es sich bei  $f_X$  tatsächlich um eine Dichtefunktion.

Jetzt soll die zugehörige Verteilungsfunktion im Intervall  $[0, 4]$  bestimmt werden:

$$F_X(x) = \int_{-\infty}^x f_X(u) du = \frac{1}{8} \cdot \int_0^x u du = \frac{1}{8} \left[ \frac{1}{2}u^2 \right]_0^x = \frac{1}{16}x^2.$$

Sie verläuft dort parabelförmig. Somit gilt, s. Abb. 4.12:

$$F_X(x) := \begin{cases} 0 & , \quad x < 0 \\ \frac{1}{16}x^2 & , \quad 0 \leq x \leq 4 \\ 1 & , \quad 4 < x \end{cases}$$

### Bemerkung

Obwohl die Zufallsvariable  $X$  das Reißen eines Gummibandes mit Hilfe der Dichtefunktion  $f_X$  adäquat beschreibt, bedeutet der Wert von  $f_X(x)$  an einer Stelle  $x$  nicht die Wahrscheinlichkeit, daß das Gummiband an dieser Stelle mit der Wahrscheinlichkeit  $f_X(x)$  reißt. Der Sachverhalt ist etwas komplizierter, vgl. dazu die Bemerkung (v) auf p. 80.

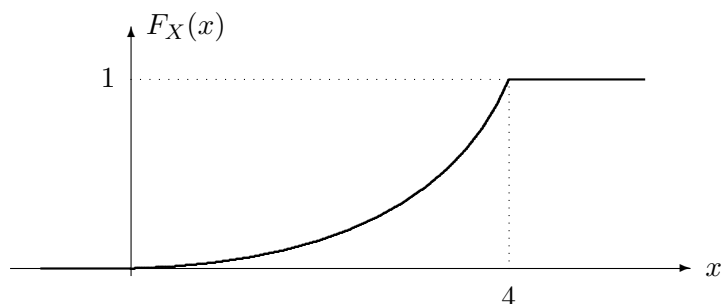


Abbildung 4.12: Parabelförmige Verteilungsfunktion einer linearen Dichtefunktion einer stetigen Zufallsvariable  $X$

### Beispiel 4.7

Die Lebensdauer  $T$  eines elektronischen Bauteiles, welches nicht verschleißbedingt sondern „zufällig“ ausfällt, kann oft als eine *exponentialverteilte* Zufallsgröße mit der Dichtefunktion

$$f(t) := \begin{cases} 0 & , t < 0 \\ a \cdot e^{-\lambda t} & , t \geq 0 \end{cases}$$

angenommen werden, wobei  $\lambda, a > 0$  positive Konstanten sind.

- (i) Wie groß ist die Konstante  $a$ ?
- (ii) Wie lautet die zugehörige Verteilungsfunktion?
- (iii) Wie groß ist der Anteil an Bauelementen, deren Lebensdauer den Wert  $t = 10$  überschreitet?

### Lösung

- (i) Die Konstante  $a$  bestimmt sich aus der *Normierungsbedingung*  $\int_{-\infty}^{\infty} f(t) dt = 1$ :

$$\int_{-\infty}^{\infty} f(t) dt = \int_0^{\infty} a \cdot e^{-\lambda t} dt = a \left[ -\frac{1}{\lambda} \cdot e^{-\lambda t} \right]_0^{\infty} = \frac{a}{\lambda} \stackrel{!}{=} 1.$$

Somit ist  $a = \lambda$ , und die Dichtefunktion besitzt die Gestalt wie in Abb. 4.13 skizziert.

- (ii) Für Werte  $t \geq 0$  erhält man damit die Verteilungsfunktion  $F$  zu

$$\begin{aligned} F(t) &= \int_{-\infty}^t f(u) du = \int_0^t \lambda \cdot e^{-\lambda u} du \\ &= \left[ -e^{-\lambda u} \right]_0^t = 1 - e^{-\lambda t}. \end{aligned}$$

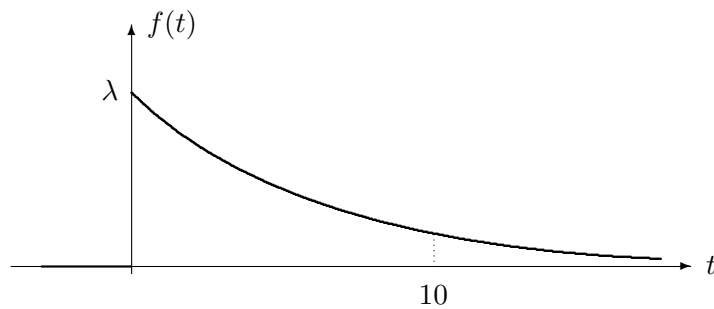


Abbildung 4.13: Exponentialfunktion als Dichtefunktion der Zufallsvariable  $T$

Daher gilt, s. auch Abb. 4.14:

$$F(t) := \begin{cases} 0 & , t < 0 \\ 1 - e^{-\lambda t} & , t \geq 0 \end{cases}$$

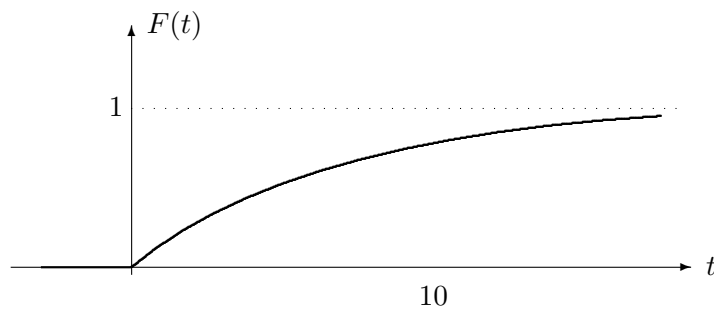


Abbildung 4.14: Verteilungsfunktion für die Zufallsvariable  $T$

- (iii) Die gesuchte Wahrscheinlichkeit entspricht der mit  $P(T > 10)$  gekennzeichneten Fläche in Abb. 4.15, sie wird wie folgt berechnet:

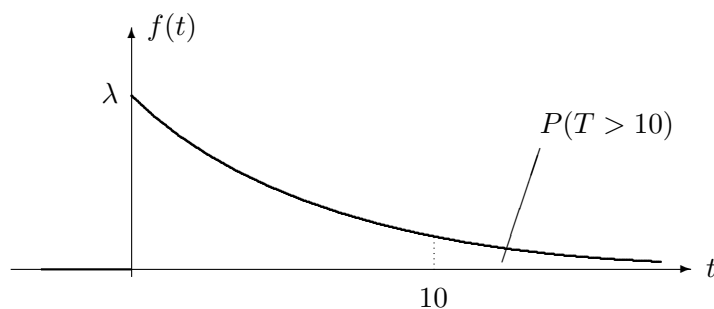


Abbildung 4.15: Exponentialfunktion als Dichtefunktion der Zufallsvariable  $T$

$$\begin{aligned} P(T > 10) &= 1 - P(T \leq 10) = 1 - F(10) \\ &= 1 - (1 - e^{-10\lambda}) = e^{-10\lambda} \end{aligned}$$

Für  $\lambda = 0.1$  beispielsweise ist  $e^{-10\lambda} = e^{-1} \approx 0.368$ , und es sind demnach rund 36,8 % der Bauteile zur Zeit  $t = 10$  noch funktionstüchtig.

### Bemerkung

- (i) Ist die Zufallsvariable  $T$  eine Lebensdauer, so pflegt man neben der *Ausfallwahrscheinlichkeit*

$$F(t) := P(T \leq t) \quad (t \in \mathbb{R})$$

auch die *Überlebenswahrscheinlichkeit* oder *Zuverlässigkeit*

$$Z(t) := P(T > t) = 1 - F(t) \quad (t \in \mathbb{R})$$

zu betrachten.  $Z$  ist eine stetige und monoton fallende Funktion mit

$$\lim_{t \rightarrow -\infty} Z(t) = 1 \quad , \quad \lim_{t \rightarrow \infty} Z(t) = 0.$$

- (ii) Die Größe

$$R(t) := -\frac{1}{Z(t)} \cdot \frac{dZ}{dt}(t) = \frac{f(t)}{Z(t)} \quad (t \in \mathbb{R})$$

bezeichnet man in diesem Zusammenhang als *Ausfallrate*.

- (iii) **Aufgabe**

Man bestimme die Zuverlässigkeit und Ausfallrate für die Exponentialverteilung mit dem Parameter  $\lambda$ .

### Bemerkung

Auch für stetige oder kontinuierliche Zufallsvariablen  $X$  sollen die definierenden Eigenschaften der Dichtefunktion  $f = f_X$  und ihrer Verteilungsfunktion  $F = F_X$  noch einmal übersichtsartig zusammengefaßt werden:

(i)  $f(x) \geq 0 \quad (x \in D(f))$

- (ii)  $F$  ist eine monoton wachsende stetige Funktion mit

(a)  $0 < F(x) < 1 \quad (x \in D(F))$

(b)  $\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad , \quad \lim_{x \rightarrow +\infty} F_X(x) = 1$

(c)  $\int_{-\infty}^{\infty} f(u) du = 1 = P(-\infty < X < \infty)$

- (iii) Ist die Dichtefunktion  $f$  stetig, so ist die Verteilungsfunktion  $F$  stetig differenzierbar, und es gilt nach dem Hauptsatz der Differential- und Integralrechnung  $F' = f$ . Wie man sieht, gilt in diesem Falle sogar  $F \in \mathcal{C}^1(\mathcal{I})$ .

- (iv) Für stetige Dichtefunktionen  $f$  ist  $F$  eine Stammfunktion, und die Wahrscheinlichkeit, daß die Zufallsvariable  $X$  einen Wert zwischen  $a$  und  $b$  annimmt, berechnet sich zu

$$\begin{aligned} P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a) = \int_a^b f(u) du. \end{aligned}$$

- (v) Die Wahrscheinlichkeit dafür, daß eine stetige Zufallsvariable  $X$  *genau* einen Wert annimmt, ist 0:

$$P(X = x) = P(\{\omega : X(\omega) = x\}) = 0.$$

Diese Eigenschaft folgt aus der Stetigkeit der Verteilungsfunktion  $F$ :

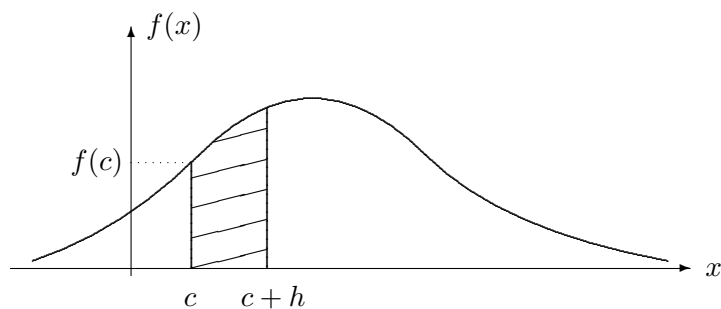
$$\begin{aligned} P(X \leq x) &= P(X = x) + P(X < x), \\ P(X \leq x) &= F(x) = \lim_{c \rightarrow 0} F(x - c) = \lim_{\substack{c \rightarrow 0 \\ c > 0}} F(x - c) \\ &= \lim_{\substack{c \rightarrow 0 \\ c > 0}} P(X \leq x - c) = P(X < x), \\ \longrightarrow & P(X = x) = 0. \end{aligned}$$

Für eine stetige Zufallsvariable hat also das Ereignis, daß  $X$  einen bestimmten Wert  $c$  annimmt, für jedes  $c \in \mathbb{R}$  die Wahrscheinlichkeit 0, auch wenn  $f(c) \neq 0$  gilt. Dennoch kann man  $f(c)$  mit dem Wert der Wahrscheinlichkeit in Verbindung bringen:

Sei  $h > 0$ . Dann gilt nach (iv) und dem Mittelwertsatz der Integralrechnung

$$\begin{aligned} P(c < X \leq c + h) &= F(c + h) - F(c) \\ &= \int_c^{c+h} f(u) du \\ &= f(\tilde{c}) \int_c^{c+h} du \quad (\tilde{c} \in (c, c + h) \text{ geeignet}) \\ &= f(\tilde{c}) \cdot h \\ &\approx f(c) \cdot h \end{aligned}$$

für kleine Werte von  $h$  und eine stetige Dichtefunktion  $f$ . Damit liegt  $X$  mit der Wahrscheinlichkeit  $\approx f(c) \cdot h$  im Intervall  $[c, c + h]$ , s. Abb. 4.16.

Abbildung 4.16: Dichtefunktion einer stetigen Zufallsvariable  $X$ 

- (vi) Für stetige Zufallsvariablen  $X$  ist zwar  $P(a < X \leq b) = P(a \leq X \leq b)$ , wegen der Rechtsstetigkeit der diskreten Verteilungsfunktion  $F$  verwendet man aber auch hier meistens den Ausdruck  $P(a < X \leq b)$ , also

$$P(a < X \leq b) = \int_a^b f_X(x) dx = F_X(b) - F_X(a).$$

*Eine Hand auf der heißen Herdplatte, einen Fuß im Eiswasser  
und im Durchschnitt fühlt man sich wohl.*

---

## Kapitel 5

# Kennwerte einer Wahrscheinlichkeitsverteilung

Die Wahrscheinlichkeitsverteilung einer diskreten oder stetigen Zufallsvariable  $X$  läßt sich eindeutig und vollständig durch folgende Größen beschreiben:

- die zugehörige Verteilungsfunktion  $F_X$  oder
- die zugehörige Dichtefunktion  $f_X$  oder
- bestimmte Parameter, welche man als *Kennwerte* oder *Maßzahlen* der Verteilung bezeichnet. Zu diesen Kennwerten zählen u.a.
  - der *Erwartungswert*  $\mu = E(X)$
  - die *Varianz*  $\sigma^2 = V(X)$
  - die *Standardabweichung*  $\sigma = \sqrt{V(X)}$

Sie sind wichtige Sonderfälle einer Gruppe von Kennwerten, die als *Momente* einer Verteilung bezeichnet werden. Der Erwartungswert ist das theoretische Gegenstück zum *Stichprobenmittel*, die Varianz ist das entsprechende Gegenstück zur *Stichprobenvarianz*.

### 5.1 Erwartungswert einer Zufallsvariable

Der Erwartungswert einer Zufallsvariablen kennzeichnet in gewisser Weise das Zentrum der Wahrscheinlichkeitsverteilung.

#### Beispiel 5.1

Einmaliger Wurf eines homogenen Würfels mit der Zufallsvariable

$$X \quad \Longleftrightarrow \quad \text{erzielte Augenzahl} \quad , \quad W(X) = \{1, 2, 3, 4, 5, 6\}$$

Die sechs möglichen Werte treten mit der gleichen Wahrscheinlichkeit

$$p_i = f(x_i) = \frac{1}{6} \quad (x_i \in W(X), i \in \{1, \dots, 6\})$$

auf, da es sich um ein Laplace-Experiment handelt; man sagt auch, sie seien *gleichverteilt*. Wenn nun das Experiment oft genug wiederholt wird, so wird man erwarten, daß die „mittlere Augenzahl“ (was ist das?) in der Nähe des arithmetischen Mittelwertes liegt:

$$\bar{x} := \sum_{i=1}^6 i \cdot \frac{1}{6} = \sum_{i=1}^6 x_i \cdot f(x_i) = 3,5.$$

Dieser Wert ist der sog. *Erwartungswert* der Zufallsvariablen  $X$  in dem Würfel-experiment. Man beachte, daß der Erwartungswert als Ergebnis keines einzigen Würfel-experimentes angenommen wird, i.a. auch nicht angenommen werden muß.

### Beispiel 5.2

Mittelwert der Noten einer Klausur. In diesem Beispiel sind die Werte der Zufallsvariable

$$X \quad \Longleftrightarrow \quad \text{erzielte Klausurnote}$$

$$W(X) = \{1.0, 1.3, 1.7, 2.0, 2.3, 2.7, 3.0, 3.3, 3.7, 4.0, 5.0\}$$

in der Regel nicht gleichverteilt. Bei der Berechnung des Erwartungswertes spielt daher die Dichtefunktion  $f_X$  eine entscheidende Rolle: sie bestimmt die Gewichtungsfaktoren, mit denen die möglichen Notenwerte  $x_i$  in die Berechnung eingehen:

$$f(x_i) = \frac{\text{Anzahl der Klausur-Studenten mit der Note } x_i}{\text{Anzahl aller Klausur-Studenten}}$$

Der (arithmetische) Mittelwert errechnet sich dann wieder zu ( $|W(X)| = 11$ )

$$\bar{x} := \sum_{i=1}^{11} x_i \cdot f(x_i).$$

### Definition 5.1

#### (i) diskreter Fall

Sei  $X$  eine diskrete Zufallsvariable mit den möglichen Werten  $x_k$ . Der *Erwartungswert* von  $X$  ist definiert durch

$$\mu := E(X) := \sum_k x_k \cdot P(X = x_k) = \sum_k x_k \cdot f(x_k).$$



Im endlichen Fall ist also ( $n \in \mathbb{N}$  geeignet)

$$E(X) = \sum_{k=1}^n x_k \cdot f(x_k),$$

im unendlichen Fall ist

$$E(X) = \sum_{k=1}^{\infty} x_k \cdot f(x_k) \quad \left( \text{oder auch z. B. } \sum_{k=0}^{\infty} x_k \cdot f(x_k) \right)$$

Im letzten Fall verlangt man die absolute Konvergenz der unendlichen Reihe, da es dann auf die Reihenfolge der Summanden nicht ankommt.

(ii) kontinuierlicher Fall

Sei  $X$  eine kontinuierliche Zufallsvariable mit der Dichtefunktion  $f$ . Der Erwartungswert von  $X$  ist definiert durch

$$\mu = E(X) := \int_{-\infty}^{\infty} x \cdot f(x) dx.$$

Auch hier wird die absolute Konvergenz des (uneigentlichen) Integrals vorausgesetzt, also die Existenz von

$$\int_{-\infty}^{\infty} |x \cdot f(x)| dx \stackrel{(f \geq 0)}{=} \int_{-\infty}^{\infty} |x| \cdot f(x) dx.$$

Ist das nicht der Fall, dann sagt man, die Verteilung besitze keinen Erwartungswert.

Der Erwartungswert einer Zufallsvariable ist somit die Summe (Integral) ihrer Werte mal der Wahrscheinlichkeit, daß ein einzelner Wert auftritt.

### Beispiel 5.3

Betrachtet wird noch einmal die zufallsbedingte Lebensdauer eines elektronischen Bauteiles, deren Dichtefunktion in guter Näherung durch eine Exponentialfunktion beschreibbar ist, s. Beispiel 4.7:

$$f(t) := \begin{cases} 0 & , t < 0 \\ \lambda \cdot e^{-\lambda t} & , t \geq 0 \end{cases} .$$

Ein Schätzwert für die zu erwartende Lebensdauer ist die durchschnittliche

(mittlere) Lebensdauer, welche durch den Erwartungswert  $E(T)$  definiert ist:

$$\begin{aligned} E(T) &= \int_{-\infty}^{\infty} t \cdot f(t) dt = \lambda \cdot \int_0^{\infty} t \cdot e^{-\lambda t} dt \\ &= \lambda \cdot \left( \left[ -\frac{t \cdot e^{-\lambda t}}{\lambda} \right]_0^{\infty} + \frac{1}{\lambda} \int_0^{\infty} e^{-\lambda t} dt \right) \\ &= \quad \quad \quad 0 \quad \quad - \frac{1}{\lambda} \left[ e^{-\lambda t} \right]_0^{\infty} \\ &= \frac{1}{\lambda} \end{aligned}$$

### Bemerkung

- (i) Benutzt man wieder die physikalische Interpretation, nach der in jedem Elementarereignis  $\omega_k$  eine Masse der Größe

$$m_k = P_X(\{x_k\}) = P(X = x_k) = f(x_k)$$

angebracht ist, so daß die Gesamtmasse 1 ergibt,

$$M := \sum_k m_k = 1,$$

dann ist der Erwartungswert gerade der *Schwerpunkt* der jeweiligen Verteilung:

$$E(X) = \sum_k x_k f(x_k) \stackrel{(M=1)}{=} \frac{1}{M} \sum_k x_k m_k =: x_S$$

bzw. im kontinuierlichen Fall

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \stackrel{(M=1)}{=} \frac{1}{M} \int_{-\infty}^{\infty} x \rho(x) dx =: x_S.$$

- (ii) Bei einer symmetrischen Dichtefunktion liegt der Erwartungswert an der Stelle ihres Maximums, s. Abb. 5.1.

Natürlich ist der Begriff des Erwartungswertes nur dann eine sinnvolle Größe, wenn die „Stichprobe“ groß und die Gewichtungsfaktoren  $f(x_k)$  nicht allzu extrem sind.

### Beispiel 5.4

Bei einem Spielautomaten gewinne man mit der Wahrscheinlichkeit

$$p = 1 - 10^{-6} \quad 2 \text{ Euro}$$

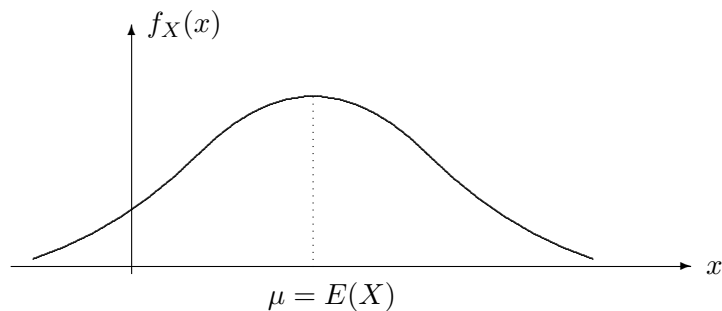


Abbildung 5.1: Kontinuierliche Verteilung mit symmetrischer Dichtefunktion

und verliere mit der Wahrscheinlichkeit

$$q := 1 - p = 10^{-6} \quad 10^6 \text{ Euro.}$$

Der Erwartungswert des Gewinnes (Verlust = negativer Gewinn) beträgt

$$E(G) = 2 \cdot (1 - 10^{-6}) + (-1) \cdot 10^6 \cdot 10^{-6} = 1 - 2 \cdot 10^{-6} > 0.$$

Es wird wohl nur wenige Spieler geben, die mit diesem Automaten ihr Glück versuchen wollen, obwohl der Erwartungswert des Gewinnes positiv ist.

### Bemerkung

- (i) Zeichnet man den Erwartungswert von Beispiel 5.4 als Schwerpunkt eines geeigneten Hebelarmes, so erhält man Zeichnung 5.2.

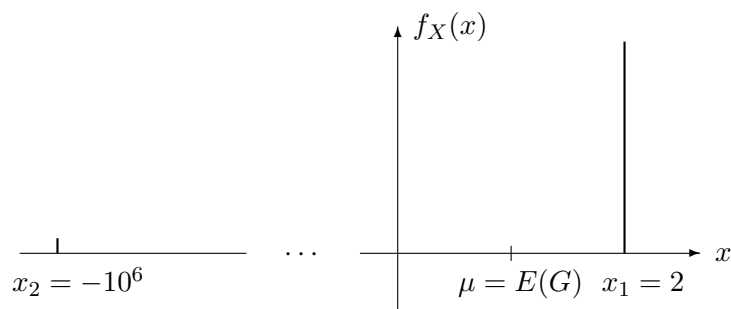


Abbildung 5.2: Glücksspielautomat als physikalischer Hebelarm

- (ii) Ein solches, mathematisch zwar korrektes, aber irgendwie doch unbefriedigendes Phänomen tritt immer dann auf, wenn man eine sehr kleine mit einer sehr großen Zahl multipliziert, um aus dem Produkt Rückschlüsse zu ziehen. Beispielsweise geschieht das bei der Risikoabschätzung von Unfällen. Das Risiko ist hierbei definiert als Produkt von Wahrscheinlichkeit, mit der ein Schaden eintritt und dem Ausmaß des Schadens.

$x_i$	2	5	10
$P(X = x_i)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

Tabelle 5.1: Verteilungstabelle der Zufallsvariable „Wegstrecke eines Fahrradkuriers“

## 5.2 Funktionen von Zufallsvariablen

Häufig treten Zufallsvariable  $Y$  auf, die vermöge einer gegebenen Funktion  $g$  aus einer anderen (i.a. leichter zugänglichen) Zufallsvariable hervorgehen:

$$Y = g(X).$$

Untersucht werden soll nun, welche Informationen über  $Y$  man aus der Wahrscheinlichkeitsverteilung von  $X$  erhalten kann.

### Beispiel 5.5

$X$  sei die Wegstrecke, die ein Fahrradkurier bei der Erledigung eines Transportauftrages zurücklegen muß. Zum Durchfahren einer Wegeinheit (etwa 1 km) werden 3 Zeiteinheiten (eine Zeiteinheit etwa 1 min) benötigt. Dann ist  $T := g(X) = 3X$  die dafür benötigte Zeit.

Nun sei  $X$  eine diskrete Zufallsvariable mit der Verteilung aus Tabelle 5.1;  $X$  kann also die drei diskreten Werte – etwa die betrachteten Standardwegstrecken  $x_1 = 2$  km,  $x_2 = 5$  km und  $x_3 = 10$  km – annehmen. Daraus resultiert die Verteilungsfunktion

$$F_X(x) := \begin{cases} 0 & , \quad x < 2 \\ \frac{1}{3} & , \quad 2 \leq x < 5 \\ \frac{5}{6} & , \quad 5 \leq x < 10 \\ 1 & , \quad 10 \leq x \end{cases} .$$

Mit Hilfe der einfachen Umformungen ( $i \in \{1, 2, 3\}$ )

$$P(X = x_i) = P(3X = 3x_i) = P(T = 3x_i) = P(T = t_i)$$

bzw.

$$F_T(t) = P(T < t) = P(3X < t) = P\left(X < \frac{t}{3}\right) = F_X\left(\frac{t}{3}\right)$$

$t_i$	6	15	30
$P(T = t_i)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

Tabelle 5.2: Verteilungstabelle der Zufallsvariable „Zeit eines Fahrradkuriers“

erhält man für die Zufallsvariable  $T = 3X$  eine Verteilung gemäß Tabelle 5.2 und die Verteilungsfunktion

$$F_T(t) := \begin{cases} 0 & , \quad t < 6 \\ \frac{1}{3} & , \quad 6 \leq t < 15 \\ \frac{5}{6} & , \quad 15 \leq t < 30 \\ 1 & , \quad 30 \leq t \end{cases} .$$

Gemäß Definition 5.1 (i) berechnen sich die Erwartungswerte von  $X$  und  $T$  zu

$$\begin{aligned} E(X) &= x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + x_3 \cdot P(X = x_3) \\ &= 2 \cdot \frac{1}{3} + 5 \cdot \frac{1}{2} + 10 \cdot \frac{1}{6} \\ &= \frac{29}{6} = 4,83 , \end{aligned}$$

$$\begin{aligned} E(T) &= t_1 \cdot P(T = t_1) + t_2 \cdot P(T = t_2) + t_3 \cdot P(T = t_3) \\ &= 6 \cdot P(T = 6) + 15 \cdot P(T = 15) + 30 \cdot P(T = 30) \\ &= 6 \cdot P(X = 2) + 15 \cdot P(X = 5) + 30 \cdot P(X = 10) \\ &= 6 \cdot \frac{1}{3} + 15 \cdot \frac{1}{2} + 30 \cdot \frac{1}{6} \\ &= \frac{87}{6} = 3 \cdot \frac{29}{6} = 3 \cdot 4.83 \\ &= 3 \cdot E(X) . \end{aligned}$$

Aus  $T = g(X) = 3 \cdot X$  folgt also  $E(T) = E[g(X)] = 3 \cdot E(X)$ , und es läßt sich die Gültigkeit des folgenden Satzes verifizieren:

### Satz 5.1

- (i) Seien  $X$  eine diskrete Zufallsvariable mit den Werten  $x_k$  und  $g$  eine reellwertige Funktion. Dann ist  $Y := g(X)$  eine diskrete Zufallsvariable mit

den möglichen Werten  $y_k := g(x_k)$ . Der Erwartungswert von  $Y$  ist gegeben durch

$$E(Y) = E[g(X)] = \sum_k g(x_k) \cdot P(X = x_k) = \sum_k g(x_k) \cdot f_X(x_k).$$

- (ii) Seien  $X$  eine stetige Zufallsvariable mit der Dichtefunktion  $f_X$  und  $g$  eine reellwertige Funktion. Dann ist  $Y := g(X)$  eine stetige Zufallsvariable mit der Dichtefunktion  $g(f_X)$ . Der Erwartungswert von  $Y$  ist gegeben durch

$$E(Y) = E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx.$$

### Satz 5.2

Sei  $X$  eine Zufallsvariable.

- (i) Ist  $X = \text{const.} = c$ , so gilt:

$$E(X) = c.$$

- (ii) Linearität des Erwartungswertes:

Sind  $g_1$  und  $g_2$  zwei von  $X$  abhängige Funktionen, so gilt für alle  $a, b \in \mathbb{R}$ :

$$E[a \cdot g_1(X) + b \cdot g_2(X)] = a \cdot E[g_1(X)] + b \cdot E[g_2(X)].$$

### Bemerkung

Die Linearitätsbeziehung gilt allgemein für Linearkombinationen von Funktionen *verschiedener* Zufallsvariablen. Dazu muß jedoch erst definiert werden, was man unter der Wahrscheinlichkeitsverteilung mehrerer Zufallsvariablen versteht, sog. *mehrdimensionaler Verteilungen*.

### Beispiel 5.6

Sei  $X$  eine diskrete Zufallsvariable mit der Wahrscheinlichkeitsverteilung wie in Tabelle 5.3 angegeben.

Welches ist der Erwartungswert von  $Y := g(X) := (X - 1)^2$ ?

### Lösung

- (i) ohne Satz 5.2

$$\begin{aligned} E(Y) &= E((X - 1)^2) = \sum_{k=0}^3 (x - 1)^2 \cdot f(x_k) \\ &= (-1)^2 \cdot f(0) + 0^2 \cdot f(1) + 1^2 \cdot f(2) + 2^2 \cdot f(3) \\ &= 1 \end{aligned}$$

$x_i$	0	1	2	3
$f(x_k)$	$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{6}$

Tabelle 5.3: Verteilungstabelle einer Zufallsvariable

(ii) mit Satz 5.2

$$\begin{aligned}
 E(Y) &= E((X-1)^2) = E(X^2 - 2X + 1) = E(X^2) - 2E(X) + E(1), \\
 E(X) &= 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{2} + 2 \cdot 0 + 3 \cdot \frac{1}{6} = 1 \\
 E(X^2) &= 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{2} + 2^2 \cdot 0 + 3^2 \cdot \frac{1}{6} = 2 \\
 E(1) &= 1 \\
 \longrightarrow E((X-1)^2) &= 2 - 2 \cdot 1 + 1 = 1
 \end{aligned}$$

**Beispiel 5.7**Sei  $X$  eine stetige Zufallsvariable mit der Dichtefunktion

$$f(x) := \begin{cases} \frac{x^2}{3} & , \quad -1 < x < 2 \\ 0 & , \quad \text{sonst} \end{cases} .$$

Welches ist der Erwartungswert von  $Y := g(X) := 2X - 1$ ?**Lösung**

(i) ohne Satz 5.2

$$E(Y) = E(2X - 1) = \int_{-1}^2 (2x - 1) \cdot \frac{x^2}{3} = \frac{3}{2}$$

(ii) mit Satz 5.2

$$\begin{aligned}
 E(X) &= \int_{-1}^2 x \cdot \frac{x^2}{3} = \frac{5}{4} \\
 \longrightarrow E(2X - 1) &= 2E(X) - 1 = 2 \cdot \frac{5}{4} - 1 = \frac{3}{2}
 \end{aligned}$$

### 5.3 Varianz und Standardabweichung

Mit dem Erwartungswert wird für die Verteilung einer Zufallsvariable ein gewisses Zentrum definiert, welches in der physikalischen Interpretation denn auch dem Schwerpunkt einer Massenverteilung entspricht; s. dazu die Bemerkung auf p. 86.

Mit einem weiteren Kennwert soll nun charakterisiert werden, wie stark die Werte der Zufallsvariable um dieses Zentrum, also um den Erwartungswert streuen.

**Definition 5.2**

Sei  $X$  eine Zufallsvariable mit dem Erwartungswert  $\mu = E(X)$ . Dann heißt

$$\sigma^2 := V(X) := E[(X - \mu)^2]$$

die *Varianz* oder *Dispersion* von  $X$ .

Die Wurzel  $\sigma = \sqrt{V(X)}$  der Varianz bezeichnet man als *Standardabweichung*.

**Bemerkung**

- (i) Die Varianz ist der Erwartungswert der Zufallsvariable

$$Z := g(X) := (X - \mu)^2,$$

welche die quadratische Abweichung der Zufallsvariable  $X$  von ihrem Erwartungswert  $\mu$  beschreibt. Daher gilt nach Satz 5.1

- (a) im diskreten Fall:

$$\sigma^2 = V(X) = E[(X - \mu)^2] = \sum_k (x_k - \mu)^2 f(x_k), \quad (5.1)$$

- (b) im kontinuierlichen Fall:

$$\sigma^2 = V(X) = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx. \quad (5.2)$$

- (ii) Die Varianz  $\sigma^2 = V(X)$  ist stets  $\geq 0$ . Sie ist ein geeignetes Maß für die *Streuung* der einzelnen Werte  $x_k$  um den Erwartungswert  $\mu$ . Bei kleiner Varianz liegen die „meisten“ Werte in der Nähe von  $\mu$ , und größere Abweichungen von  $\mu$  treten nur mit geringen Wahrscheinlichkeiten auf.

Die beiden Abbildungen 5.3 und 5.4 verdeutlichen diesen Sachverhalt für zwei Dichtefunktionen von kontinuierlichen Zufallsvariablen mit demselben Erwartungswert.



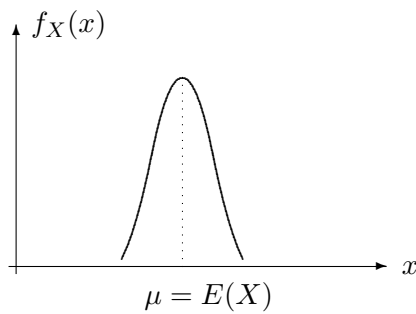


Abbildung 5.3: kleine Varianz

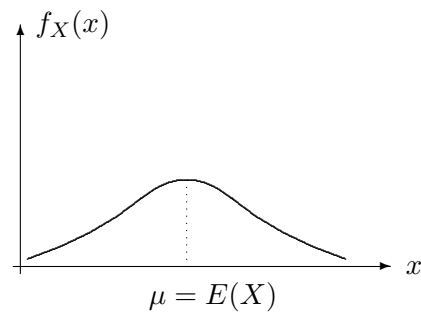


Abbildung 5.4: große Varianz

- (iii) Man kann auch die Standardabweichung  $\sigma$  als Streuungsmaß verwenden: sie beschreibt die Abweichung der Zufallsvariable  $X$  von ihrem Erwartungswert  $\mu$ . Gegenüber der Varianz hat sie den Vorteil, daß sie die gleiche Einheit hat wie die Zufallsvariable  $X$  selbst.
- (iv) Die der Zufallsvariable  $X$  zugeordneten Kennwerte  $\mu, \sigma, \sigma^2$  werden häufig auch als *Kennwerte der Verteilung* bezeichnet.

**Beispiel 5.8**

Es soll die Varianz für das Würfelbeispiel 5.1 berechnet werden. Man erhält mit  $\mu = E(X) = 3,5$  und Formel (5.1):

$$V(X) = E[(X - \mu)^2] = \sum_{k=1}^6 (k - 3,5)^2 \cdot \frac{1}{6} = 2,92$$

**Satz 5.3** (Verschiebungsregel für die Varianz)

$$V(X) = E(X^2) - [E(X)]^2 \quad (5.3)$$

*Beweis*

Aufgrund der Linearität des Erwartungswertes  $\mu = E(X)$  (Satz 5.2) erhält man

$$\begin{aligned} V(X) &= E((X - \mu)^2) \\ &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 E(1) \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

**Bemerkung**

Das physikalische Äquivalent zur Verschiebungsregel für die Varianz ist der *Trägheitsmomentensatz* oder *Satz von Steiner*:

Sind

$J$	:	Trägheitsmoment eines starren Körpers in bezug auf eine vorgeschriebene Achse
$J_S$	:	Trägheitsmoment in bezug auf die dazu parallele Achse durch den Schwerpunkt $S$
$d$	:	Abstand beider Achsen
$M$	:	Gesamtmasse des starren Körpers

so gilt:

$$J = J_S + M \cdot d^2.$$

**Folgerung**

Unter allen auf parallele Achsen bezogene Trägheitsmomente ist das auf die Schwerpunktsachse bezogene Trägheitsmoment am kleinsten.

Verbindung zur Verschiebungsregel für die Varianz:

$$\begin{aligned} V(X) &= E(X^2) - [E(X)]^2 \\ &\iff \\ \sum_k (x_k - d)^2 m_k &= \sum_k x_k^2 m_k - d^2 \\ \int_{-\infty}^{\infty} (x - d)^2 \rho(x) dx &= \int_{-\infty}^{\infty} x^2 \rho(x) dx - d^2 \end{aligned}$$

Man beachte:

$$M = \sum_k m_k = 1 \quad , \quad M = \int_{-\infty}^{\infty} \rho(x) dx = 1$$

**Aufgabe**

Man berechne die Varianzen für

Beispiel 4.3	(Lösung: $V(X) = 5,83$ )
Beispiel 4.4	(Lösung: $V(X) = 0,72$ )
Beispiel 4.6	(Lösung: $V(X) = 8/9$ )
Beispiel 4.7	(Lösung: $V(T) = 1/\lambda^2$ )

zur Übung jeweils ohne und mit Satz 5.3.

Der folgende Satz beleuchtet die Bedeutung der Varianz quantitativ.

**Satz 5.4** (Tschebyschev-Ungleichung)

Sei  $X$  eine beliebige (!) Zufallsvariable mit dem Erwartungswert  $\mu = E(X)$  und der Varianz  $\sigma^2 = V(X)$ . Dann gilt die folgende Ungleichung:

$$P(\mu - t \cdot \sigma < X < \mu + t \cdot \sigma) \geq 1 - \frac{1}{t^2} \quad (t > 0)$$

oder in äquivalenter Formulierung

$$P(|X - \mu| \geq t \cdot \sigma) \leq \frac{1}{t^2} \quad (t > 0)$$

**Interpretation**

Für  $t = 10$  erhält man etwa im zweiten Fall

$$P(|X - \mu| \geq 10 \sigma) \leq \frac{1}{10^2} = 1\%.$$

Es ist also recht unwahrscheinlich, daß die Werte der Zufallsvariable  $X$  von ihrem Erwartungswert um mehr als  $10 \sigma$  abweichen. Positiv ausgedrückt: Es ist ziemlich sicher, daß gilt:

$$\mu - 10 \sigma < X < \mu + 10 \sigma.$$

Man nennt  $\sigma$  daher die *Streuung* von  $X$ .

*Beweis*

(i) diskreter Fall

Sei  $t > 0$ . dann ist

$$\begin{aligned}
 V(X) = \sigma^2 &= \sum_k (x_k - \mu)^2 f(x_k) \\
 &= \sum_{|x_k - \mu| < t\sigma} (x_k - \mu)^2 f(x_k) + \sum_{|x_k - \mu| \geq t\sigma} (x_k - \mu)^2 f(x_k) \\
 &\geq \sum_{|x_k - \mu| \geq t\sigma} (x_k - \mu)^2 f(x_k) \\
 &\geq t^2 \sigma^2 \sum_{|x_k - \mu| \geq t\sigma} f(x_k) \\
 &= t^2 \sigma^2 \sum_{|x_k - \mu| \geq t\sigma} P(\{\omega : X(\omega) = x_k\}) \\
 &= t^2 \sigma^2 \cdot P(\{\omega : |X(\omega) - \mu| \geq t\sigma\}) \\
 &= t^2 \sigma^2 \cdot P(|X - \mu| \geq t\sigma)
 \end{aligned}$$

Daraus ergeben sich für  $t > 0$  die beiden äquivalenten Beziehungen

$$P(|X - \mu| \geq t\sigma) \leq \frac{1}{t^2} \quad \text{sowie}$$

$$\begin{aligned}
 P(\mu - t\sigma < X < \mu + t\sigma) &= P(|X - \mu| < t\sigma) \\
 &= 1 - P(|X - \mu| \geq t\sigma) \\
 &\geq 1 - \frac{1}{t^2}
 \end{aligned}$$

(ii) kontinuierlicher Fall

Hausaufgabe!

### Beispiel 5.9

Für eine Serie von Glühbirnen betrage die durchschnittliche Lebensdauer 3 000 Stunden mit einer Standardabweichung von 250 Stunden. In einem Bürohaus seien 5 000 dieser Glühbirnen installiert.

Wieviele von ihnen müssen innerhalb einer Periode von 2 000 bis 4 000 Stunden ersetzt werden?

**Lösung**

Sei  $X$  die Zufallsvariable

$$X \quad :\Leftrightarrow \quad \text{Anzahl ausfallender Glühbirnen,}$$

dann gilt mit  $\mu = 3000$  ,  $\sigma = 250$  ,  $t \cdot \sigma = 1000 = 4 \cdot 250$  :

$$\begin{aligned} P(\mu - t\sigma < X < \mu + t\sigma) &= P(|X - \mu| < t \cdot \sigma) \geq 1 - \frac{1}{t^2} \\ \longrightarrow P(|X - 3000| < 4 \cdot 250) &\geq 1 - \frac{1}{4^2} = 0.9375 \\ \longrightarrow 5000 \cdot 0.9375 &= 4687.5 \end{aligned}$$

Innerhalb einer Zeitspanne von 2000 bis 4000 Stunden werden 4688 Glühbirnen erwartungsgemäß ausfallen, funktionieren also am Ende dieser Zeitspanne – nach 4000 Stunden – nicht mehr.

**Bemerkung**

Der Vorteil der Tschebyschev-Ungleichung – eine Ungleichung für eine *beliebige* Zufallsvariable – ist bei quantitativen Aussagen natürlich ein Nachteil. Man veranschauliche sich diese Ungleichung etwa für den Fall  $t = 1$  und gebe eine Interpretation. Setzt man hier zusätzlich die Kenntnis spezieller Verteilungen ein, so erhält man z.T. wesentlich bessere Resultate, s. dazu auch das folgende Beispiel.

**Beispiel 5.10**

Eine stetige Zufallsvariable  $X$  habe die Dichtefunktion (wie sieht diese aus?)

$$f(x) := \begin{cases} \frac{3}{4}x(2-x) & , \quad 0 \leq x \leq 2 \\ 0 & , \quad \text{sonst} \end{cases} .$$

Es soll die Wahrscheinlichkeit  $P(\mu - 2\sigma < X < \mu + 2\sigma)$  explizit berechnet und mit dem Ergebnis verglichen werden, welches sich mit Hilfe der Tschebyschev-Ungleichung ergibt.

**Lösung**

(i) Explizite Berechnung:

$$(a) \quad f \geq 0 \quad \text{und} \quad \int_{-\infty}^{\infty} f(x) dx = 1, \quad \text{somit ist } f \text{ eine Dichtefunktion.}$$

$$(b) \quad \mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_0^2 x \cdot \frac{3}{4}x(2-x) dx = 1$$

$$(c) \sigma^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx = \int_0^2 (x - 1)^2 \cdot \frac{3}{4} x (2 - x) dx = \frac{6}{5}$$

$$(d) P(\mu - 2\sigma < X < \mu + 2\sigma) = \int_{1-2\sqrt{\frac{6}{5}}}^{1+2\sqrt{\frac{6}{5}}} f(x) dx = \int_0^2 f(x) dx = 1$$

(ii) Anwendung der Tschebyschev-Ungleichung:

$$\begin{aligned} P(\mu - 2\sigma < X < \mu + 2\sigma) &= P\left(1 - 2 \cdot \frac{6}{5} < X < 1 + 2 \cdot \frac{6}{5}\right) \\ &\geq 1 - \frac{1}{4} = \frac{3}{4} \end{aligned}$$

## 5.4 Momente einer Verteilung

Seien  $X$  eine Zufallsvariable mit der Dichtefunktion  $f$  und  $g$  eine reellwertige Funktion. Dann gilt nach Satz 5.1 für den Erwartungswert der Zufallsvariable  $Y := g(X)$ :

$$E(Y) = E[g(X)] = \begin{cases} \sum_k g(x_k) \cdot f(x_k) & , \quad X \text{ diskret} \\ \int_{-\infty}^{\infty} g(x) \cdot f(x) dx & , \quad X \text{ stetig} \end{cases}$$

Für die Funktion  $g$  werden einige Spezialfälle betrachtet:

### Definition 5.3

- (i)  $E(X^k)$  heißt *k-tes Moment* von  $X$
- (ii)  $E(|X|^k)$  heißt *k-tes absolutes Moment* von  $X$
- (iii)  $E((X-c)^k)$  heißt *k-tes Moment bzgl. des Zentrums  $c$*  oder *k-tes zentriertes Moment* von  $X$
- (iv)  $E(|X-c|^k)$  heißt *k-tes absolutes Moment bzgl. des Zentrums  $c$*  oder *k-tes zentriertes absolutes Moment bzgl.  $c$*  von  $X$
- (v) Das auf die dritte Potenz der Standardabweichung bezogene zentrierte Moment 3. Ordnung wird als *Schiefe* von  $X$  bezeichnet:

$$\text{Schiefe von } X := \frac{E((X - \mu)^3)}{\sigma^3}$$

- (vi) Das auf die vierte Potenz der Standardabweichung bezogene zentrierte Moment 4. Ordnung wird als *Exzeß* von  $X$  bezeichnet:

$$\text{Exzeß von } X := \frac{E((X - \mu)^4)}{\sigma^4} - 3$$

**Bemerkung**

- (i) Die Momente sind Parameter der Verteilung von  $X$ , deshalb spricht man häufig auch von den *Momenten der Verteilung* statt von den Momenten der Zufallsvariable  $X$ .
- (ii) Existieren alle Momente einer Verteilung, so kann man daraus die Verteilung rekonstruieren, die Verteilung ist also durch ihre Momente festgelegt (allerdings u.U. nicht eindeutig).
- (iii) Der Erwartungswert einer Zufallsvariable ist gerade ihr erstes Moment (Schwerpunkt), ihre Varianz ihr zweites zentriertes Moment bzgl. des Schwerpunktes (Trägheitsmoment).
- (iv) Kennt man allerdings die *Art* der Verteilung, so genügen zu ihrer Festlegung häufig nur einige wenige dieser Momente:
  - (a) eine Exponentialverteilung ist durch ihr erstes Moment allein bestimmt, also durch ihren Erwartungswert. Ebenso eine diskrete Gleichverteilung.
  - (b) eine Normalverteilung ist bestimmt durch
    - ihr erstes Moment (Erwartungswert)
    - ihr zweites zentriertes Moment bzgl.  $\mu$ , also durch ihre Varianz:

$$E((X - \mu)^2) = V(X) = \sigma^2$$

Der nächste Satz stellt häufig verwendbare bequeme Abschätzungen für gewisse Wahrscheinlichkeiten bereit:

**Satz 5.5** (Markov-Ungleichung)

Seien  $X$  eine Zufallsvariable,  $\gamma \in \mathbb{R}$ . Die Momente  $E(|X - \gamma|^k)$  mögen für alle  $k \in \mathbb{N}$  existieren.

Dann gilt die folgende Ungleichung:

$$P(|X - \gamma| \geq t) \leq \frac{E(|X - \gamma|^k)}{t^k} \quad (t > 0).$$

Für  $\gamma := E(X)$  und  $k := 2$  erhält man aus der Markov-Ungleichung als Spezialfall die Tschebyschev-Ungleichung:

$$P(|X - E(X)| \geq t) \leq \frac{V(X)}{t^2} \quad (t > 0).$$

**5.5 Standardtransformation**

Sei  $X$  eine (diskrete oder stetige) Zufallsvariable mit den Kennwerten  $\mu = E(X)$  und  $\sigma^2 = V(X)$ . Seien  $a, b \in \mathbb{R}$  und sei

$$Z := g(X) := aX + b,$$

also  $g$  eine lineare Funktion von  $X$ .

Welches sind Erwartungswert und Varianz von  $Z$ ?

**Satz 5.6** (Kennwerte bei Skalierung und Translation)

- (i)  $\mu_Z := E(Z) = a \cdot \mu_X + b$
- (ii)  $\sigma_Z^2 := V(Z) = a^2 \cdot \sigma_X^2$
- (iii)  $\sigma_Z := \sqrt{V(Z)} = |a| \cdot \sigma_X$

**Bemerkung**

Der Erwartungswert  $\mu_X$  wird auf die gleiche Weise transformiert wie die Zufallsvariable  $X$  selbst.

Die Transformation der Varianz  $\sigma_X^2$  ist nur von  $a$ , nicht aber von dem konstanten Summanden  $b$  abhängig. Das erkennt man sofort an der für  $a \neq 0$  gültigen Darstellung  $Z = g(X) = aX + b = a(X + \frac{b}{a})$ .

*Beweis*

Zu (i) Linearität des Erwartungswertes (Satz 5.2 (ii))

Zu (ii) (a) Nach der Verschiebungsregel (Satz 5.3) gilt

$$V(Z) = E((Z - \mu_Z)^2) = E(Z^2) - \mu_Z^2$$

(b) Nach der Definition von  $Z$  gilt

$$\begin{aligned} E(Z^2) &= E((aX + b)^2) \\ &= E(a^2X^2 + 2abX + b^2) \\ &= a^2E(X^2) + 2abE(X) + b^2 \end{aligned}$$

(c) Gleichung (b) wird in (a) eingesetzt, und es wird (i) berücksichtigt:

$$\begin{aligned} V(Z) &= E(Z^2) - \mu_Z^2 \\ &= a^2E(X^2) + 2ab\mu_X + b^2 - (a\mu_X + b)^2 \\ &= a^2 \cdot E(X^2) - a^2\mu_X^2 \\ &= a^2[E(X^2) - \mu_X^2] \\ &= a^2 \cdot V(X) \end{aligned}$$

Zu (iii) Folgt unmittelbar aus (ii)

**Beispiel 5.11** (Standardtransformation)

$X$  sei eine Zufallsvariable mit dem Erwartungswert  $\mu_X = E(X)$  und der Varianz  $\sigma_X^2 = V(X)$ . Man betrachte die lineare Transformation

$$Z := g(X) := \frac{X - \mu_X}{\sigma_X} = \frac{1}{\sigma_X} \cdot X - \frac{\mu_X}{\sigma_X}.$$



Dann hat die Zufallsvariable  $Z$  nach Satz 5.6 den Erwartungswert

$$\mu_Z := E(Z) = \frac{1}{\sigma_X} \cdot \mu_X - \frac{\mu_X}{\sigma_X} = 0$$

und die Varianz

$$\sigma_Z^2 := V(Z) = a^2 \cdot \sigma_X^2 = \left(\frac{1}{\sigma_X}\right)^2 \cdot \sigma_X^2 = 1.$$

#### Definition 5.4

Die im letzten Beispiel vorgestellte Transformation heißt *Standardtransformation*, und die dort definierte Zufallsvariable  $Z = \frac{X-\mu}{\sigma}$  heißt die zu  $X$  gehörende *standardisierte Zufallsvariable*. Den Effekt der Standardtransformation einer Zufallsvariable  $X$  auf Erwartungswert und Varianz bzw. Standardabweichung kann man an Abb. 5.5 erkennen.

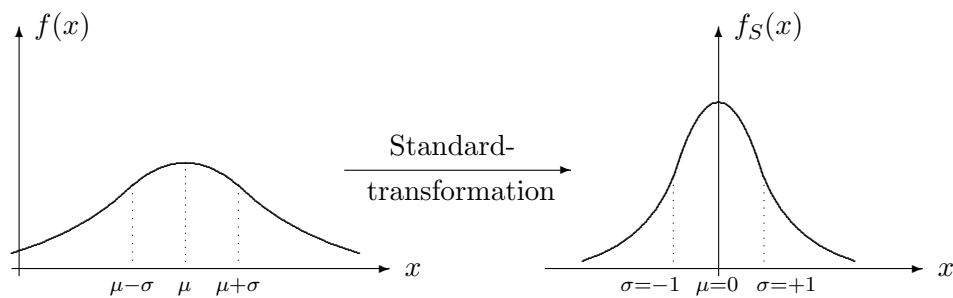


Abbildung 5.5: Standardtransformation einer Zufallsvariable  $X$

# Kapitel 6

## Spezielle Wahrscheinlichkeitsverteilungen

In diesem Kapitel werden einige spezielle häufig auftretende Wahrscheinlichkeitsverteilungen betrachtet. Diese Verteilungen werden beschrieben, indem ihre Dichtefunktion  $f$  oder ihre Verteilungsfunktion  $F$  angegeben werden. Da es mehrere Zufallsvariablen geben kann, welche dieselbe Verteilung, also auch dieselbe Dichte haben (Münzwurf, Ziehen einer roten Karte aus einem Stapel von zwei verschiedenfarbigen Karten), bezeichnet man oft die Dichte durch das Symbol  $f$  allein, ohne einen Index  $X$  anzufügen. Man spricht deshalb auch von der *Dichte einer Verteilung* anstatt von der Dichte einer Zufallsvariable.

### 6.1 Diskrete Wahrscheinlichkeitsverteilungen

#### 6.1.1 Diskrete Gleichverteilung

Die einfachste aller diskreten Wahrscheinlichkeitsverteilungen ist die *Gleichverteilung*: Alle Werte der Zufallsvariable  $X$  werden mit derselben Wahrscheinlichkeit  $\frac{1}{n}$  angenommen ( $n \in \mathbb{N}$  geeignet):

$$W(X) := \{x_1, \dots, x_n\} \quad \implies$$
$$f(x) := f(x; n) := \begin{cases} \frac{1}{n} & , \quad x \in W(X) \\ 0 & , \quad \text{sonst} \end{cases}$$

Die Dichte  $f$  einer solchen Verteilung hängt nur von einem Parameter ab, nämlich von  $n = |W(X)|$ . Eine diskrete Gleichverteilung setzt somit einen endlichen Wertebereich  $W(X)$  von  $X$  voraus.

#### Beispiele

- (i) Würfeln mit einem homogenen Würfel
- (ii) Werfen einer homogenen Münze

- (iii) Ziehen einer Karte aus einem Kartenspiel mit  $n$  verschiedenen Karten
- (iv) Ziehen einer bestimmten Kugel aus einer Urne mit  $n$  verschiedenfarbigen Kugeln

Die graphische Veranschaulichung einer diskreten Gleichverteilung mit Hilfe eines Histogrammes besteht aus  $n = |W(X)|$  Rechtecken gleicher Höhe, s. dazu Abb. 6.1.

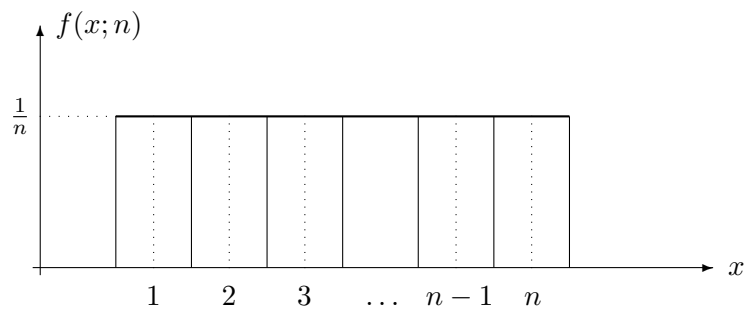


Abbildung 6.1: Histogramm einer Gleichverteilung

### Satz 6.1

Sei  $X$  eine diskrete gleichverteilte Zufallsvariable mit  $W(X) := \{1, \dots, n\}$ . Dann gilt:

- (i) 
$$E(X) = \frac{1}{n} \sum_{k=1}^n k = \frac{n+1}{2}$$
- (ii) 
$$V(X) = \frac{1}{n} \sum_{k=1}^n \left(k - \frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}$$

*Beweis*

- (i) Für den Erwartungswert gilt

$$E(X) = \sum_{k=1}^n k \cdot \frac{1}{n} = \frac{1}{n} \sum_{k=1}^n k = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

- (ii) Für die Varianz erhält man mit dem Verschiebungssatz (Satz 5.3)

$$\begin{aligned}
V(X) &= E(X^2) - [E(X)]^2 \\
&= \frac{1}{n} \sum_{k=1}^n k^2 - \left(\frac{n+1}{2}\right)^2 \\
&= \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \\
&= \frac{n+1}{12} [2(2n+1) - 3(n+1)] \\
&= \frac{n^2 - 1}{12}.
\end{aligned}$$

**Bemerkung**

Die diskrete Gleichverteilung ist sowohl durch den (einen) Parameter Erwartungswert als auch durch den (einen) Parameter Varianz eindeutig festgelegt:

$$\begin{aligned}
E(X) = \frac{n+1}{2} &\iff n = |W(X)| = 2 \cdot E(X) - 1 \\
V(X) = \frac{n^2 - 1}{12} &\iff n = |W(X)| = \sqrt{12V(X) + 1}
\end{aligned}$$

**6.1.2 Multinomialverteilung****6.1.2.1 Bernoulli-Experiment****Definition 6.1**

Ein *Bernoulli-Experiment* oder *Null-Eins-Experiment* ist ein Experiment, bei dem nur *zwei* verschiedene sich gegenseitig ausschließende Ereignisse mit den beiden konstanten Wahrscheinlichkeiten

$$\begin{aligned}
p &:= P(A) && \text{und} \\
q &:= P(\bar{A}) = 1 - P(A)
\end{aligned}$$

eintreten können.

**Beispiele**

(i) Werfen einer homogenen Münze

$$\begin{aligned}
&K \iff \text{Kopf} \quad , \quad W \iff \text{Wappen} \\
\longrightarrow \quad p = P(K) = \frac{1}{2} \quad , \quad q = P(\bar{K}) = P(W) = \frac{1}{2}
\end{aligned}$$

- (ii) In einer Urne befinden sich sechs weiße und zwei schwarze Kugeln. Bei der zufälligen Entnahme einer Kugel sind nur die beiden sich gegenseitig ausschließenden Ereignisse

$$\begin{aligned} W & : \iff \text{Ziehung einer weißen Kugel} \\ S & : \iff \text{Ziehung einer schwarzen Kugel} \end{aligned}$$

möglich. Sie treten auf mit den Wahrscheinlichkeiten

$$p := P(W) = \frac{6}{8} = \frac{3}{4} \quad \text{und} \quad q := P(S) = \frac{2}{8} = \frac{1}{4}.$$

Eine entsprechende Zufallsvariable  $X$  sei definiert durch

$$X : \iff \text{Ziehen einer weißen Kugel.}$$

Sie hat den Wertebereich  $W(X) = \{0, 1\}$ . Der Parameter  $p$  heißt auch *Erfolgswahrscheinlichkeit*, der Parameter  $q$  dementsprechend *Mißerfolgswahrscheinlichkeit*. Die folgenden Größen ergeben sich sehr einfach:

- (i) Dichtefunktion eines Bernoulli-Experimentes:

$$f(x) := \begin{cases} q & , \quad x = 0 \\ p & , \quad x = 1 \\ 0 & , \quad \text{sonst} \end{cases} = \begin{cases} p^x \cdot (1-p)^{1-x} & , \quad x \in \{0, 1\} \\ 0 & , \quad \text{sonst} \end{cases}$$

- (ii) Erwartungswert eines Bernoulli-Experimentes: (Abb. 6.2)

$$E(X) = 1 \cdot p + 0 \cdot q = p$$

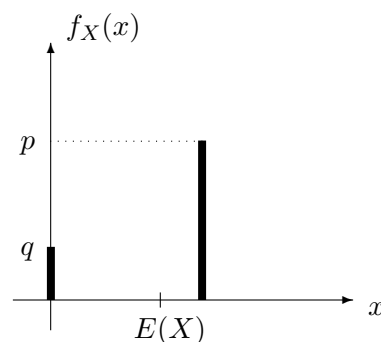


Abbildung 6.2: Erwartungswert bei einem Bernoulli-Experiment

- (iii) Varianz eines Bernoulli-Experimentes:

$$\begin{aligned} V(X) &= E(X - E(X))^2 \stackrel{\text{(Satz 5.3)}}{=} E(X^2) - (E(X))^2 \\ &= 1^2 \cdot p + 0^2 \cdot q - p^2 \\ &= p(1-p) = p \cdot q \end{aligned}$$

### 6.1.2.2 Binomialverteilung

#### Definition 6.2

Ein Experiment, welches aus einer  $n$ -fachen Ausführung eines Bernoulli-Experimentes mit den beiden möglichen (und sich gegenseitig ausschließenden) Ereignissen  $A$  und  $\bar{A}$  besteht, heißt ein *Bernoulli-Experiment vom Umfang  $n$*  oder eine *Bernoulli-Kette vom Umfang  $n$* .

#### Beispiele

$n$ -fache Ausführung des Experimentes

- (i) Werfen einer homogenen Münze
- (ii) Ziehen einer Kugel aus einer Urne

#### Bemerkung

Beispiel (ii) ist nur dann ein Bernoulli-Experiment, wenn nach jeder Ziehung die Ausgangslage wiederhergestellt wird, die gezogene Kugel vor dem nächsten Zug also wieder zurückgelegt wird.

Versuche mit mehreren Ausgängen (Würfeln) führen zu sog. *Multinomialverteilungen*.

#### Beispiel 6.1

In einer Urne befinden sich  $w = 14$  weiße und  $s = 6$  schwarze Kugeln. Ein Versuch bestehe in der  $n = 3$ -maligen Ausführung des Experimentes „Ziehen einer Kugel mit Zurücklegen“. Das Ziehen einer weißen Kugel gelte dabei als Erfolg, das Ziehen einer schwarzen Kugel als Mißerfolg.

Die Zufallsvariable  $X$  sei definiert durch

$$X \quad :\Leftrightarrow \quad \begin{array}{l} \text{Anzahl der gezogenen weißen Kugeln} \\ \text{bei } n = 3 \text{ Ziehungen mit Zurücklegen.} \end{array}$$

Die Wahrscheinlichkeit, eine weiße Kugel zu ziehen, ist dann gegeben durch

$$p := \frac{w}{s+w} = \frac{14}{6+14} = \frac{7}{10} \quad (\text{Erfolgswahrscheinlichkeit});$$

die Wahrscheinlichkeit, eine schwarze Kugel zu ziehen, ist dementsprechend gegeben durch

$$q := 1 - p = \frac{s}{s+w} = \frac{6}{6+14} = \frac{3}{10} \quad (\text{Mißerfolgswahrscheinlichkeit}).$$

Die Erfolgswahrscheinlichkeit  $p$  bzw. die Mißerfolgswahrscheinlichkeit  $q$  sollen nun berechnet werden, wenn mehr als ein Versuch ausgeführt wird. Dazu wird eine Bernoulli-Kette vom Umfang  $n$  betrachtet.

**Bernoulli-Kette vom Umfang  $n$** 

Ein Versuch bestehe aus der  $n$ -mal wiederholten Durchführung eines Bernoulli-Experimentes mit der Erfolgswahrscheinlichkeit  $p$ . Die Menge der Elementarereignisse wird dann adäquat beschrieben durch den Merkmalraum

$$\Omega := \underbrace{\{0, 1\} \times \{0, 1\} \times \dots \times \{0, 1\}}_{n \text{ Faktoren}} = \{0, 1\}^n.$$

Das Bernoulli-Experiment  $i$  sei unabhängig vom Bernoulli-Experiment  $j$ . Die Zufallsvariable ist für jedes der  $n$  Experimente natürlich dieselbe, nämlich  $X$ . Um zu bezeichnen, welches der Experimente man denn nun gerade auswertet, hat sich etabliert,  $X$  zu indizieren und mit  $X_i$  gerade diejenige Zufallsvariable zu bezeichnen, welche für das  $i$ -te Experiment „zuständig“ ist.

Für jedes  $i \in \{1, \dots, n\}$  wird gesetzt:

$$X_i(\omega) := \begin{cases} 1 & , \text{ wenn das } i\text{-te Experiment erfolgreich war} \\ 0 & , \text{ sonst} \end{cases},$$

sowie

$$\begin{aligned} A_i^1 &:= \{\omega : \omega \in \Omega = \{0, 1\}^n, X_i(\omega) = 1\} =: \{„X_i = 1“\}, \\ A_i^0 &:= \overline{A_i^1}. \end{aligned}$$

Dann ist jedes Elementarereignis  $\omega \in \Omega$  von der Form

$$\{\omega\} = A_1^{x_1} \cap A_2^{x_2} \cap \dots \cap A_n^{x_n},$$

wobei  $x_i \in \{0, 1\}$  ist und den jeweils beobachteten Wert von  $X_i$  bedeutet, also  $x_i = X_i(\omega)$ . Dieser Wert  $x_i$  ergibt sich somit als mögliche Auswertung eines Elementarereignisses  $\omega$ , man sagt: er kann an dem Elementarereignis  $\omega$  *beobachtet* werden.

Sei nun  $x$  die Anzahl der Erfolge in  $n$  Versuchen, also

$$x := x_1 + x_2 + \dots + x_n,$$

dann gehört  $x$  zum Wertebereich  $W(X)$  der Zufallsvariable  $X$ , welche die Anzahl der Erfolge in  $n$  Versuchen zählt:

$$X := X_1 + X_2 + \dots + X_n.$$

Weiterhin gilt wegen der Unabhängigkeit der Experimente für die Ereignisse  $A_i^{x_i}$  die Produktregel, und man erhält

$$P(\{\omega\}) = P\left(\bigcap_{i=1}^n A_i^{x_i}\right) = \prod_{i=1}^n P(A_i^{x_i}) = p^x \cdot q^{n-x}.$$

Das ist die Wahrscheinlichkeit,  $x$  weiße und  $n - x$  schwarze Kugeln in einer bestimmten Reihenfolge zu ziehen.

$x$	0	1	2	...	$n$
$f(x)$	$q^n$	$\binom{n}{1} q^{n-1} p$	$\binom{n}{2} q^{n-2} p^2$	...	$p^n$

Tabelle 6.1: Verteilungstabelle einer binomialverteilten Zufallsvariable

Jeder Wert  $x$  kann nun auf  $\binom{n}{x}$  verschiedene Arten zustande kommen, denn die  $n$  gezogenen Kugeln können auf  $n!$  verschiedene Arten angeordnet werden, welche ihrerseits in zwei Klassen zu je  $x$  weißen - und  $n - x$  schwarzen Kugeln zerfallen. Die Anzahl unterschiedlicher Permutationen (mit Wiederholung) lautet daher

$$P(n; x, n - x) = \frac{n!}{x! (n - x)!} = \binom{n}{x} = K(n; x)$$

und ist eine Kombination ohne Wiederholung.

### Definition 6.3

Die Wahrscheinlichkeitsdichte

$$f(x) := b(x; n, p) := \begin{cases} \binom{n}{x} p^x q^{n-x} & , \quad x \in \{0, \dots, n\} \\ 0 & , \quad \text{sonst} \end{cases}$$

heißt Wahrscheinlichkeitsdichte der *Binomialverteilung*. Sie besitzt die Verteilungstabelle 6.1.

### Bemerkung

- (i) Der Name *Binomialverteilung* erklärt sich aus der Tatsache, daß die in der Verteilungstabelle angegebenen Wahrscheinlichkeiten der Reihe nach den Summanden in der binomischen Entwicklung von  $(q + p)^n$  entsprechen:

$$\begin{aligned} 1 &= (q + p)^n = q^n + \binom{n}{1} q^{n-1} p + \binom{n}{2} q^{n-2} p^2 + \dots + p^n \\ &= f(0) + f(1) + f(2) + \dots + f(n) \end{aligned}$$

- (ii) Erwartungswert und Varianz einer binomialverteilten Zufallsvariable  $X$  sind gegeben durch

$$\begin{aligned} E(X) &= \sum_{x=0}^n x \cdot \binom{n}{x} p^x \cdot q^{n-x} = np, \\ V(X) &= \sum_{x=0}^n x^2 \cdot \binom{n}{x} p^x \cdot q^{n-x} - (np)^2 = npq. \end{aligned}$$



- (iii) Für die Gültigkeit einer Binomialverteilung müssen die folgenden Bedingungen erfüllt sein:
- (a) Feste Anzahl  $n$  von Experimenten
  - (b) Zwei Ausgänge für jedes Experiment
  - (c) Bei jedem Experiment dieselbe Wahrscheinlichkeit
  - (d) Voneinander unabhängige Versuche

Beispielsweise liegt keine Binomialverteilung vor, wenn die Anzahl  $n$  der Experimente von dem Ausgang des vorherigen Experimentes abhängig gemacht wird: Solange Ziehen einer Kugel mit Zurücklegen, bis man eine weiße erhalten hat ...

### Beispiel 6.2

Die Wahrscheinlichkeit, daß ein bestimmtes Maschinenteil einen Belastungstest besteht, sei  $p = 75\%$ . Wie groß ist die Wahrscheinlichkeit, daß von den nächsten vier zu testenden Maschinenteilen

- (i) genau zwei
- (ii) mehr als zwei

den Test bestehen?

### Lösung

Die Zufallsvariable

$X$   $:\Leftrightarrow$  Anzahl der den Belastungstest bestehenden Maschinenteile

ist binomialverteilt mit den Parametern

$n$   $:=$  4 (Anzahl der Maschinenteile, also der Stufen des Versuches)  
 $p$   $:=$   $\frac{3}{4}$  (Erfolgswahrscheinlichkeit)  
 $x$   $:=$  2 (aktueller Wert der Zufallsvariable  $X$ )

- (i) Genau zwei Maschinenteile bestehen den Test mit der Wahrscheinlichkeit

$$\begin{aligned}
 P(X = 2) &= b(x = 2; n = 4, p = 0.75) \\
 &= \binom{4}{2} \cdot \left(\frac{3}{4}\right)^2 \cdot \left(\frac{1}{4}\right)^2 = \frac{4!}{2!2!} \cdot \frac{3^2}{4^4} \\
 &= \frac{27}{128} = 21,1\%.
 \end{aligned}$$

- (ii) Mehr als zwei Maschinenteile bestehen den Test mit der Wahrscheinlichkeit

$$\begin{aligned}
 P(X > 2) &= P(X = 3) + P(X = 4) \\
 &= b(x = 3; n = 4, p = 0.75) + b(x = 4; n = 4, p = 0.75) \\
 &= \binom{4}{3} \cdot \left(\frac{3}{4}\right)^3 \cdot \left(\frac{1}{4}\right)^1 + \binom{4}{4} \cdot \left(\frac{3}{4}\right)^4 \cdot \left(\frac{1}{4}\right)^0 \\
 &= 4 \cdot \frac{3^3}{4^4} + 1 \cdot \frac{3^4}{4^4} = \frac{27}{64} + \frac{81}{256} \\
 &= \frac{189}{256} = 73,8\%.
 \end{aligned}$$

### Beispiel 6.3

Für den Fall einer binomialverteilten Zufallsvariable ist das von Johann Bernoulli im 18. Jahrhundert entdeckte *schwache Gesetz der großen Zahlen* recht erhellend.

Sei  $X$  eine binomialverteilte Zufallsvariable, also eine Zufallsvariable mit der Dichtefunktion

$$f(x) := b(x; n, p) \quad (x \in \mathbb{R}).$$

Nun betrachte man die Zufallsvariable

$$Y := g(X) := \frac{X}{n},$$

welche den Anteil der Erfolge in  $n$  Versuchen beschreibt, die also den relativen Erfolgsanteil mißt. Es gelten dann

$$\begin{aligned}
 E(X) = n \cdot p &\longrightarrow E(Y) = \frac{1}{n} \cdot E(X) = p \\
 V(X) = n \cdot p \cdot q &\longrightarrow V(Y) = \frac{1}{n^2} \cdot V(X) = \frac{p \cdot q}{n}
 \end{aligned}$$

Auf die Zufallsvariable  $Y$  und ihre Parameter wird nun die Tschebyschev-Ungleichung (Satz 5.4) angewandt,

$$P(|Y - \mu| \geq t \cdot \sqrt{V(Y)}) \leq \frac{1}{t^2} \quad (t > 0) :$$

$$Y = \frac{X}{n}, \quad \mu = E(Y) = p, \quad s := t \cdot \sqrt{V(Y)} \iff t = \frac{s}{\sqrt{V(Y)}}$$

$$\begin{aligned}
 \longrightarrow & P\left(\left|\frac{X}{n} - p\right| \geq s = t \cdot \sqrt{V(Y)}\right) \\
 & \leq \frac{1}{t^2} = \frac{V(Y)}{s^2} = \frac{pq}{n s^2} \stackrel{(*)}{\leq} \frac{1}{n s^2} \quad (s > 0),
 \end{aligned}$$

wobei sich die letzte Ungleichung (\*) aus  $pq \in (0, 1)$  für  $p \in (0, 1)$  ergibt.<sup>1</sup>

### Aussage

Die Wahrscheinlichkeit, daß der Anteil der positiven Versuchsausgänge vom Erwartungswert abweicht, konvergiert mit der Anzahl  $n$  der Versuche gegen Null.

Für binomialverteilte Zufallsvariablen läßt sich somit die Stabilität der relativen Häufigkeiten für  $n \rightarrow \infty$  präzisieren:

Mit wachsendem Versuchsumfang konvergiert der Anteil der positiven Versuchsausgänge gegen die „wahre“ Erfolgswahrscheinlichkeit.

Wesentlich hierbei ist, daß die Wahrscheinlichkeit eines zufälligen Ereignisses nicht der Grenzwert der relativen Häufigkeit im Sinne der klassischen Analysis ist:

$$\lim_{n \rightarrow \infty} |Y_n - p| = 0,$$

sondern es konvergiert „nur“ die Wahrscheinlichkeit eines Ereignisses gegen Null:

$$\lim_{n \rightarrow \infty} P(\{\omega : |Y_n(\omega) - p| \geq \varepsilon\}) = 0 \quad (\varepsilon > 0).$$

Man sagt hierzu,  $Y_n := \frac{X}{n}$  konvergiere gegen  $p$  in Wahrscheinlichkeit, oder die Folge  $\{Y_n = \frac{X}{n}\}_{n \in \mathbb{N}}$  unterliege dem schwachen Gesetz der großen Zahlen.

### Beispiel 6.4

Mit einer Meinungsumfrage soll der Anteil der Personen an der Bevölkerung abgeschätzt werden, welcher die Arbeitsmarktpolitik der Regierung unterstützt. Wie groß muß die Stichprobe mindestens sein, damit der geschätzte vom wahren Anteil mit einer Wahrscheinlichkeit von mindestens 90 % um nicht mehr als 5 % abweicht?

### Lösung

Hier wird vorausgesetzt, daß die Meinungen innerhalb der Bevölkerung binomialverteilt sind, eine durchaus fragwürdige Annahme. Was bedeutet es beispielsweise, daß die Wahrscheinlichkeit bei jedem Experiment, also bei jeder Befragung dieselbe ist (Bemerkung (iii) (c) auf p. 109)? Es bedeutet, daß die Wahrscheinlichkeit  $p$  für alle befragten Personen konstant ist, unabhängig von Alter, Wohnort, Geschlecht, ...

<sup>1</sup>Man beweise, daß sich diese Abschätzung wesentlich verbessern läßt, indem man die Gültigkeit der Implikation

$$p \in [0, 1] \implies pq \leq \frac{1}{4}$$

nachweist.

Nach dem schwachen Gesetz der großen Zahlen aus Beispiel 6.3,

$$P\left(\left|\frac{X}{n} - p\right| \geq s\right) \leq \frac{1}{n s^2} \quad (s > 0)$$

wird eine untere Schranke für  $n$  zum Wert  $s = 0.05$  gesucht, so daß gilt:

$$P\left(\left|\frac{X}{n} - p\right| < 0.05\right) \geq 0.90 \quad \text{bzw.} \quad P\left(\left|\frac{X}{n} - p\right| \geq 0.05\right) \leq 0.10.$$

Die letzte Ungleichung wird erfüllt, falls gilt:

$$\frac{1}{n \cdot (0.05)^2} \leq 0.10 \quad \longleftrightarrow \quad n \geq 4000.$$

### Bemerkung

Obige Abschätzung hat man erhalten unter Benutzung der groben Abschätzung  $pq = p(1-p) < 1$ . Unter Beachtung der Fußnote aus Beispiel 6.3 läßt sich das Ergebnis noch um den Faktor  $1/4$  verbessern, die Anzahl der zu befragenden Personen kann also auf  $n \geq 1000$  gedrückt werden.

### 6.1.3 Hypergeometrische Verteilung

Eine  $n$ -fache Ausführung eines Bernoulli-Experimentes führt zu einer Binomialverteilung, die im letzten Abschnitt studiert worden ist. Eine Binomialverteilung liegt jedoch nur dann vor, wenn die Wahrscheinlichkeit bei jedem Einzelexperiment dieselbe ist. Ist das nicht der Fall, so kann man keine Binomialverteilung mehr erwarten.

#### Typische Beispiele

$n$ -fache Ausführung eines Auswahlexperimentes *ohne* Zurücklegen, etwa

- (i) Ziehen einer Karte aus einem Kartenspiel
- (ii) Ziehen einer Kugel aus einer Urne (Lottospiel „6 aus 49“)
- (iii) Qualitätskontrolle (Suchen und ggfs. Entfernen eines defekten Teils)

#### Beispiel 6.5

In einer Urne befinden sich fünf weiße und drei schwarze Kugeln. Nacheinander werden drei Kugeln *ohne* Zurücklegen gezogen.

Wie groß ist die Wahrscheinlichkeit, bei drei Zügen zwei weiße und eine schwarze Kugel zu ziehen?

**Lösung**

- (i) Einsammeln der Wahrscheinlichkeiten mit Hilfe eines Ereignisbaumes:

$$P(\langle W, W, S \rangle) = \frac{5}{8} \cdot \frac{4}{7} \cdot \frac{3}{6} = \frac{5}{28}$$

$$P(\langle W, S, W \rangle) = \frac{5}{8} \cdot \frac{3}{7} \cdot \frac{4}{6} = \frac{5}{28}$$

$$P(\langle S, W, W \rangle) = \frac{3}{8} \cdot \frac{5}{7} \cdot \frac{4}{6} = \frac{5}{28}$$

$$\longrightarrow P(2W \text{ und } 1S) = 3 \cdot \frac{5}{28} = \frac{15}{28}$$

- (ii) Erstellung einer Formel für die Verteilung:

Insgesamt gibt es

$\binom{5}{2}$  Möglichkeiten, aus fünf weißen Kugeln zwei weiße Kugeln auszuwählen;

$\binom{3}{1}$  Möglichkeiten, aus drei schwarzen Kugeln eine schwarze Kugel auszuwählen;

$\binom{8}{3}$  Möglichkeiten, aus acht Kugeln drei Kugeln auszuwählen.

Damit erhält man für die gesuchte Wahrscheinlichkeit:

$$\begin{aligned} P(2W \text{ und } 1S) &= \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} \\ &= \frac{\binom{5}{2} \cdot \binom{3}{1}}{\binom{8}{3}} = \frac{15}{28} \end{aligned}$$

**Definition 6.4**

- (i) Ein Experiment heißt *hypergeometrisches Experiment*, wenn es die folgenden Eigenschaften hat:
- Eine Zufallsstichprobe vom Umfang  $n$  wird aus einer Gesamtheit von  $N$  Elementen ausgewählt;
  - Von den  $N$  Elementen besitzen  $k$  die Eigenschaft  $A$  (Erfolg) und  $N - k$  die Eigenschaft  $\bar{A}$  (Mißerfolg).
- (ii) In einem hypergeometrischen Experiment sei die Zufallsvariable  $X$  definiert durch

$$X \quad : \iff \quad \text{Anzahl der Erfolge}$$

Dann heißt  $X$  eine *hypergeometrische Zufallsvariable*.

(iii) Die Wahrscheinlichkeitsdichte

$$f(x) := h(x; N, n, k) := \frac{\binom{k}{x} \cdot \binom{N-k}{n-x}}{\binom{N}{n}} \quad (x \in \{0, \dots, n\})$$

heißt Wahrscheinlichkeitsdichte der *hypergeometrischen Verteilung*.

Erwartungswert und Varianz einer hypergeometrischen Verteilung sind gegeben durch

$$\begin{aligned} \mu = E(X) &= \frac{nk}{N} && \text{und} \\ \sigma^2 = V(X) &= \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N} \left(1 - \frac{k}{N}\right). \end{aligned}$$

### Beispiel 6.6

Eine Warenlieferung von elektronischen Bauteilen gelte als einwandfrei, wenn sie auf Pakete à hundert Stück nicht mehr als drei defekte Teile enthalte. Zur Kontrolle werden einer Einheit fünf Bauteile entnommen und geprüft. Die Ware wird zurückgewiesen, wenn ein defektes Teil gefunden wird.

- (i) Welches ist die Wahrscheinlichkeit, in einem Paket genau ein defektes Bauteil zu finden, wenn das ganze Paket drei defekte Bauteile enthält?
- (ii) Man berechne und interpretiere das Intervall  $[\mu - 2\sigma, \mu + 2\sigma]$  mit Hilfe der Tschebyschev-Ungleichung.

### Lösung

- (i) Bei dieser Prüfmethode genügt die Zufallsvariable

$$X : \iff \text{Anzahl der defekten elektronischen Bauteile}$$

einer hypergeometrischen Verteilung mit den Parametern  $N = 100$ ,  $n = 5$ ,  $k = 3$  und  $x = 1$ . Damit ist die Wahrscheinlichkeit, in einem Paket genau ein defektes Bauteil zu finden, wenn das ganze Paket drei defekte Teile enthält, gegeben durch

$$h(x=1; N=100, n=5, k=3) = \frac{\binom{3}{1} \cdot \binom{100-3}{5-1}}{\binom{100}{5}} = 0,1380 = 13,8\%.$$

(ii) Für Erwartungswert, Varianz und Standardabweichung berechnet man

$$\begin{aligned}\mu &= \frac{5 \cdot 3}{100} = \frac{15}{100} = 0.15 \\ \sigma^2 &= \frac{100-5}{39} \cdot 5 \cdot \frac{3}{100} \cdot \left(1 - \frac{3}{100}\right) = 0.1396 \\ \sigma &= 0.374 \\ (\mu - 2\sigma, \mu + 2\sigma) &= (0.15 - 2 \cdot 0.374, 0.15 + 2 \cdot 0.374) \\ &= (-0.598, 0.898)\end{aligned}$$

Nach der Tschebyschev-Ungleichung gilt für die Zufallsvariable  $X$  die Abschätzung

$$P(\{|X - \mu| < 2\sigma\}) \geq 1 - \frac{1}{4} = 0.75.$$

### Interpretation

Für eine Gruppe von hundert elektronischen Bauteilen mit drei defekten gilt bei zufälliger Entnahme von fünf Teilen, daß mit einer Wahrscheinlichkeit von wenigstens 75 % die Anzahl der defekten Bauteile zwischen

$$-0.598 \text{ (also 0)} \quad \text{und} \quad 0.898$$

liegt.

Positiv ausgedrückt:

In wenigstens 75 % der Fälle enthalten die fünf entnommenen Teile weniger als ein defektes, die Lieferung gilt demnach als einwandfrei.

### Bemerkung

(i) Grundsätzlich gilt:

$$\begin{array}{ll} \text{Ziehung mit Zurücklegen} & \longrightarrow \text{ Binomialverteilung} \\ \text{Ziehung ohne Zurücklegen} & \longrightarrow \text{ hypergeometrische Verteilung} \end{array}$$

(ii) Für eine Ziehung mit Zurücklegen ist die Wahrscheinlichkeit eines Ereignisses bei allen Versuchen konstant  $p = \frac{k}{N}$ .

Für eine Ziehung ohne Zurücklegen gilt:

$$\begin{array}{ll} 1. \text{ Ziehung} & : \quad p = \frac{k}{N} \\ 2. \text{ Ziehung} & : \quad p = \frac{k-1}{N-1} \quad \text{oder} \quad p = \frac{k}{N-1} \end{array}$$

je nachdem, ob im ersten Zug ein defektes Teil gezogen wurde oder nicht.

- (iii) Für  $N \gg n$  sind diese Änderungen so gering, daß sie keine nennenswerte Rolle spielen. Die hypergeometrische Verteilung läßt sich dann näherungsweise durch die rechnerisch bequemere Binomialverteilung ersetzen:

$$N \gg n \quad \implies \quad h(x; N, n, k) \approx b(x; n, p = \frac{k}{N}).$$

Für konstante Werte  $p = \frac{k}{N}$  für  $N \rightarrow \infty$  gilt:

$$\lim_{\substack{N \rightarrow \infty \\ (p = \frac{k}{N} \text{ konstant)}}} h(x; N, n, k) = b(x; n, p) \quad (x \in \mathbb{R}).$$

- (iv) Für Erwartungswert und Varianz gilt ( $h$ : hypergeometrische Verteilung,  $b$ : Binomialverteilung):

$$\begin{aligned} E_h(X) &= n \cdot \frac{k}{N} = n \cdot p = E_b(X) \\ V_h(X) &= \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N} \cdot \left(1 - \frac{k}{N}\right) \\ &= \frac{N-n}{N-1} \cdot n \cdot p \cdot q \\ &= \frac{N-n}{N-1} \cdot V_b(X) \\ &= \frac{1-n/N}{1-1/N} \cdot V_b(X) \\ &\xrightarrow{(N \rightarrow \infty)} V_b(X) \end{aligned}$$

### Beispiel 6.7

Eine Lieferung enthält  $N = 100$  Transistoren, die aus einer Massenproduktion mit 5% Ausschuß stammen. Bei Anlieferung der Ware wird vom Kunden eine Abnahmekontrolle in Form einer Stichprobe vom Umfang  $n = 4$  ohne Zurücklegen durchgeführt. Die entnommenen Transistoren werden dabei auf ihre Funktionstüchtigkeit überprüft.

Mit welcher Wahrscheinlichkeit enthält die durchgeführte Stichprobe nur einwandfreie Ware?

### Lösung

- (i) Korrekte Berechnung mit der hypergeometrischen Verteilung:

Die Zufallsvariable

$X \iff$  Anzahl der pro Stichprobe vorhandenen defekten Transistoren

ist hypergeometrisch verteilt. Mit den Parametern

$$N = 100, \quad k = 5 \text{ (5\% Ausschuß)}, \quad n = 4$$



lautet die Wahrscheinlichkeitsfunktion dieser Verteilung

$$f_h(x) = P_h(X = x) = \frac{\binom{5}{x} \cdot \binom{100-5}{4-x}}{\binom{100}{4}} = \frac{\binom{5}{x} \cdot \binom{95}{4-x}}{\binom{100}{4}}$$

( $x \in \{0, 1, 2, 3, 4\}$ ). Für  $x = 0$  folgt daraus

$$P_h(X = 0) = \frac{\binom{5}{0} \cdot \binom{95}{4}}{\binom{100}{4}} = \frac{95 \cdot 94 \cdot 93 \cdot 92}{100 \cdot 99 \cdot 98 \cdot 97} = 0.8119 \approx 81,2\%$$

(ii) Angenäherte Berechnung mit der approximierenden Binomialverteilung:

Zu einem ähnlichen Ergebnis gelangt man, wenn man die hypergeometrische Verteilung näherungsweise durch die Binomialverteilung mit den Parametern  $n = 4$  und  $p = \frac{k}{N} = 0.05$  ersetzt. Wegen  $n = 5 \ll 100 = N$  ist eine solche Näherung zulässig.

Die Wahrscheinlichkeitsfunktion für die Binomialverteilung lautet dann

$$f_b(x) = P_b(X = x) = \binom{4}{x} \cdot 0.05^x \cdot 0.95^{4-x} \quad (x \in \{0, 1, 2, 3, 4\}).$$

Aus ihr erhält man für die gesuchte Wahrscheinlichkeit  $P_b(X = 0)$  den Näherungswert

$$P_b(X = 0) = \binom{4}{0} \cdot 0.05^0 \cdot 0.95^4 = 0.8145 \approx 81,5\%,$$

in guter Übereinstimmung mit dem exakten Wert von 81,2% aus Teil (i).

## 6.1.4 Weitere diskrete Verteilungen

Es gibt weitere diskrete Verteilungen, von denen zwei hier nur kurz angesprochen werden sollen. Für weitergehende Studien sei auf die Spezialliteratur verwiesen.

### 6.1.4.1 Poisson-Verteilung

#### (i) Auftreten

Die Poisson-Verteilung ist eine typische Verteilung einer sog. „Zählvariable“  $X$ . In vielen Fällen liefert sie ein adäquates Modell, z. B. bei den folgenden Erscheinungen:

- Anzahl der Verkehrsunfälle in einer bestimmten Region und in einer bestimmten Periode;

- Anzahl der von einer radioaktiven Substanz ausgestrahlten Partikel pro Zeiteinheit;
- Anzahl der Anrufe in der Telephonzentrale einer Firma an einem Tag;
- Anzahl der Jobs, welche in einem bestimmten Zeitintervall an einem Server eintreffen;
- Zunahme der Bakterien in einer Kultur in einem bestimmten Zeitintervall, etc.

Zwar könnte man alle diese Zufallsgrößen prinzipiell mit Hilfe der Binomialverteilung beschreiben, jedoch wird die Bestimmung ihrer Einzelwahrscheinlichkeiten durch folgende Besonderheiten erschwert, welche dem der Binomialverteilung zugrundeliegenden Bernoulli'schen Versuchsschema innewohnen:

- Die Anzahl der Versuche ist sehr groß (im 2. Beispiel die Anzahl der am Zerfallsprozeß beteiligten Atomkerne);
- Bei einer Serie von  $n$  Versuchen ist die Wahrscheinlichkeit  $p_n = p(A)$  des interessierenden Ereignisses  $A$  in jedem einzelnen Versuch sehr klein<sup>2</sup> (im 2. Beispiel die Wahrscheinlichkeit für den Zerfall eines einzelnen Atomkerns im betrachteten Zeitintervall).

### (ii) Voraussetzungen und Eigenschaften

- (1) Die Anzahl der Versuchsausgänge in einem bestimmten Zeitintervall oder in einem bestimmten Raumgebiet ist unabhängig von derjenigen in einem dazu disjunkten Zeitintervall oder in einem dazu disjunkten Raumgebiet. Man sagt, ein Poisson-Prozeß habe kein „Gedächtnis“.
- (2) Die Wahrscheinlichkeit für das Auftreten eines Einzelereignisses in einem sehr kurzen Zeitintervall oder in einem sehr kleinen Raumgebiet ist proportional zur Länge des Zeitintervalles oder dem Volumen des Raumgebietes. Sie ist unabhängig von der Anzahl der Versuchsausgänge außerhalb dieses Zeitintervalles oder dieses Raumgebietes.
- (3) Die Wahrscheinlichkeit dafür, daß in einem sehr kurzen Zeitintervall oder in einem sehr kleinen Raumgebiet mehr als ein Ereignis auftritt, ist vernachlässigbar.

### (iii) Dichtefunktion

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

Dabei ist  $\lambda$  eine beliebige positive reelle Zahl. Die Dichtefunktion heißt auch *Poisson-Dichte*.

---

<sup>2</sup>Die Poisson-Verteilung wird aus diesem Grunde häufig auch als Gesetz der „seltenen“ Ereignisse bezeichnet.

**(iv) Erwartungswert und Varianz**

$$E(X) = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda} \lambda^x}{x!} = \lambda$$

$$V(X) = \sum_{x=0}^{\infty} x^2 \cdot \frac{e^{-\lambda} \lambda^x}{x!} - \lambda^2 = \lambda$$

**(v) Grenzwerteigenschaft**

Die Poisson-Verteilung ist der Grenzwert einer Binomialverteilung in folgendem Sinne:

Sei  $X$  eine binomialverteilte Zufallsvariable mit der Dichtefunktion  $b(\cdot; n, p)$ . Falls für  $n \rightarrow \infty$  und  $p \rightarrow 0$  der Erwartungswert  $\mu = np$  konstant bleibt, so gilt

$$b(x; n, p) \longrightarrow p(x; \mu).$$

**Beispiel 6.8**

Bei der Herstellung von Gegenständen aus Glas können Blasen auftreten, welche dazu führen, daß das Produkt für den Verkauf ungeeignet ist. Bekannt sei, daß im Mittel eins von tausend produzierten Stücken eine oder mehrere Blasen habe.

Wie groß ist die Wahrscheinlichkeit, daß sich in einer Zufallsstichprobe von achttausend Stücken weniger als sieben defekte Teile befinden?

**Lösung**

Dieser Versuch ist im wesentlichen ein Experiment, welches von einer binomialverteilten Zufallsvariable mit  $n = 8\,000$  und  $p = 0.001$  gesteuert ist. Da  $n$  recht groß und  $p$  sehr klein ist, kann dieser Versuch durch einen Poisson-Prozeß mit dem Parameter

$$\mu = np = 8\,000 \cdot 0.001 = 8$$

modelliert werden. Für die Zufallsvariable

$$X \quad :\iff \quad \text{Anzahl der Blasen}$$

erhält man dann näherungsweise

$$P(X < 7) = \sum_{x=0}^6 b(x; 8\,000, 0.001)$$

$$\approx \sum_{x=0}^6 p(x; 8) = \sum_{x=0}^6 \frac{8^x}{x!} e^{-8}$$

$$= 0.3134 = 31,34\%.$$

## 6.1.4.2 Geometrische Verteilung

## (i) Auftreten

Binomialverteilte Zufallsvariablen lassen sich interpretieren als die Anzahl der Erfolge bei  $n$ -maliger unabhängiger Durchführung eines Bernoulli-Experimentes. Häufig interessiert man sich nun für die Anzahl der Versuche, die mit der Erfolgswahrscheinlichkeit  $p$  ( $p \in (0, 1)$  geeignet) durchgeführt werden müssen, bis eine *feste* Anzahl von Erfolgen auftritt. Anstatt die Wahrscheinlichkeit von  $x$  Erfolgen in  $n$  Versuchen zu bestimmen, ist man nun an der Wahrscheinlichkeit interessiert, daß der  $k$ -te Erfolg im  $x$ -ten Versuch auftritt. Zufallsvariable, die derartige Versuche beschreiben, heißen *negativ binomialverteilte Zufallsvariable*. Für eine negativ binomialverteilte Zufallsvariable ist  $W(X) = \{k, k + 1, k + 2, \dots\}$ .

## (ii) Dichtefunktion der negativen Binomialverteilung

Voneinander unabhängige und wiederholt ausgeführte Versuche mögen mit Wahrscheinlichkeit  $p$  zum Erfolg und mit Wahrscheinlichkeit  $q = 1 - p$  zum Mißerfolg führen. Dann hat die Zufallsvariable

$X \iff$  Anzahl der Versuche, bei denen  
im  $k$ -ten Versuch ein Erfolg vorliegt

die Dichtefunktion

$$f(x) := b^*(x; k, p) := \begin{cases} \binom{x-1}{k-1} p^k q^{x-k} & , \quad (x \in \{k, k+1, k+2, \dots\}) \\ 0 & , \quad \text{sonst} \end{cases}$$

Die negativ binomialverteilte Zufallsvariable hat ihren Namen daher, daß jeder Term in der Binomialentwicklung von  $p^k(1 - q)^{-k}$  zu dem entsprechenden Wert von  $b^*(x; k, p)$  für  $x \in \{k, k + 1, k + 2, \dots\}$  korrespondiert.

## (iii) Geometrische Verteilung

Ein Spezialfall der negativen Binomialverteilung liegt für  $k = 1$  vor und heißt *geometrische Verteilung*. Hier liegt eine Wahrscheinlichkeitsverteilung vor für die Anzahl der erforderlichen Versuche, bis zum *ersten* Mal Erfolg eintritt.

## (iv) Dichtefunktion der geometrischen Verteilung

$$f(x) := b^*(x; k = 1, p) := g(x; p) := \begin{cases} pq^{x-1} & , \quad (x \in \mathbb{N}) \\ 0 & , \quad \text{sonst} \end{cases}$$

## (v) Erwartungswert und Varianz der geometrischen Verteilung

$$E(X) = \frac{1}{p} \quad , \quad V(X) = \frac{q}{p^2}$$

**Beispiel 6.9**

Zu einer bestimmten Tageszeit mögen die Mitarbeiter in der Telephonzentrale einer Versicherungsgesellschaft nahe ihrer Kapazitätsgrenze arbeiten. Jeder neue Anruf landet dann entweder in einer Warteschleife oder ist sofort besetzt. Von Interesse ist hier die Anzahl der Versuche, welche ein Anrufer unternehmen muß, bis er eine Verbindung erhält.

Sei  $p := 0.05$  die (Erfolgs-)Wahrscheinlichkeit einer sofortigen Verbindung. Welches ist die Wahrscheinlichkeit dafür, im 5. Versuch eine erfolgreiche Verbindung zu erlangen?

**Lösung**

Benutzt man die geometrische Verteilung mit den Parametern  $x = 5$  und  $p = 0.05$ , so erhält man

$$P(X = 5) = g(x = 5; p = 0.05) = 0.05 \cdot 0.95^4 = 0.041 = 4,1\%.$$

**6.2 Stetige Wahrscheinlichkeitsverteilungen**

In diesem Abschnitt werden häufig auftretende stetige Verteilungen betrachtet; zur Definition einer stetigen Zufallsvariable s. Definition 4.4, p. 76.

**6.2.1 Gleichverteilung**

Unter der *Gleichverteilung über dem Intervall  $(a, b)$*  versteht man die Verteilung einer Zufallsvariable  $X$ , deren Dichtefunktion gegeben ist durch<sup>3</sup>

$$f(x) := f(x; a, b) := \frac{1}{b-a} \cdot \chi_{(a,b)}(x) := \begin{cases} \frac{1}{b-a} & , \quad x \in (a, b) \\ 0 & , \quad \text{sonst} \end{cases}.$$

Die Dichte dieser Verteilung ist also nur über dem Intervall  $(a, b)$  von Null verschieden, und sie hat dort den Wert  $\frac{1}{b-a}$ . Üblich sind auch die Begriffe *Rechteckverteilung* oder *uniforme Verteilung*. Vollkommen analog läßt sich die Gleichverteilung über dem abgeschlossenen Intervall  $[a, b]$  definieren.

Der Erwartungswert und die Varianz einer stetigen gleichverteilten Zufallsvariable  $X$  sind gegeben durch

$$E(X) = \frac{1}{b-a} \int_a^b u \, du = \frac{a+b}{2}$$

$$V(X) = \frac{1}{b-a} \int_a^b u^2 \, du - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

---

<sup>3</sup> $\chi_{\mathcal{M}}$  bezeichne, wie in der mathematischen Literatur üblich, die sog. *charakteristische Funktion* der Menge  $\mathcal{M}$ :  $\chi_{\mathcal{M}}(x) := \begin{cases} 1 & , \quad x \in \mathcal{M} \\ 0 & , \quad x \notin \mathcal{M} \end{cases}$

**Aufgabe**

Warum läßt sich auf einem unbeschränkten Intervall, etwa auf  $(-\infty, 0)$ ,  $(0, \infty)$  oder  $\mathbb{R} = (-\infty, \infty)$ , keine stetige Gleichverteilung definieren?

**6.2.2 Normalverteilung**

Die mit Abstand wichtigste stetige Verteilung bzw. Verteilungsart ist die sog. *Gauß-Verteilung* oder *Normalverteilung*.

**Definition 6.5**

Eine Zufallsvariable  $X$  mit Wertebereich  $W(X) = \mathbb{R}$  heißt *normalverteilt mit den Parametern  $\mu$  und  $\sigma$*  oder auch *Gauß-verteilt*, wenn ihre Dichtefunktion  $f$  gegeben ist durch (Abb. 6.3)

$$f(x) := n(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (x \in \mathbb{R}).$$

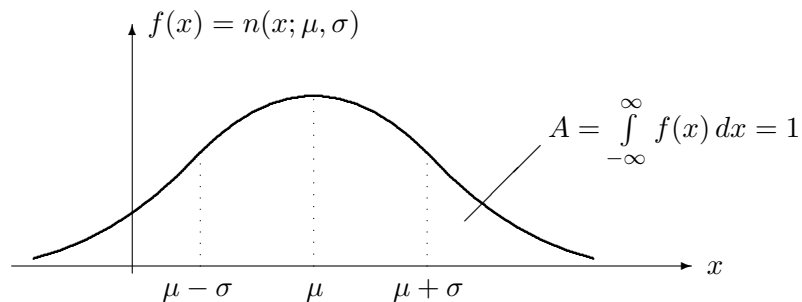


Abbildung 6.3: Dichtefunktion einer normalverteilten Zufallsvariable

Die Normalverteilung spielt eine zentrale Rolle in der Statistik. Die Erfahrung lehrt, daß die Normalverteilung ein adäquates Modell für viele Zufallsvariablen darstellt, die bei Experimenten und Beobachtungen in der Praxis auftreten.

**Beispiele**

- Meßfehler bei physikalischen Messungen, die von „zufälligen“ Ungenauigkeiten der Meßapparatur herrühren, also nicht durch apparaturbedingte systematische Fehler erzeugt werden.
- Oft ist die Abweichung von der Normgröße bei der Produktion von Artikeln durch eine Maschine normalverteilt.

**Bemerkung**

- (i) Der Funktionsgraph  $\Gamma_f$  von  $f$  ist symmetrisch in bezug auf die vertikale Achse  $x = \mu$ .

- (ii) An der Stelle  $x = \mu$  hat  $f$  ein globales Maximum, an den Stellen  $x = \mu \pm \sigma$  jeweils einen Wendepunkt.

$$(iii) \quad \int_{-\infty}^{\infty} n(x; \mu, \sigma) dx = 1 \quad (\mu \in \mathbb{R}, \sigma \in (0, \infty))$$

- (iv) Ist  $X$  eine stetige Zufallsvariable mit einer Normalverteilung als Dichtefunktion, so gilt

$$E(X) = \mu \quad , \quad V(X) = \sigma^2.$$

- (v) Eine Normalverteilung ist durch die beiden Kenngrößen  $\mu$  und  $\sigma$  eindeutig festgelegt.
- (vi) Die Verteilungsfunktion der allgemeinen Normalverteilung besitzt die folgende Integraldarstellung:

$$F(x) := N(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^x e^{-\frac{1}{2} \left(\frac{t-\mu}{\sigma}\right)^2} dt \quad (x \in \mathbb{R}).$$

- (vii) *Beweis* von (i) - (vi): Hausaufgabe!

### Problem

Sei  $X$  eine normalverteilte Zufallsvariable. Wie groß ist die Wahrscheinlichkeit, daß  $X$  einen Wert im Intervall  $(x_1, x_2]$  annimmt?

**Lösung** 
$$P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} n(x; \mu, \sigma) dx$$

Die Frage bleibt, wie man dieses Integral denn nun ausrechnet, also etwa eine Stammfunktion zu  $n(\cdot; \mu, \sigma)$  bestimmt. Nun ist bekannt, daß die Funktion

$$\exp(x; a) := e^{-ax^2} \quad (x \in \mathbb{R})$$

für keine Zahl  $a \neq 0$  eine elementare Stammfunktion besitzt.

Man muß das Integral also numerisch lösen oder auf tabellierte Werte zurückgreifen. Da man schlecht für jede Kombination von Parametern  $\mu$  und  $\sigma$  eine Tabelle halten kann, tut man das für die ausgezeichnete Kombination  $\mu = 0$ ,  $\sigma = 1$  und transformiert alle anderen Normalverteilungen auf diese.

## 6.2.2.1 Standard-Normalverteilung

**Definition 6.6**

Die Normalverteilung mit den Parametern  $\mu = 0$  und  $\sigma = 1$  heißt *Standardnormalverteilung*, oder auch *standardisierte Normalverteilung*. Ihre Dichtefunktion ist (Abb. 6.4)

$$f(x) := \phi(x) := n(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} \quad (x \in \mathbb{R}).$$

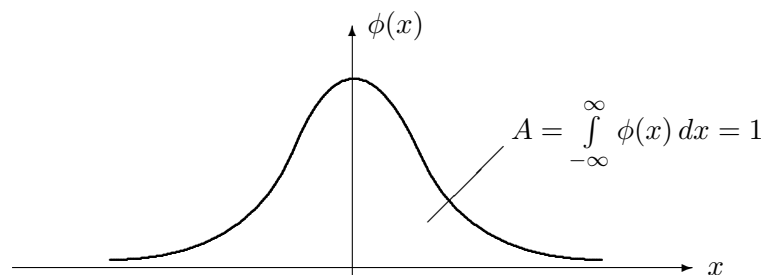


Abbildung 6.4: Dichtefunktion einer standardnormalverteilten Zufallsvariable mit der zugehörigen Verteilungsfunktion

$$F(x) := \Phi(x) := N(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt \quad (x \in \mathbb{R}).$$

**Satz 6.2**

Sei  $X$  eine normalverteilte Zufallsvariable mit den Parametern  $\mu := \mu_X$  und  $\sigma^2 := \sigma_X^2$ . Seien  $a, b \in \mathbb{R}$  mit  $a \neq 0$  und sei

$$Y := g(X) := aX + b.$$

Dann ist  $Y$  ebenfalls normalverteilt mit den Parametern  $\mu_Y = a\mu + b$  und  $\sigma_Y^2 = a^2\sigma^2$ .

*Beweis*

Sei  $a > 0$ . Dann sind die folgenden Ereignisse äquivalent:

$$Y \leq y \quad \iff \quad X = \frac{Y - b}{a} \leq \frac{y - b}{a},$$

und letzteres hat die Wahrscheinlichkeit

$$P\left(X \leq \frac{y - b}{a}\right) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\frac{y-b}{a}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt.$$



Mit der Variablensubstitution

$$t := \frac{u-b}{a}, \quad dt = \frac{du}{a}$$

geht dieses Integral über in das Integral

$$P(Y \leq y) = \frac{1}{\sqrt{2\pi} a \sigma} \int_{-\infty}^y e^{-\frac{1}{2} \frac{(u-(a\mu+b))^2}{a^2 \sigma^2}} du.$$

Das zeigt, daß auch  $Y$  normalverteilt ist mit den Parametern  $\mu_Y = a\mu + b$  und  $\sigma_Y^2 = a^2 \sigma^2$ . Eine analoge Rechnung führt auch bei  $a < 0$  zum Ziel.

$Y$  ist eine lineare Funktion von  $X$ , daher gilt für den Erwartungswert und die Varianz nach Satz 5.6

$$E(Y) = a\mu + b, \quad V(Y) = a^2 \sigma^2.$$

Auf die normalverteilte Zufallsvariable  $X$  wird nun die Standard-Transformation angewandt. Das ergibt dann den folgenden Satz:

### Satz 6.3

Sei  $X$  eine normalverteilte Zufallsvariable mit den Parametern  $E(X) = \mu := \mu_X$  und  $V(X) = \sigma^2 := \sigma_X^2$ . Sei

$$Z := \frac{X - \mu}{\sigma} = \frac{1}{\sigma} \cdot X - \frac{\mu}{\sigma}.$$

Dann ist  $Z$  standardnormalverteilt, hat also die Parameter  $E(Z) = \mu_Z = 0$  und  $V(Z) = \sigma_Z^2 = 1$ .

Die Transformation aus Satz 6.3 bildet also jede normalverteilte Zufallsvariable  $X$  auf die standardnormalverteilte Zufallsvariable  $Z$  ab; sie heißt daher *Standardtransformation*.

Für eine normalverteilte Zufallsvariable  $X$  genügt es daher, die spezielle Funktion  $\Phi(\cdot) = N(\cdot, \mu = 0, \sigma^2 = 1)$  zu kennen, alle anderen Funktionen  $N(\cdot; \mu, \sigma)$  werden nicht benötigt.

### Satz 6.4

Sei  $X$  eine normalverteilte Zufallsvariable mit den Parametern  $E(X) = \mu := \mu_X$  und  $V(X) = \sigma^2 := \sigma_X^2$ . Sei

$$\Phi(x) := N(x; \mu = 0, \sigma^2 = 1).$$

Dann gilt

$$P(a < X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right). \quad (6.1)$$

Diese Beziehung gestattet es, Wahrscheinlichkeiten für *beliebige* Normalverteilungen auf solche für die Standardnormalverteilung umzurechnen, s. Abb. 6.5 und Abb. 6.6.

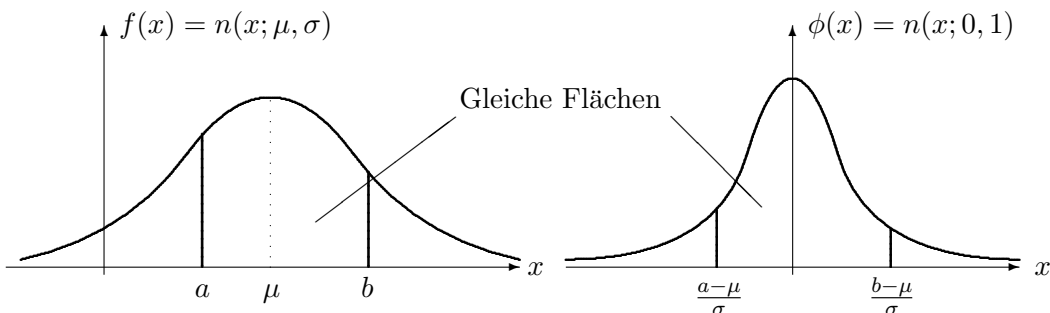


Abbildung 6.5: Dichtefunktion und Wahrscheinlichkeiten einer normalverteilten Zufallsvariable

Abbildung 6.6: Dichtefunktion und Wahrscheinlichkeiten einer standardnormalverteilten Zufallsvariable

*Beweis*

Für die standardnormalverteilte Zufallsvariable  $Z := \frac{X - \mu}{\sigma}$  gilt:

$$\begin{aligned} a < X = \sigma Z + \mu &\iff \frac{a - \mu}{\sigma} < Z \\ b \geq X = \sigma Z + \mu &\iff \frac{b - \mu}{\sigma} \geq Z \end{aligned}$$

$$\longrightarrow P(a < X \leq b) = P\left(\frac{a - \mu}{\sigma} < Z \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

Die Funktionswerte der Verteilungsfunktion  $\Phi = N(\cdot; 0, 1)$  der standardnormalverteilten Zufallsvariable  $Z$  sind tabelliert. Die Berechnung des Funktionswertes  $\Phi(-x)$  erfolgt dabei anhand der Formel (Abb. 6.7)

$$\Phi(-x) = 1 - \Phi(x) \quad (x \in \mathbb{R}).$$

### Beispiel 6.10

Die Zufallsvariable  $X$  sei normalverteilt mit dem Erwartungswert  $E(X) = 10$  und der Varianz  $V(X) = 4$ .

Welcher prozentuale Anteil der Werte liegt dann im Intervall  $[5, 12]$ ?

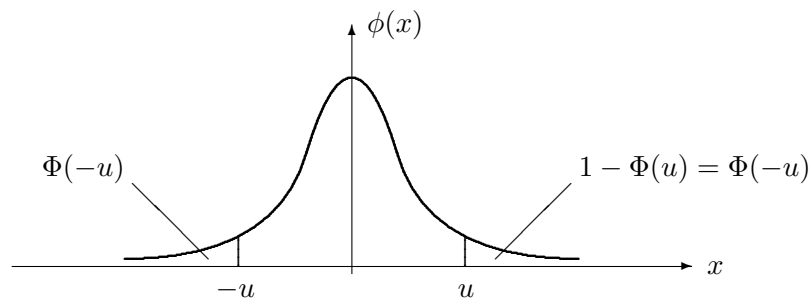


Abbildung 6.7: Dichtefunktion und Wahrscheinlichkeiten einer standardnormalverteilten Zufallsvariable

### Lösung

$$\begin{aligned}
 P(5 < X \leq 12) &= P\left(\frac{5-10}{2} < Z \leq \frac{12-10}{2}\right) \\
 &= P\left(-\frac{5}{2} < Z \leq 1\right) \\
 &= \Phi(1) - \Phi\left(-\frac{5}{2}\right) \\
 &= \Phi(1) - \left[1 - \Phi\left(\frac{5}{2}\right)\right] \\
 &= 0.8413 + 0.9938 - 1 \\
 &= 0.8351 = 83.5\%
 \end{aligned}$$

### Beispiel 6.11

Eine Maschine packe Lebkuchenpakete ab, deren Nettomasse normalverteilt sei mit dem Erwartungswert  $E(X) = 248$  g und der Varianz  $V(X) = (4 \text{ g})^2$ .

Für eine Lieferung sei ein Toleranzbereich von  $(250 \pm 5)$  g vorgesehen. Welcher Anteil ist dann außerhalb dieses Bereiches zu erwarten (der sog. *Ausschußanteil*, s. Abb. 6.8)?

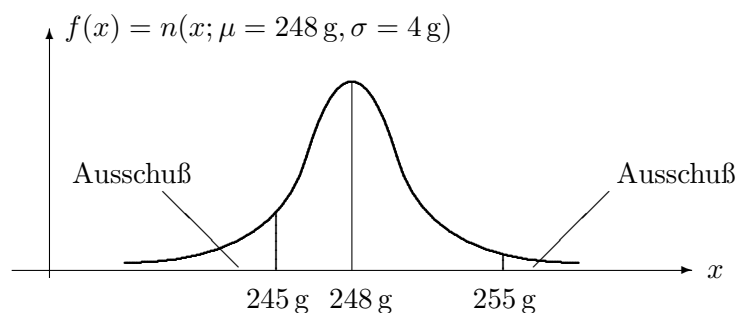


Abbildung 6.8: Ausschußanteil bei der Verpackung von Lebkuchenpaketen

**Lösung**

$$\begin{aligned}
P(245 \text{ g} < X \leq 255 \text{ g}) &= P\left(\frac{245 \text{ g} - 248 \text{ g}}{4 \text{ g}} < Z \leq \frac{255 \text{ g} - 248 \text{ g}}{4 \text{ g}}\right) \\
&= \Phi\left(\frac{7}{4}\right) - \Phi\left(-\frac{3}{4}\right) \\
&= \Phi(1.75) - [1 - \Phi(0.75)] \\
&= 0.9599 + 0.7734 - 1 \\
&= 0.7333 = 73.33\% = 1 - 26.67\%
\end{aligned}$$

Genauer:

- $P(X \leq 245 \text{ g}) = \Phi(-0.75) = 0.2266 = 22.7\%$
- $P(X > 255 \text{ g}) = 1 - P(X \leq 255 \text{ g}) = 1 - [1 - \Phi(1.75)] = 0.0401 = 4\%$

Ein Anteil von 4% ist zu schwer – was den Abnehmer vermutlich nicht stören würde –, ein Anteil von 22.7% ist zu leicht. Beide Anteile zusammen ergeben 26.7%.

**Beispiel 6.12**

Eine Maschine packe Lebkuchenpakete ab, deren Nettomasse normalverteilt sei mit dem Erwartungswert  $E(X) = 1000 \text{ g}$  und der Varianz  $V(X) = (5 \text{ g})^2$ . Für eine Lieferung sei ein Toleranzbereich von  $(1000 \pm 15) \text{ g}$  vorgesehen. Der Gesetzgeber schreibe vor, daß 95% aller Pakete ein Gewicht von mindestens 985 g aufweisen müssen.

- (i) Welcher Anteil der Pakete enthält mindestens 985 g?
- (ii) Auf welchen Sollwert läßt sich die Verpackungseinstellung bei dieser Genauigkeit absenken, damit immer noch 95% aller Pakete ein Gewicht von mindestens 985 g aufweisen; s. Abb. 6.9?

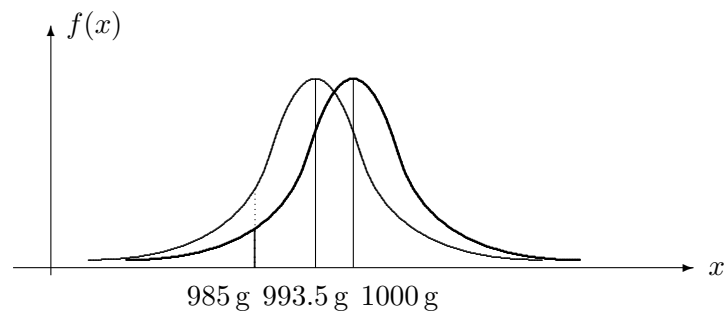


Abbildung 6.9: Absenkung des Sollwertes bei der Verpackung von Lebkuchenpaketen

**Lösung**

(i) Anteil der Pakete, die mindestens 985 g enthalten:

$$\begin{aligned}
P(X > 985 \text{ g}) &= 1 - P(X \leq 985 \text{ g}) \\
&= 1 - \Phi\left(\frac{985 - 1000}{5}\right) \\
&= 1 - \Phi(-3) = 1 - [1 - \Phi(3)] = \Phi(3) \\
&= 0.9987 = 99.87\%
\end{aligned}$$

(ii) Absenkung des Verpackungssollwertes:

$$\begin{aligned}
P(X > 985 \text{ g}) &= 1 - P(X \leq 985 \text{ g}) \\
&= 1 - \Phi\left(\frac{985 - m}{5}\right) \\
&\stackrel{!}{\geq} 0.95 \\
&\Leftrightarrow \\
\Phi\left(\frac{985 - m}{5}\right) &\leq 1 - 0.95 = 0.05 \\
&\stackrel{(1)}{\Leftrightarrow} \\
\frac{985 - m}{5} &\leq \Phi^{-1}(0.05) \stackrel{(2)}{=} -\Phi^{-1}(1 - 0.05) = -1.645 \\
&\Leftrightarrow \\
m &\geq 985 + 8.225 = 993.225
\end{aligned}$$

wobei in (1) benutzt wurde, daß mit  $\Phi$  auch die Umkehrfunktion  $\Phi^{-1}$  streng monoton wachsend ist. (2) kommt zustande, weil für die Werte  $u$  der Umkehrfunktion  $\Phi^{-1}$ , also die Argumente der Verteilungsfunktion  $\Phi$  selbst, wegen der Symmetrie von  $\Phi$  die folgende Beziehung gilt (Abb. 6.10):

$$\begin{aligned}
\Phi(u) &= p && \longleftrightarrow && u &= \Phi^{-1}(p) \\
\Phi(-u) &= 1 - p && \longleftrightarrow && -u &= \Phi^{-1}(1 - p) \\
\longrightarrow & u_p := \Phi^{-1}(p) &= & -\Phi^{-1}(1 - p) &= & -u_{1-p}
\end{aligned}$$

Der vom Gesetzgeber zugelassene Ausschußanteil wird in diesem Beispiel also auch dann nicht unterschritten, wenn man den Sollwert der Maschine auf 993.225 g einstellt, statt auf den geforderten Wert von 1000 g.

**Bemerkung**

Aus der Tabelle der Standardnormalverteilung liest man die folgende Faustregel für eine  $N(\mu, \sigma^2)$ -verteilte Zufallsvariable  $X$  ab:

$$\begin{aligned}
P(\mu - \sigma < X \leq \mu + \sigma) &\approx 0.680 = 68.0\% \\
P(\mu - 2\sigma < X \leq \mu + 2\sigma) &\approx 0.955 = 95.5\% \\
P(\mu - 3\sigma < X \leq \mu + 3\sigma) &\approx 0.997 = 99.7\%
\end{aligned}$$

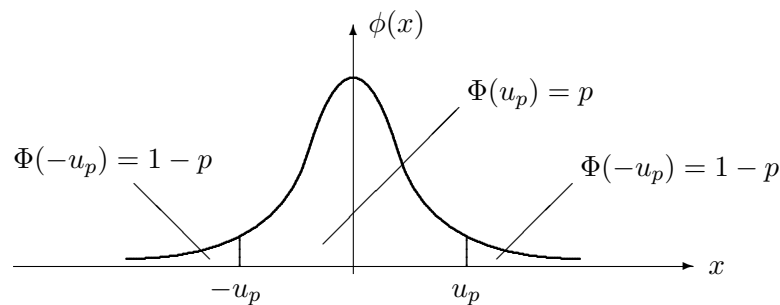


Abbildung 6.10: Dichtefunktion und Wahrscheinlichkeiten einer standardnormalverteilten Zufallsvariable

Somit liegen etwa 99.7% aller Werte einer normalverteilten Zufallsvariable in einem Intervall der Breite  $\pm 3\sigma$  um den Erwartungswert  $\mu$  herum.

### Aufgabe

Man schätze die Wahrscheinlichkeiten

$$\begin{aligned} P(\mu - \sigma < X \leq \mu + \sigma) \\ P(\mu - 2\sigma < X \leq \mu + 2\sigma) \\ P(\mu - 3\sigma < X \leq \mu + 3\sigma) \end{aligned}$$

mit Hilfe der Tschebyschev-Ungleichung (Satz 5.4) ab.

#### 6.2.2.2 Quantile der Standard-Normalverteilung

In den letzten Beispielen sind u.a. schon folgende Probleme aufgetreten:

eine normalverteilte Zufallsvariable  $X$  soll mit einer vorgegebenen Wahrscheinlichkeit  $p$  einen Wert

- unterhalb,
- oberhalb oder
- innerhalb

bestimmter Grenzen (Schranken) annehmen. Aufgaben dieser Art treten im Zusammenhang mit Parameterschätzungen und statistischen Prüfverfahren (Signifikanztests) auf. Bei einer Abgrenzung nach oben beispielsweise muß dann die zur vorgegebenen Wahrscheinlichkeit  $p$  gehörende Schranke  $x_p$  bestimmt werden ( $F_X$  : Verteilungsfunktion von  $X$ ):

$$P(X \leq x_p) = F_X(x_p) = p$$

**Definition 6.7**

Seien  $X$  eine Zufallsvariable und  $p \in (0, 1)$ . Ein Wert  $x_p$  mit

$$P(X \leq x_p) \geq p \quad \text{und} \quad P(X \geq x_p) \geq 1 - p$$

heißt  $p$ -Quantil der Zufallsvariable  $X$  (bzw.  $p$ -Quantil der Verteilungsfunktion  $F_X$  von  $X$ ).

Das  $p$ -Quantil heißt auch *Percentil der Ordnung  $p$* . Statt 0.1-Quantil sagt man auch *10 %-Quantil*. Die 25 %- bzw. 75 %-Quantile heißen auch *Quartile*, das 50 %-Quantil *Median* der Verteilung.

**Satz 6.5**

Die Ungleichungen von Definition 6.7 sind äquivalent zu

$$p \leq F_X(x_p) \leq p + P(X = x_p).$$

*Beweis*

$$\begin{aligned} \text{a) } P(X \leq x_p) &= F_X(x_p) \\ \\ \text{b) } P(X \geq x_p) &= 1 - P(X < x_p) \\ &= 1 - P(X \leq x_p) + P(X = x_p) \\ &= 1 - F_X(x_p) + P(X = x_p) \\ &\geq 1 - p \\ &\iff \\ F_X(x_p) &\leq p + P(X = x_p) \end{aligned}$$

Für eine stetige Zufallsvariable, insbesondere also für die (stetige) Normalverteilung, vereinfacht sich die obige Bedingung wegen

$$P(X = a) = 0 \quad (a \in \mathbb{R})$$

zu der Bedingung

$$F_X(x_p) = p = P(X \leq x_p).$$

**Bemerkung**

Für eine stetige Zufallsvariable  $X$  ist die Verteilungsfunktion  $F_X$  stetig und streng monoton wachsend, also injektiv und damit umkehrbar. Das  $p$ -Quantil ist somit das Urbild von  $p$  unter  $F_X$ :

$$F_X(x_p) = p \quad \iff \quad x_p = F_X^{-1}(p)$$

Die Quantile der Standardnormalverteilung  $Z$  lassen sich prinzipiell aus der Wertetabelle der Verteilungsfunktion  $\Phi = F_Z$  ermitteln. Da solche Probleme in

den Anwendungen jedoch häufig auftreten, ist es sinnvoll, die gängigen Werte in einer zweiten Tabelle zu sammeln.

**Beispiel 6.13** (einseitige Abgrenzung nach oben)

Die standardnormalverteilte Zufallsvariable  $Z$  soll mit der Wahrscheinlichkeit  $p = 0.95$  einen Wert aus dem Intervall  $(-\infty, c)$  annehmen. Gesucht ist also das zum Wert  $p = 0.95$  gehörende Quantil  $c = x_p$  der Standardnormalverteilung:

$$P(Z \leq c) = \Phi(c) = 0.95 \quad \longrightarrow \quad c = x_{0.95} = 1.645$$

**Beispiel 6.14** (einseitige Abgrenzung nach unten)

Die Werte der normalverteilten Zufallsvariable  $X$  mit  $\mu = E(X) = 8$  und  $\sigma^2 = V(X) = 16$  sollen mit der Wahrscheinlichkeit  $p = 0.9$  oberhalb der unbekannt Schranke  $x_p$  liegen:

$$\begin{aligned} P(X > x_p) &= 1 - P(X \leq x_p) \\ &= 1 - P\left(Z \leq \frac{x_p - 8}{\sqrt{16}}\right) \\ &= 1 - \Phi\left(\frac{x_p - 8}{4}\right) \\ &\stackrel{!}{=} 0.9 \\ &\iff \\ \Phi\left(\frac{x_p - 8}{4}\right) &= 1 - 0.9 = 0.1 \\ &\iff \\ \frac{x_p - 8}{4} &= \Phi^{-1}(0.1) = -\Phi^{-1}(1 - 0.1) = -1.282 \\ &\iff \\ x_p &= 8 - 4 \cdot 1.282 = 2.872 \end{aligned}$$

**Beispiel 6.15** (zweiseitige Abgrenzung)

Eine zweiseitige Abgrenzung läßt sich sinnvollerweise nur für einen symmetrischen Bereich um den Erwartungswert  $\mu$  einer normalverteilten Zufallsvariable durchführen: man sucht hierbei ein symmetrisches Intervall mit  $\mu$  als Mittelpunkt, in welchem die Werte der Zufallsvariable mit einer vorgegebenen Wahrscheinlichkeit liegen.

Seien also  $X$  eine normalverteilte Zufallsvariable,  $h_p > 0$  und  $p \in (0, 1)$ . Gesucht ist das Intervall  $[\mu - h_p, \mu + h_p]$ , in welchem die Werte von  $X$  mit der



Wahrscheinlichkeit  $p$  liegen, etwa mit  $p := 0.95$ :

$$\begin{aligned}
 P(\mu - h_p < X \leq \mu + h_p) &= P\left(\frac{(\mu - h_p) - \mu}{\sigma} < Z \leq \frac{(\mu + h_p) - \mu}{\sigma}\right) \\
 &= \Phi\left(\frac{h_p}{\sigma}\right) - \Phi\left(-\frac{h_p}{\sigma}\right) \\
 &= \Phi\left(\frac{h_p}{\sigma}\right) - \left[1 - \Phi\left(\frac{h_p}{\sigma}\right)\right] \\
 &= 2\Phi\left(\frac{h_p}{\sigma}\right) - 1 \\
 &\stackrel{!}{=} p \\
 &\iff \\
 \Phi\left(\frac{h_p}{\sigma}\right) &= \frac{1+p}{2} \\
 &\iff \\
 \frac{h_p}{\sigma} &= \Phi^{-1}\left(\frac{1+p}{2}\right) \\
 &= \Phi^{-1}\left(\frac{1+0.95}{2}\right) = \Phi^{-1}(0.975) = 1.960 \\
 &\iff \\
 h_p &= 1.960 \sigma
 \end{aligned}$$

Damit ist das gesuchte Intervall gegeben durch (Abb. 6.11)

$$\mathcal{I} := [\mu - h_p, \mu + h_p] = [\mu - 1.960 \sigma, \mu + 1.960 \sigma].$$

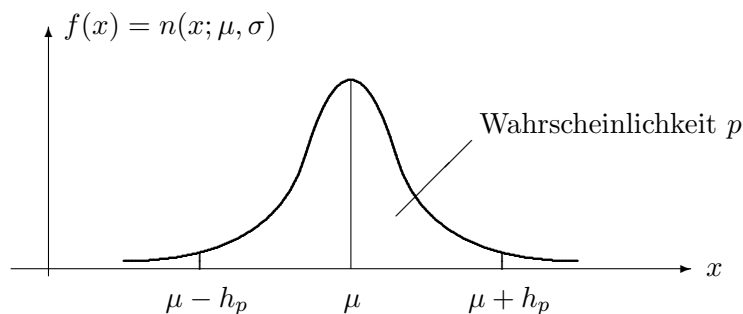


Abbildung 6.11: Zweiseitige Abgrenzung bei einer Normalverteilung

### 6.2.2.3 Logarithmische Normalverteilung

Viele Verteilungen laufen als *positiv schiefe* oder *linkssteile Verteilungen* rechts flach aus, s. Abb. 6.12. Ein bekanntes Beispiel ist die Häufigkeitsverteilung der Einkommen in einem Lande: je weiter man nach oben steigt, desto dünner

wird die Luft. Außerdem kann das Einkommen als Merkmal einen bestimmten Schrankenwert (hier: Null) nicht unterschreiten und ist daher nach dieser Seite hin in seiner Variationsmöglichkeit gehemmt.

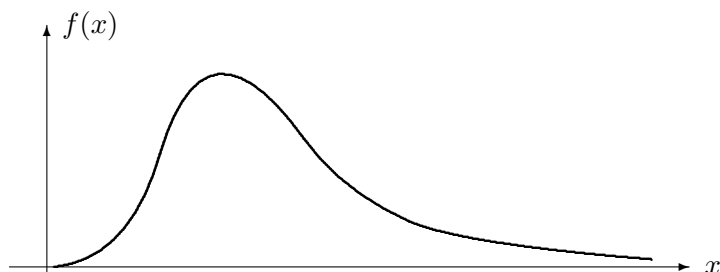


Abbildung 6.12: Positiv schiefe, linkssteile Verteilung

### Beispiele 6.16

(i) Linkssteile Verteilungen:

Wartezeiten, Einkommen in Euro, Alter bei der Eheschließung

(ii) Rechtssteile Verteilungen:

Tragzeit bei Säugetieren, Kopfumfang von Neugeborenen

Besonders dann, wenn die Verteilung links durch den Wert Null begrenzt ist, kommt man durch Logarithmieren annähernd zu normalverteilten Werten. Durch das Logarithmieren wird der Bereich im Intervall  $(0, 1)$  in den Bereich  $(-\infty, 0)$  überführt und stark gestreckt, während der Bereich im Intervall  $(1, \infty)$  stark gestaucht wird.

Die Entstehung einer logarithmischen Normalverteilung, kurz: *Lognormalverteilung* genannt, läßt sich darauf zurückführen, daß viele Zufallsgrößen *multiplikativ* zusammenwirken. Dagegen kommt die Normalverteilung selbst durch *additives* Zusammenwirken vieler Zufallsgrößen zustande. Daher ist die Lognormalverteilung insbesondere bei Merkmalen aus der Biologie und der Wirtschaft vorherrschend, da sie multiplikative Wirkungen in additive transformiert.

### Definition 6.8

Eine stetige positive Zufallsvariable  $X$  heißt *logarithmisch normalverteilt* (*lognormal-verteilt*), wenn  $\ln X$  normalverteilt ist. Sie hat daher die Wahrscheinlichkeitsdichte

$$f(x) := \begin{cases} \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\ln x - \mu}{\sigma} \right)^2} & , \quad x > 0 \\ 0 & , \quad x \leq 0 \end{cases}$$

Es gilt

$$E(X) = e^{\mu + \sigma^2/2} \quad , \quad V(X) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1).$$

**Aufgabe**

Man leite die Verteilungsfunktion der logarithmischen Normalverteilung aus der Verteilungsfunktion einer „gewöhnlichen“ Normalverteilung her.

Hinweis Aus  $X = \ln Y \Leftrightarrow Y = e^X$  folgt

$$P(X \leq x = \ln y) = P(Y \leq y = e^x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x=\ln y} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt.$$

Jetzt führe man die Substitution  $t := \ln z$  durch.

**6.2.3 Approximation der Binomialverteilung durch die Normalverteilung**

In Abschnitt 6.1.4.1 wurde beschrieben, daß sich die (diskrete) Binomialverteilung durch die (diskrete) Poisson-Verteilung approximieren läßt, falls  $n$  hinreichend groß und  $p$  hinreichend klein sind, so daß das Produkt  $\mu := np$  konstant bleibt:

$$\begin{aligned} n &\longrightarrow \infty, & p &\longrightarrow 0, & \mu = np &\text{ konstant} \\ \implies & & b(x; n, p) &\longrightarrow p(x; \mu) & & (x \in \mathbb{R}) \end{aligned}$$

Manchmal läßt sich die (diskrete) Binomialverteilung auch gut durch die (stetige) Normalverteilung approximieren, nämlich dann, wenn die Anzahl  $n$  der unabhängigen Bernoulli-Versuche hinreichend groß ist. Den diesbezüglich genauen Sachverhalt beschreibt der nächste Satz.

**Satz 6.6** (Grenzwertsatz von Moivre-Laplace)

Sei  $\{X_n\}_{n \in \mathbb{N}}$  eine Folge von binomialverteilten Zufallsvariablen mit dem Erwartungswert  $\mu_n = E(X_n) = np$  und der Varianz  $\sigma_n^2 = V(X) = npq$ . Sei  $\{Z_n\}_{n \in \mathbb{N}}$  eine Folge von Zufallsvariablen, definiert durch

$$Z_n := \frac{X_n - \mu_n}{\sigma_n} = \frac{X_n - np}{\sqrt{npq}} \quad (n \in \mathbb{N}).$$

Dann ist die Folge  $\{Z_n\}_{n \in \mathbb{N}}$  asymptotisch normalverteilt, d.h. es gilt die Konvergenz  $\lim_{n \rightarrow \infty} Z_n = \Phi(\cdot)$  in folgendem Sinne:

$$\boxed{\lim_{n \rightarrow \infty} P\left(Z_n = \frac{X_n - np}{\sqrt{npq}} \leq x\right) = \Phi(x)} \quad (6.2)$$

für alle  $(x \in \mathbb{R})$

In Worten:

Wenn bei dem der Binomialverteilung zugrundeliegenden Bernoulli'schen Versuchsschema die Anzahl der Versuche gegen unendlich strebt, dann konvergiert die Verteilungsfunktion der standardisierten binomialverteilten Zufallsvariable gegen die Verteilungsfunktion der Standard-Normalverteilung. Man sagt, die Binomialverteilung sei *asymptotisch normalverteilt*.

### Bemerkung

Dieser Satz läßt sich dahingehend erweitern, daß die Glieder der Folge  $\{X_n\}_{n \in \mathbb{N}}$  von (identisch verteilten) Zufallsvariablen nicht notwendig mehr binomialverteilt zu sein brauchen. Er heißt dann *Zentraler Grenzwertsatz* oder *Zentraler Grenzwertsatz von Lindeberg und Levy*.

Auch schon dann, wenn  $n$  nicht so sehr groß ist, stellt die Normalverteilung eine ziemlich gute Approximation einer Binomialverteilung dar; s. dazu das folgende Beispiel.

### Beispiel 6.17

Man betrachte die Binomialverteilung  $b(\cdot; n, p)$  für  $n = 15$  und  $p = 0.4$ . Eine entsprechende Zufallsvariable hat

$$\begin{aligned} \text{den Erwartungswert} \quad \mu &= np = 6 \\ \text{und die Varianz} \quad \sigma^2 &= np(1-p) = 3.6 . \end{aligned}$$

Die entsprechende approximierende Normalverteilung ist in Abb. 6.13 zu sehen.

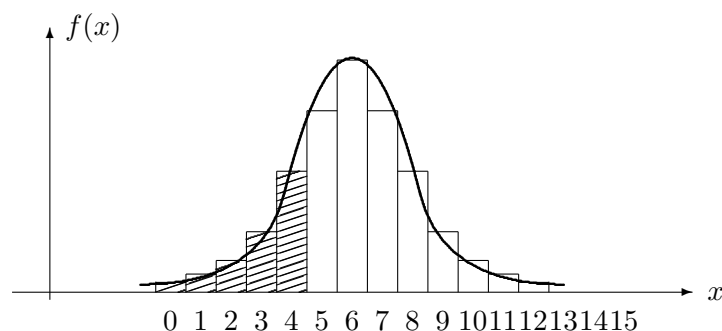


Abbildung 6.13: Approximation einer Binomialverteilung durch eine Normalverteilung

Sei nun  $X$  die binomialverteilte Zufallsvariable, dann gilt, etwa für den Wert  $X \leq 4$ :

(i) Korrekte Berechnung:

$$\begin{aligned}
 P(X \leq 4) &= \sum_{x=0}^4 b(x; n=15, p=0.4) \\
 &= \sum_{x=0}^4 \binom{15}{x} \cdot \left(\frac{4}{10}\right)^x \cdot \left(\frac{6}{10}\right)^{15-x} \\
 &= \left(\frac{6}{10}\right)^{15} \cdot \frac{1}{10^4} \cdot \left[ 1 \cdot 1 \cdot 6^4 + 15 \cdot 4 \cdot 6^3 + \frac{15 \cdot 14}{1 \cdot 2} \cdot 4^2 \cdot 6^2 \right. \\
 &\quad \left. + \frac{15 \cdot 14 \cdot 13}{1 \cdot 2 \cdot 3} \cdot 4^3 \cdot 6 + \frac{15 \cdot 14 \cdot 13 \cdot 12}{1 \cdot 2 \cdot 3 \cdot 4} \cdot 4^4 \right] \\
 &= 0.2172776 = 21.73\%
 \end{aligned}$$

(ii) Approximative Berechnung mit Hilfe der Normalverteilung:

$$\begin{aligned}
 P\left(z_1 = \frac{0-6}{\sqrt{3.6}} < Z \leq \frac{4-6}{\sqrt{3.6}} = z_2\right) &= P(-3.16 < Z \leq -1.05) \\
 &= \Phi(-1.05) - \Phi(-3.16) \\
 &= 1 - \Phi(1.05) - (1 - \Phi(3.16)) \\
 &= \Phi(3.16) - \Phi(1.05) \\
 &= 0.9992 - 0.8531 \\
 &= 0.1461 = 14.61\%
 \end{aligned}$$

Ein Vergleich mit (i) zeigt, daß dieses Ergebnis so richtig überzeugend nicht ist. Man kann es allerdings verbessern:

(iii) Bei der Approximation einer Binomialverteilung durch eine Normalverteilung sollte eine sog. *Stetigkeitskorrektur* erfolgen in dem Sinne, daß die Intervallgrenzen  $a$  und  $b$  jeweils um 0.5 nach außen hin verschoben werden. Das liegt daran, daß die Binomialverteilung diskret, die Normalverteilung hingegen stetig ist, cf. Abb. 6.13:

$$P(a < X \leq b) \approx \Phi\left(\frac{b+0.5-\mu}{\sigma}\right) - \Phi\left(\frac{a-0.5-\mu}{\sigma}\right).$$

Von daher wird jetzt die approximative Berechnung mit Hilfe der Normalverteilung unter Berücksichtigung der o.a. Stetigkeitskorrektur noch einmal wiederholt:

$$\begin{aligned}
 P\left(z_1 = \frac{0-0.5-6}{1.897} < Z \leq \frac{4+0.5-6}{1.897} = z_2\right) &= P(-3.43 < Z \leq -0.79) \\
 &= \Phi(-0.79) - \Phi(-3.43) \\
 &= 0.9997 - 0.7852 \\
 &= 0.2145 = 21.45\%
 \end{aligned}$$

Die Abweichung vom wahren Wert  $P(X \leq 4) = 0.2173$  ist kleiner als 1.3 %.

### Beispiel 6.18

Mit Hilfe unabhängiger Versuche ist die Wahrscheinlichkeit eines zufälligen Ereignisses näherungsweise durch seine relative Häufigkeit zu bestimmen. Mit dem Grenzwertsatz von Moivre-Laplace soll nun untersucht werden, wie groß die Anzahl der unabhängigen Versuche sein muß, damit mit einer Wahrscheinlichkeit von mindestens 99 % das Ergebnis mit einem Fehler behaftet ist, der kleiner als 0.05 ist.

Die Lösung dieser Aufgabe beruht auf der Tatsache, daß die relative Häufigkeit eines Ereignisses  $A$  wegen  $h_n(A) = \frac{X_n}{n}$  mit Hilfe der binomialverteilten Zufallsvariable  $X_n$  beschrieben wird. Sei  $p := P(A)$ , dann ist die Anzahl  $n$  der unabhängigen Versuche aus der Beziehung

$$P\left(\left|\frac{X_n}{n} - p\right| < 0.05\right) \geq 0.99$$

zu bestimmen. Dazu:

$$\begin{aligned} P\left(\left|\frac{X_n}{n} - p\right| < 0.05\right) &= P(|X_n - np| < 0.05n) \\ &= P\left(\left|\frac{X_n - np}{\sqrt{npq}}\right| < \frac{0.05n}{\sqrt{npq}}\right) \\ &= P\left(-\frac{\sqrt{n} \cdot 0.05}{\sqrt{p(1-p)}} < \frac{X_n - np}{\sqrt{np(1-p)}} < \frac{\sqrt{n} \cdot 0.05}{\sqrt{p(1-p)}}\right) \\ &\approx \Phi\left(\frac{\sqrt{n} \cdot 0.05}{\sqrt{p(1-p)}}\right) - \Phi\left(-\frac{\sqrt{n} \cdot 0.05}{\sqrt{p(1-p)}}\right) \\ &= 2\Phi\left(\frac{\sqrt{n} \cdot 0.05}{\sqrt{p(1-p)}}\right) - 1 \\ &\geq 2\Phi\left(\frac{\sqrt{n} \cdot 0.05}{\sqrt{0.25}}\right) - 1 \end{aligned}$$

Hierbei wurde, da  $p$  in der Regel unbekannt ist, die Abschätzung

$$p(1-p) = \frac{1}{4} - \left(p - \frac{1}{2}\right)^2 \leq \frac{1}{4}$$

sowie das monotone Wachstum von  $\Phi$  benutzt. Damit lautet eine hinreichende Bedingung zur Erfüllung der Aufgabenstellung

$$2\Phi\left(\frac{\sqrt{n} \cdot 0.05}{\sqrt{0.25}}\right) - 1 = 2\Phi\left(\frac{\sqrt{n} \cdot 0.05}{0.5}\right) - 1 \geq 0.99,$$

also

$$\Phi\left(\frac{\sqrt{n}}{10}\right) \geq 0.995.$$

Das 0.995-Quantil der Standardnormalverteilung ist gegeben durch

$$x_0 = \Phi^{-1}(0.995) = 2.576 \quad \Longleftrightarrow \quad \Phi(x_0) = 0.995.$$

Gesucht ist also die kleinste natürliche Zahl  $n$  mit der Eigenschaft

$$\frac{\sqrt{n}}{10} \geq 2.576 \quad \iff \quad n \geq 664$$

**Beispiel 6.19** (Galton'sches Brett)

Das *Galton'sche Brett* liefert eine experimentelle Bestätigung der Möglichkeit, die Binomialverteilung  $b(\cdot; n, p = 1/2)$  durch eine Normalverteilung zu approximieren.

Das Galton'sche Brett enthält  $n$  Reihen von Pflöcken s. Abb. 6.14. Eine durch die  $n$  Reihen herunterrollende Kugel geeigneter Größe trifft in jeder Reihe auf einen Pflock und wird mit der Wahrscheinlichkeit  $p = 1/2$  nach links oder nach rechts abgelenkt. Danach trifft sie auf einen Pflock der nächsten Reihe und so weiter, bis sie schließlich in einem der  $n + 1$  Fächer unterhalb der Pflockreihen landet.

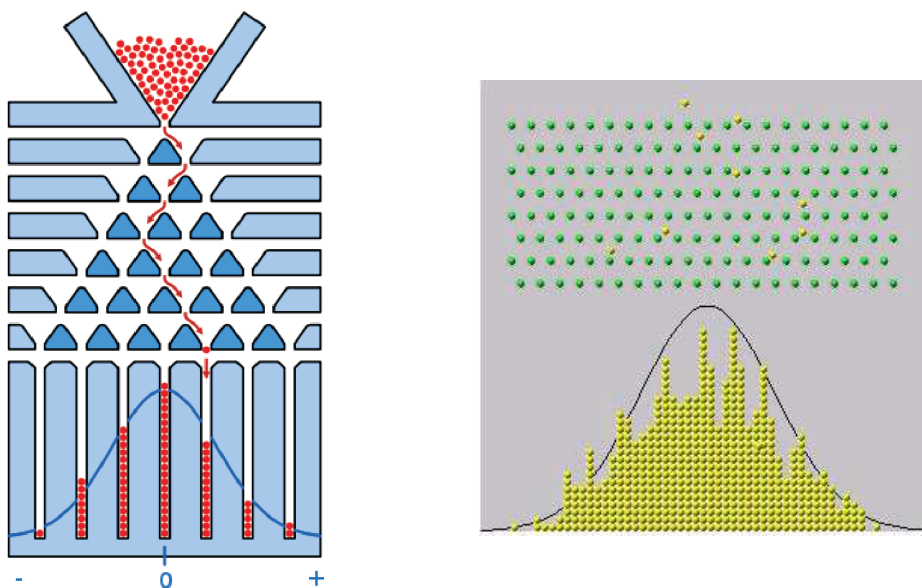


Abbildung 6.14: Galton'sches Brett

Diese  $n + 1$  Fächer seien von links nach rechts mit den Nummern  $0, 1, 2, \dots, n$  durchnummeriert. Wird die Kugel an den  $n$  Pflöcken, die sie unterwegs trifft,  $k$  mal nach rechts und  $n - k$  mal nach links abgelenkt, so landet sie im Fach mit der Nummer  $k$  ( $k \in \{0, \dots, n\}$  geeignet).

Dies erfolgt mit der Wahrscheinlichkeit

$$p_k := \binom{n}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(1 - \frac{1}{2}\right)^{n-k} = \binom{n}{k} \cdot \left(\frac{1}{2}\right)^n.$$

Nach dem Bernoulli'schen (schwachen) Gesetz der großen Zahlen (Beispiel 6.3, p. 110) stimmt der Anteil der Kugeln im Fach mit der Nummer  $k$  an der Zahl

aller Kugeln ungefähr mit der Wahrscheinlichkeit  $p_k$  überein. Im Experiment zeigt dann die Verteilung der Kugeln über die  $n + 1$  Fächer die typische Gestalt der Gauß'schen Glockenkurve.

#### 6.2.4 Exponentialverteilung

Sei  $\lambda > 0$ . Eine Zufallsvariable  $X$  mit dem Wertebereich  $W(X) = (0, \infty)$  heißt *exponentialverteilt mit dem Parameter  $\lambda$* , wenn ihre Dichtefunktion gegeben ist durch

$$f(x) := f(x; \lambda) := \lambda e^{-\lambda x} \cdot \chi_{(0, \infty)}(x) = \begin{cases} \lambda e^{-\lambda x} & , \quad x \in (0, \infty) \\ 0 & , \quad \text{sonst} \end{cases} .$$

Die Exponentialverteilung ist als Modell für die Lebensdauer verschiedenster Objekte in Gebrauch (Beispiel 4.7, p. 78) und auch eng mit der Poisson-Verteilung verbunden. Die Poisson-Verteilung (Abschnitt 6.1.4.1) ist eine Verteilung der verschiedensten Zählvariablen, welche die Häufigkeit bestimmter Ereignisse während einer bestimmten Zeitperiode beschreiben. Es läßt sich zeigen:

Ist  $X$  eine Zählvariable mit einer Poisson-Verteilung, so hat die Variable  $T$ , welche die Zeitspanne zwischen zwei aufeinanderfolgenden Ereignissen mißt, die durch  $X$  gezählt werden, eine Exponentialverteilung.

Auch die Exponentialverteilung läßt sich durch die Eigenschaft der „Gedächtnislosigkeit“ charakterisieren:

#### Satz 6.7

Seien  $X$  eine Zufallsvariable und  $\lambda, x, y \in \mathbb{R}$ . Dann gilt

$$X \text{ ist } \lambda\text{-exponentialverteilt} \iff P(X > x + y \mid X > y) = P(X > x).$$

*Beweis*

Greiner/Tinhofer: Stochastik für Studienanfänger der Informatik 3.5.3, p. 78

#### Bemerkung

Die „Restverteilung“  $P(X \leq x + y \mid X > y)$  ist also mit der ursprünglichen Verteilung identisch:

$$P(X \leq x + y \mid X > y) = P(X \leq x).$$

Diese Eigenschaft gilt nur für die Exponentialverteilung, d.h. durch die Eigenschaft der „Gedächtnislosigkeit“ ist die Exponentialverteilung bereits eindeutig festgelegt. Umgangssprachliche Bemerkungen wie „Das Gerät ist ja wie neu“



deuten bereits an, daß exponentialverteilte Zufallsvariable besonders gut zur Modellierung der Lebensdauer von technisch ausgereiften Produkten (solchen ohne „Kinderkrankheiten“) bis zum ersten Auftreten von Alterserscheinungen geeignet sind.

### Erwartungswert und Varianz der Exponentialerteilung

$$E(X) = \int_0^{\infty} u \cdot \lambda e^{-\lambda u} du = \frac{1}{\lambda}$$

$$V(X) = \int_0^{\infty} u^2 \cdot \lambda e^{-\lambda u} du - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

## 6.2.5 Weitere stetige Verteilungen

### 6.2.5.1 Gamma-Verteilung

Seien  $\alpha, \beta > 0$ . Eine stetige Zufallsvariable  $X$  hat eine  $\Gamma$ -Verteilung mit den Parametern  $\alpha$  und  $\beta$ , wenn ihre Dichtefunktion gegeben ist durch

$$f(x) := f(x; \alpha, \beta) := \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} e^{-x/\beta} & , x > 0 \\ 0 & , \text{sonst} \end{cases} .$$

Dabei ist die *Euler'sche Gamma-Funktion* definiert durch

$$\Gamma(\alpha) := \int_0^{\infty} x^{\alpha-1} e^{-x} dx.$$

Für  $\alpha = 1$  erhält man die Exponentialverteilung.

### 6.2.5.2 Weibull-Verteilung

Seien  $\alpha, \beta > 0$ . Eine stetige Zufallsvariable  $X$  hat eine *Weibull-Verteilung*<sup>4</sup> mit den Parametern  $\alpha$  und  $\beta$ , wenn ihre Dichtefunktion gegeben ist durch

$$f(x) := f(x; \alpha, \beta) := \begin{cases} \alpha \beta x^{\beta-1} e^{-\alpha x^\beta} & , x > 0 \\ 0 & , \text{sonst} \end{cases} .$$

Auch diese Verteilung wird häufig zur Beschreibung von Ermüdungserscheinungen von Werkstoffen herangezogen. Die Lebensdauer  $T$  eines Gerätes oder einer Maschine besitzt dann eine Verteilungsfunktion vom Typ

$$F(x) = 1 - e^{-\alpha x^\beta} \quad (x > 0).$$

<sup>4</sup>Nach dem Schwedischen Physiker Waloddi Weibull, der diese Verteilung 1939 einführte

**Aufgabe**

Man zeige:  $F'(x) = f(x) \quad (x > 0)$ .

Für  $\beta = 1$  erhält man die Exponentialverteilung.

# Kapitel 7

## Mehrdimensionale Wahrscheinlichkeitsverteilungen

### 7.1 Stichprobenvariable und Testverteilungen

Wenn Werte einer Zufallsvariable  $X$  in einem Versuch ermittelt werden sollen, der aus mehreren identischen Einzelexperimenten besteht, so beschreibt man diesen Versuch durch eine Folge von identischen Zufallsvariablen  $X_1, \dots, X_n$ , die angeben, welchen Wert man bei dem ersten, zweiten,  $\dots$ ,  $n$ -ten Experiment gefunden hat. Man benutzt den Index damit nur zur Kennzeichnung desjenigen Einzelexperimentes, für welches die Zufallsvariable  $X_i$  „zuständig“ ist.

Die  $X_i$  heißen *Stichprobenvariable* für  $X$ . Jedes  $X_i$  ( $i \in \{1, \dots, n\}$ ) besitzt daher

- denselben Wertebereich:  $W(X_i) = W(X)$ ,
- dieselbe Dichtefunktion:  $f_{X_i} = f_X$ ,
- dieselbe Verteilung:  $F_{X_i} = F_X$

wie  $X$  selbst. Werden die Einzelexperimente des Versuches unabhängig voneinander durchgeführt, so spiegelt sich das in der Unabhängigkeit der Zufallsvariablen  $X_i$  wider, und die gemeinsame Verteilung hat dann die Dichtefunktion

$$\tilde{f}(x_1, \dots, x_n) := f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_n) = \prod_{i=1}^n f(x_i) \quad (7.1)$$

und die Verteilungsfunktion

$$\tilde{F}(x_1, \dots, x_n) := F(x_1) \cdot F(x_2) \cdot \dots \cdot F(x_n) = \prod_{i=1}^n F(x_i). \quad (7.2)$$

Im folgenden wird unter dem Begriff *Stichprobenvariable* einer Zufallsvariable  $X$  stets eine Folge  $X_1, \dots, X_n$  von unabhängigen und identisch wie  $X$  verteilten Zufallsvariablen verstanden, deren gemeinsame Verteilung durch (7.1) und (7.2) gegeben ist.

### 7.1.1 Chi-Quadrat-Verteilung

Die Chi-Quadrat-Verteilung ist eine der beiden sog. *Testverteilungen*, welche in der Statistik eine große Rolle spielt.

$X_1, \dots, X_n$  seien stochastisch unabhängige Zufallsvariable, welche alle der Standardnormalverteilung  $N(\mu = 0, \sigma = 1)$  genügen. Die Zufallsvariable

$$\chi_n^2 := X_1^2 + \dots + X_n^2 \quad (n \in \mathbb{N} \text{ geeignet})$$

ist eine stetige Zufallsvariable mit dem Wertebereich  $[0, \infty)$ . Sie heißt *Chi-Quadrat-Verteilung* und besitzt die Dichtefunktion (Abb. 7.1 bis 7.3)

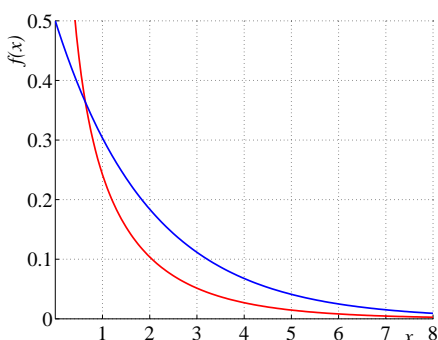


Abbildung 7.1: Chi-Quadrat-Verteilung für  $n = 1$  (rot) und  $n = 2$  (blau)

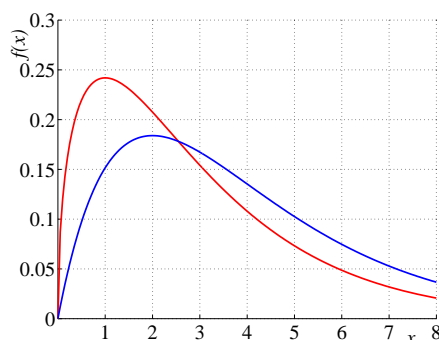


Abbildung 7.2: Chi-Quadrat-Verteilung für  $n = 3$  (rot) und  $n = 4$  (blau)

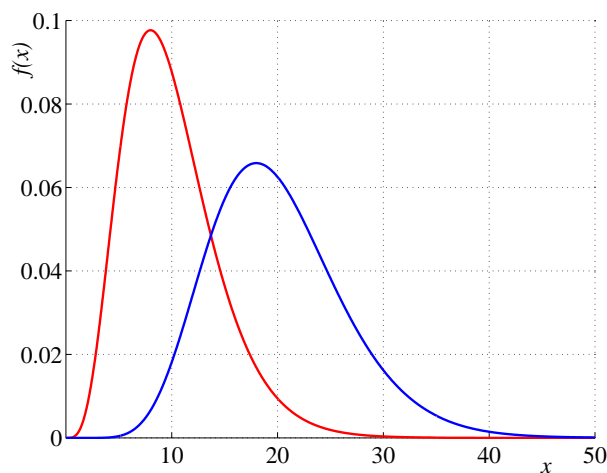


Abbildung 7.3: Chi-Quadrat-Verteilung für  $n = 10$  (rot) und  $n = 20$  (blau)

$$f_{\chi_n^2}(x) := \begin{cases} k_n \cdot x^{\frac{n-2}{2}} \cdot e^{-\frac{x}{2}} & , x > 0 \\ 0 & , x \leq 0 \end{cases} .$$

Die Konstante  $k_n$  ist eine Normierungskonstante, damit die Fläche unter dem Graphen der Dichtefunktion den Wert 1 bekommt. Sie berechnet sich daher aus

der Bedingung  $\int_{-\infty}^{\infty} f_{\chi_n^2}(x) dx = 1$ , und man erhält

$$k_n = \frac{1}{2^{\frac{n}{2}} \cdot \Gamma(\frac{n}{2})}$$

mit der *Euler'schen Gammafunktion*, definiert durch

$$\Gamma(\alpha) := \int_0^{\infty} e^{-x} \cdot x^{\alpha-1} dx \quad (\alpha > 0).$$

Die natürliche Zahl  $n$  heißt *Anzahl der Freiheitsgrade der Verteilung*. Sie bedeutet die Anzahl unabhängiger Informationen zur Berechnung von statistischen Aussagen. Für  $n \in \{1, 2\}$  fallen die Kurven streng monoton, für  $n > 2$  haben sie ein absolutes Maximum bei  $x = n - 2$ .

Die  $\chi_n^2$ -Verteilung besitzt die folgenden Kennwerte, welche nur von der Anzahl  $n$  der Freiheitsgrade abhängen:

$$\begin{array}{ll} \text{Erwartungswert} & : \quad E(\chi_n^2) = n \\ \text{Varianz} & : \quad V(\chi_n^2) = 2n \end{array}$$

Für große Freiheitsgrade  $n$  ( $n \geq 100$ ) läßt sich die  $\chi^2$ -Verteilung durch eine Normalverteilung mit dem Erwartungswert  $\mu = n$  und der Varianz  $\sigma^2 = 2n$  approximieren.

### 7.1.2 *t*-Verteilung von „Student“

Eine weitere wichtige Testverteilung ist die sog. *t-Verteilung von Student*<sup>1</sup>.

$X$  und  $Y$  seien zwei stochastisch unabhängige Zufallsvariable, für die gilt:

- $X$  ist standardnormalverteilt, also eine  $N(0, 1)$ -Zufallsvariable;
- $Y$  ist  $\chi^2$ -verteilt mit  $n$  Freiheitsgraden, d.h. es existieren  $n$   $N(0, 1)$ -Zufallsvariablen  $Y_1, \dots, Y_n$  mit  $Y := \sum_{i=1}^n Y_i^2$ .

Die aus diesen Zufallsvariablen gebildete Größe

$$T_n := \frac{X}{\sqrt{Y/n}}$$

<sup>1</sup>Diese Verteilung wurde 1908 von dem Mathematiker William Sealey Gosset entwickelt. Er hatte festgestellt, daß der standardisierte Erwartungswert normalverteilter Daten nicht mehr normalverteilt ist, wenn die Varianz des Merkmals unbekannt ist und mit Hilfe der Stichprobenvarianz geschätzt werden muß. Parametertests in der Statistik, bei denen *t*-Verteilungen Verwendung finden, bezeichnet man als *t-Tests*.

Gosset war zu dieser Zeit bei der irischen Brauerei Guinness angestellt, welche Forschungspublikationen ihrer Angestellten verboten hatte. Er publizierte seine Arbeit daher unter dem Pseudonym „Student“.

ist eine stetige Zufallsvariable mit der Dichtefunktion (Abb. 7.4)

$$f_{T_n}(x) := k_n \cdot \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}} \quad (x \in \mathbb{R}),$$

$$k_n := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \cdot \Gamma\left(\frac{n}{2}\right)}.$$

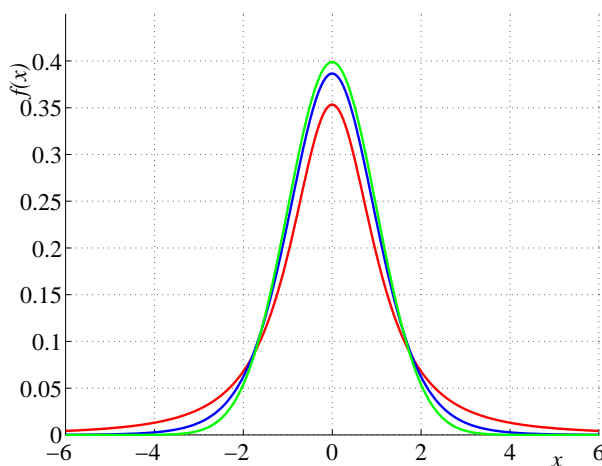


Abbildung 7.4:  $t$ -Verteilung für  $f = 2$  (rot),  $f = 8$  (blau) und Standardnormalverteilung (grün)

### Bemerkung

- (i) Die  $t$ -Verteilung hat ungefähr das Aussehen der Standardnormalverteilung, ist allerdings für kleine Werte von  $n$  breiter als diese.
- (ii) Für  $n = 1$  heißt die  $t$ -Verteilung auch *Cauchy-Verteilung*. Die Cauchy-Verteilung hat keinen Erwartungswert und keine Varianz.
- (iii) Für  $n > 1$  hat die  $T$ -Verteilung den Erwartungswert  $E(T_n) = 0$ .
- (iv) Für  $n \in \{1, 2\}$  hat die  $t$ -Verteilung keine Varianz. Für die anderen  $n$  gilt

$$V(T_n) = \frac{n}{n-2} \quad (n \in \mathbb{N}, n > 2).$$

- (v) Für  $n \geq 30$  darf die  $t$ -Verteilung in guter Näherung durch die Standardnormalverteilung ersetzt werden; es gilt

$$\lim_{n \rightarrow \infty} f_{T_n}(x) = \phi(x) \quad (x \in \mathbb{R});$$

hierbei ist  $\phi$  die Dichtefunktion der Standardnormalverteilung.

**Teil II**

**Statistik**

# Kapitel 8

## Inhalt der Statistik

### 8.1 Klassifikation der Statistik

Die Statistik läßt sich wie folgt unterteilen:

I Deskriptive Statistik

II Induktive Statistik

- (i) Festlegung von Zufallsbereichen
- (ii) Parameterschätzungen
- (iii) Hypothesentests

Zu I: Deskriptive Statistik

Zahlreiche Erkenntnisse entstammen der Analyse großer Datenmengen:

- Wahlen
- Meinungsumfragen
- Laboruntersuchung vieler Proben
- ...

In der Regel sammelt man Daten über ein *Merkmal*  $X$ , das an den Elementen einer umfangreichen *Beobachtungsmenge* (von Individuen oder Objekten) in unterschiedlicher *Ausprägung* zu erkennen ist. Anstelle von Beobachtungsmenge sagt man auch *Grundgesamtheit*. Ein erstes Ziel der Datenanalyse ist die Gewinnung einer möglichst aufschlußreichen numerischen oder graphischen Beschreibung der *Verteilung* dieser Ausprägung über die gesamte Beobachtungsmenge. Unter einer *statistischen Aussage* über  $X$  versteht man dann eine Aussage über diese Verteilung.



**Beispiel 8.1**

Der Fachbereich Elektrotechnik einer großen Hochschule startet eine Umfrage unter seinen ehemaligen Studenten, welche in den letzten fünf Jahren ihren Abschluß erreicht hatten.

- (i) Wieviele Monate dauerte es vom Erwerb Ihres Abschlusses bis zu Ihrer ersten Anstellung?
- (ii) Sind Sie
  - selbständig erwerbstätig,
  - an einem Hochschul- oder Forschungsinstitut,
  - in einem großen ( $n > 200$ ) Unternehmen,
  - in einem mittelgroßen ( $30 < n < 200$ ) Unternehmen,
  - in einer kleinen Firma ( $n < 30$ ),
  - noch ohne Anstellung?
- (iii) Wie schätzen Sie Ihr Gehalt im Vergleich zu Informatikern in Ihrem oder in einem vergleichbaren Unternehmen ein?
  - besser,
  - etwa gleich    oder
  - niedriger?

Die Umfrage hat zum Ziel, ein Berufsbild der Elektrotechniker zu zeichnen. Die gestellten Fragen beziehen sich dabei auf unterschiedliche Größen:

- (i) Quantitatives Merkmal  $X_1$
- (ii) Qualitatives Merkmal  $X_2$
- (iii) Rangmerkmal  $X_3$

Quantitative Merkmale werden weiterhin unterschieden in

- Diskrete Merkmale (Frage (i) des Beispiels)
- Stetige Merkmale (Zeit, Länge, Fläche, etc.)

Aufgabe der deskriptiven (= beschreibenden) Statistik ist es, Daten in eine übersichtliche Form zu bringen, sie auch graphisch aufzubereiten.

**Zu II: Induktive Statistik**

Hier soll eine Verbindung hergestellt werden zwischen der bis jetzt entwickelten Theorie (der Wahrscheinlichkeit) und der Wirklichkeit. Dabei stellt man die folgenden Fragen:

- (i) Wie lassen sich „Zufallsbereiche“ definieren? Diese erhält man aus einer bekannten Grundgesamtheit dadurch, daß man ihr Stichproben entnimmt und bestimmte Stichprobenkenngrößen bildet. Dann fragt man nach Bereichen, in denen diese Stichprobenkenngrößen mit einer festgelegten Wahrscheinlichkeit (der *statistischen Sicherheit*) auftreten werden (direkter Schluß).

- (ii) Welche Schlüsse lassen sich von einer *Stichprobe* auf die zugehörige Grundgesamtheit ziehen (indirekter Schluß)?
- (iii) Welche Zuverlässigkeit besitzen solche Schlüsse?

Zu (i) Beispiel für die Festlegung eines Zufallsbereiches:

Anhand einer erfolgten Wahl sollen für die nächste Wahl Stimmbezirke bestimmt werden, in denen das erzielte Ergebnis mit dem Gesamtergebnis in bezug auf möglichst viele Determinanten übereinstimmt, sog. „repräsentative“ Stimmbezirke.

Zu (ii) Beispiel für eine Parameterschätzung:

Ein Politiker möchte die Chancen seiner Wiederwahl an Hand einer Meinungsumfrage unter hundert zufällig ausgesuchten Wählern ermitteln (Schätzen des Erwartungswertes).

Zu (iii) Beispiel für einen Hypothesentest:

Ein 100-maliges Werfen einer Münze liefert die *statistische Hypothese*

$$P(\text{Kopf}) = P(\text{Wappen}) = \frac{1}{2}.$$

Wie sicher ist dieser „Schluß“?

### Zusammenfassung

Ziel der deskriptiven Statistik:

Vorhandene Daten sollen in eine übersichtliche Form gebracht und auf gemeinsame Merkmale hin untersucht und angeordnet werden.

beschreiben  $\longrightarrow$  deskriptive Statistik

Ziel der induktiven Statistik:

Aus der Beobachtung und Auswertung eines relativ kleinen Teiles (Stichprobe) einer großen Menge (Grundgesamtheit) von Daten soll auf Gesetzmäßigkeiten geschlossen werden, welche dann auch in der großen Menge gelten (sollen!).

schließen  $\longrightarrow$  induktive Statistik

## 8.2 Statistik und Wahrscheinlichkeitsrechnung

Die Statistik faßt Daten zu sinnvollen Größen zusammen und ermittelt Parameter, aus denen sich mit den Methoden der Wahrscheinlichkeitsrechnung Voraussagen für Ergebnisse zukünftiger Versuche ableiten lassen, vgl. die Übersicht in den Tabellen 8.1 und 8.2.

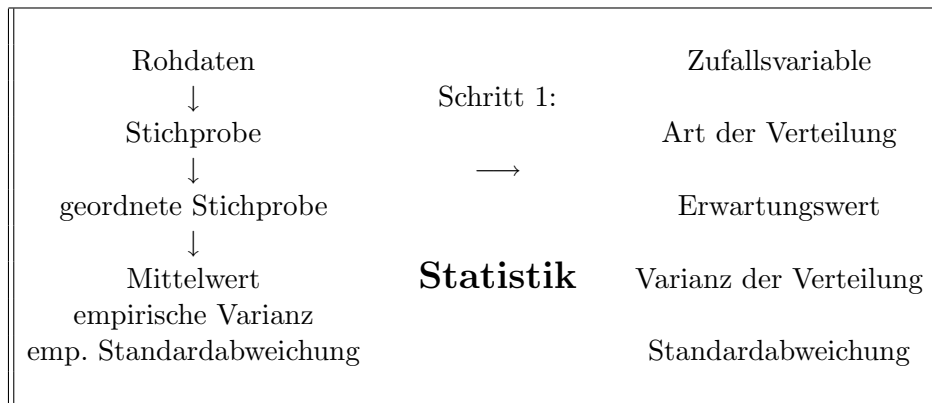


Tabelle 8.1: Von den Daten zur theoretischen Verteilung

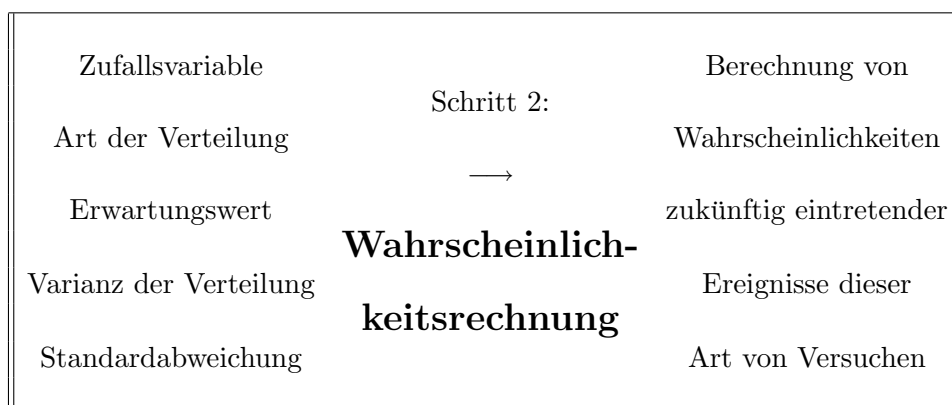


Tabelle 8.2: Von der Verteilung zu Wahrscheinlichkeiten

# Kapitel 9

## Deskriptive Statistik

### 9.1 Stichprobe und Grundgesamtheit

#### Definition 9.1

(i) *Grundgesamtheit*

Eine Menge  $\Omega$  von Elementen, welche hinsichtlich eines bestimmten Merkmals untersucht werden soll. Das interessierende Merkmal wird dabei durch eine Zufallsvariable  $X$  beschrieben.

(ii) *Stichprobe*

Darunter versteht man jede Teilmenge einer Grundgesamtheit. Eine nach dem Zufallsprinzip herausgegriffene Teilmenge von  $n$  Elementen heißt eine *Zufallsstichprobe vom Umfang  $n$* .

Die beobachteten Merkmalswerte  $x_1, \dots, x_n$  der  $n$  Elemente einer Stichprobe sind Realisierungen einer Zufallsvariable  $X$  und heißen *Stichprobenwerte*. Man beachte, daß diese Stichprobenwerte  $x_i$  oder, wie man auch sagt, die *Merkmalsausprägungen*, nicht alle verschieden sein müssen.

(iii) *verzerrt, unverzerrt (biased sample)*

Eine Stichprobe heißt *verzerrt*, wenn ihr Zustandekommen vom untersuchten Merkmal selbst abhängt.

#### Beispiel

Die Meinung der Bevölkerung in bezug auf die Politik der SPD wird anhand einer Zufallsstichprobe ermittelt, welche aus den Mitgliedern einer Parteiversammlung der CDU gebildet wird. Hier ist die Grundgesamtheit schon vorsortiert.

Haben alle Elemente der Grundgesamtheit die gleiche Chance, ausgewählt zu werden, so heißt die Stichprobe *unverzerrt*.

(iv) *repräsentative Stichprobe*

Der Begriff „repräsentative Stichprobe“ nimmt den Zufall schon etwas an die Leine; hier wird das Terrain der Mathematik ein wenig verlassen, da die Stichprobe absichtlich im Hinblick auf das Zustandekommen ihres Ergebnisses ausgewählt wird. Mit repräsentativen Stichproben wird insbesondere in der empirischen Sozialforschung gearbeitet.

**Beispiel**

Vor der Durchführung der Befragung einer Stichprobe von Personen in bezug auf ihre Meinung zur Arbeitsmarktpolitik der Regierung wird eine Vorauswahl getroffen anhand von

- Alter
- Wohnsitz
- Beruf
- ???

- (v) Für  $x \in \mathbb{R}$  sei  $g(x)$  die Anzahl derjenigen Werte in einer Stichprobe vom Umfang  $n$ , welche gleich  $x$  sind. Es gilt also

$$g(x) \neq 0 \quad \iff \quad x \in \{x_1, \dots, x_n\}$$

$g(x)$  heißt *absolute Häufigkeit* von  $x$ , und  $h(x) := \frac{1}{n} \cdot g(x)$  heißt *relative Häufigkeit* von  $x$ . Die dadurch definierte Funktion  $h : \mathbb{R} \rightarrow \mathbb{R}$  heißt *Häufigkeitsfunktion* oder *empirische Dichte* des Merkmals  $X$  bei der Stichprobe  $\{x_1, \dots, x_n\}$ .

- (vi) Für  $x \in \mathbb{R}$  sei  $G(x)$  die Anzahl derjenigen Werte in einer Stichprobe vom Umfang  $n$ , welche kleiner oder gleich  $x$  sind.  $G(x)$  heißt *absolute Summenhäufigkeit* von  $x$ , und die Größe  $H(x) := \frac{1}{n} \cdot G(x)$  heißt *relative Summenhäufigkeit* an der Stelle  $x$ . Die dadurch definierte Funktion  $H : \mathbb{R} \rightarrow \mathbb{R}$  heißt *empirische Verteilungsfunktion* des Merkmals  $X$  bei der Stichprobe  $\{x_1, \dots, x_n\}$ . Es handelt sich dabei um eine monoton steigende Treppenfunktion, jeder Stichprobenwert  $x_i$  ist eine Sprungstelle von  $H$ .

(vii) **Beispiel**

Gegeben sei die Grundgesamtheit  $\Omega := \{1, 2, \dots, 10\}$  und die Stichprobe  $A := \{x_1, \dots, x_7\}$  mit  $x_1 = x_3 = x_5 = x_7 = 1$ ,  $x_2 = 2$ ,  $x_4 = 4$  und  $x_6 = 6$ . Dann gelten:

$$g(x) := \begin{cases} 4 & , \quad x = 1, \\ 1 & , \quad x = 2, \\ 1 & , \quad x = 4, \\ 1 & , \quad x = 6, \\ 0 & , \quad \text{sonst} \end{cases} \quad h(x) := \begin{cases} \frac{4}{7} & , \quad x = 1, \\ \frac{1}{7} & , \quad x = 2, \\ \frac{1}{7} & , \quad x = 4, \\ \frac{1}{7} & , \quad x = 6, \\ 0 & , \quad \text{sonst} \end{cases}$$

$$g(x) := \begin{cases} 0 & , \quad x < 1, \\ 4 & , \quad 1 \leq x < 2, \\ 5 & , \quad 2 \leq x < 4, \\ 6 & , \quad 4 \leq x < 6, \\ 7 & , \quad 6 \leq x \end{cases} \quad G(x) := \begin{cases} 0 & , \quad x < 1, \\ \frac{4}{7} & , \quad 1 \leq x < 2, \\ \frac{5}{7} & , \quad 2 \leq x < 4, \\ \frac{6}{7} & , \quad 4 \leq x < 6, \\ 1 & , \quad 6 \leq x \end{cases}$$

Tabelle 9.1 liefert eine Übersicht über die Zusammenhänge zwischen Größen der Wahrscheinlichkeitstheorie und solchen der empirischen Statistik.

### Gruppierung der Stichprobenwerte in Klassen

Bei umfangreichen Stichproben mit vielen verschiedenen Werten gruppiert man diese zweckmäßigerweise in sog. *Klassen*.

- (i) Die Stichprobe wird geordnet, und es werden ihr kleinster und ihr größter Wert  $x_{\min}$  und  $x_{\max}$  bestimmt. Dadurch wird ein Intervall  $\mathcal{I}$  definiert, in dem alle Stichprobenwerte liegen.
- (ii) Das Intervall  $\mathcal{I}$  wird in  $k$  Teilintervalle  $\Delta\mathcal{I}$  der gleichen Breite  $\Delta x$  zerlegt (sog. *Klassen gleicher Breite*).
- (iii) Die Mitte  $\tilde{x}_i$  eines jeden Klassenintervalles  $\Delta\mathcal{I}$  heißt *Klassenmitte*. Die Einteilung sollte dabei so gewählt werden, daß die Klassenmitten durch möglichst einfache Zahlen (z. B. ganze Zahlen) charakterisiert werden.
- (iv) Fällt ein Stichprobenwert in einen der beiden Randpunkte einer Klasse, so zählt man ihn je zur Hälfte den beiden angrenzenden Klassen zu.
- (v) Wenn man über den Klassen Rechtecke gleicher Breite  $\Delta x = \Delta x_i$  errichtet, deren Höhen den relativen Klassenhäufigkeiten entsprechen, so entsteht ein sog. *Histogramm*, s. Abb. 9.1. Die Flächeninhalte der Rechtecke sind dann den relativen Klassenhäufigkeiten proportional.

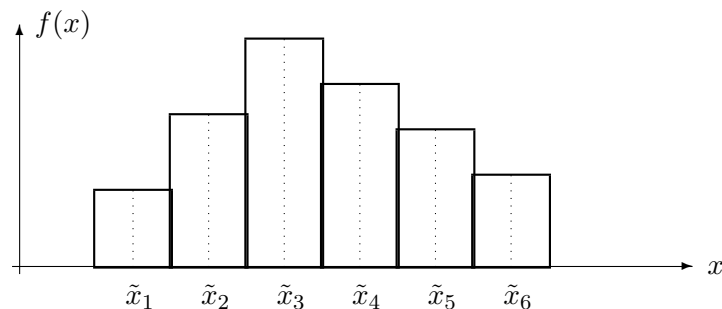


Abbildung 9.1: Histogramm einer gruppierten Stichprobe

Wahrscheinlichkeitstheorie	empirische Statistik
Ergebnismenge $\Omega$	Grundgesamtheit $\Omega$
Ereignis $A \subseteq \Omega$	Stichprobe $A \subseteq \Omega$
Zufallsvariable $X$	Merkmal $X$
$W(X)$ : Realisierungen von $X$	Stichprobenwerte $x_1, \dots, x_n$
Dichtefunktion $f_X$	relative Häufigkeit $h$
Verteilungsfunktion $F_X$	empirische Verteilungsfunktion $H$
$F_X(x) = \sum_{\{i: x_i \leq x\}} f_X(x_i)$	$H(x) = \sum_{\{i: x_i \leq x\}} h(x_i)$
Erwartungswert $\mu = E(X)$	(empirischer) Mittelwert $\bar{x}$
$E(X) = \sum_i x_i f_X(x_i)$	$\bar{x} = \frac{1}{n} \sum_{i=1}^n h(x_i)$
Varianz $\sigma^2 = V(X)$	(empirische) Varianz
$V(X) = \sum_i (x_i - \mu)^2 f_X(x_i)$	$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
	oder
	$\hat{s}_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$

Tabelle 9.1: Übersicht über wahrscheinlichkeitstheoretische - und statistische Größen

## 9.2 Statistische Parameter

Jede Stichprobe läßt sich durch ihre Häufigkeitsverteilung vollständig beschreiben, also entweder durch ihre

- empirische Dichtefunktion  $h$       oder
- empirische Verteilungsfunktion  $H$ .

Man kann eine Stichprobe aber auch durch bestimmte *statistische Kennwerte* (auch *Maßzahlen* genannt) charakterisieren, wenn auch, wie schon bei manchen Zufallsvariablen, in unvollständiger Weise. In diesem Zusammenhang spricht man von *empirischen Kennwerten*, im Unterschied zu den Kennwerten einer Wahrscheinlichkeitsverteilung. Diese Kennwerte sollen darüber Auskunft geben, wo die „Mitte“ der beobachteten Werte ist und wie groß der Bereich ist, über den sich die Werte „im wesentlichen“ erstrecken.

### 9.2.1 Mittelwert

#### Definition 9.2

Seien  $n \in \mathbb{N}$  und  $A := \{x_1, \dots, x_n\}$  eine Stichprobe, also  $A$  eine Stichprobe vom Umfang  $n$ .

- (i) Die Größe

$$\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$$

heißt *arithmetischer Mittelwert* der Stichprobe, auch *Stichprobenmittel*.

- (ii) Ordnet man die Werte der Stichprobe der Größe nach, so heißt die entstehende Zahlenreihe

$$x_{(1)}, \dots, x_{(n)}$$

mit  $x_{(1)} \leq \dots \leq x_{(n)}$  die zugehörige *geordnete Stichprobe*.

- (iii) Die Zahl

$$\tilde{x} := \begin{cases} x_{(\frac{n+1}{2})} & , \quad n \text{ ungerade} \\ \frac{1}{2} \cdot [x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}] & , \quad n \text{ gerade} \end{cases}$$

heißt *Median* der Stichprobe.

- (iv) Falls ein Wert in einer Stichprobe am häufigsten auftritt, so heißt dieser Wert *Modalwert*.

#### Beispiel 9.1

Kleinwagen haben laut Liste einen durchschnittlichen Benzinverbrauch von 5,5 l/100 km, Mittelklassewagen einen von 7,5 l/100 km und Geländewagen einen



von 10,5 l/100 km. Ein Mietwagenhändler besitze eine Flotte von 100 Kraftfahrzeugen, davon 50 Kleinwagen, 30 Mittelklassewagen und 20 Geländewagen. Je nach Wahl des Durchschnittswertes hat diese Flotte einen unterschiedlichen Benzinverbrauch:

(i) arithmetischer Mittelwert:	Verbrauch auf 100 km:	7,1 l
(ii) Median:	Verbrauch auf 100 km:	6,5 l
(iii) Modalwert:	Verbrauch auf 100 km:	5,5 l

Je nachdem, ob man in einer Argumentation an einem hohen oder niedrigen Benzinverbrauch interessiert ist, zieht man dann den einen oder anderen Mittelwert zu Rate.

### Bemerkung

- (i) Stichprobenmittelwert, Median oder Modalwert geben Auskunft über die Lage der beobachteten Werte auf der Zahlengerade. Sie heißen daher auch *Lageparameter*. Der Median „teilt“ eine geordnete Stichprobe in zwei gleich umfangreiche Teile, so daß sich links von ihm genauso viele Stichprobenwerte befinden wie rechts von ihm.
- (ii) Jeder der drei Parameter hat eine gewisse Optimalitätseigenschaft. Sie ergeben sich aus der Forderung, die Stichprobe „optimal“ durch eine einzige Zahl zu repräsentieren. Ein solcher Repräsentant sollte also „möglichst nahe“ an allen  $x_i$  gleichzeitig liegen.
- (iii) Der Median  $\tilde{x}$  minimiert die Summe der betragsmäßigen Abweichungen von den Werten  $x_i$  der Stichprobe:

$$\tilde{x} = \min_{x \in \mathbb{R}} \left\{ \sum_{i=1}^n |x_i - x| \right\}.$$

- (iv) Das Stichprobenmittel  $\bar{x}$  minimiert die Summe der quadratischen Abweichungen von den Werten  $x_i$  der Stichprobe:

$$\bar{x} = \min_{x \in \mathbb{R}} \left\{ \sum_{i=1}^n (x_i - x)^2 \right\}.$$

### Aufgabe

Man beweise Teil (iv) der voranstehenden Bemerkung durch eine Extremwertberechnung.

### 9.2.2 Empirische Varianz und Standardabweichung

Empirische Verteilungen mit demselben Mittelwert können völlig verschieden aussehen. Daher reicht die Angabe des Mittelwertes  $\bar{x}$  zur Charakterisierung der Häufigkeitsverteilung allein nicht aus. Da die einzelnen Stichprobenwerte  $x_i$  um ihren Mittelwert  $\bar{x}$  – ihren Schwerpunkt – *streuen*, benötigt man noch ein geeignetes *Streuungsmaß*, welches die Größe „aller“ dieser Abweichungen mißt.

#### Definition 9.3

Seien  $n \in \mathbb{N}$  und  $A := \{x_1, \dots, x_n\}$  eine Stichprobe, also  $A$  eine Stichprobe vom Umfang  $n$ .

- (i) Die *Spannweite* der Stichprobe ist die Differenz zwischen größtem und kleinstem Wert:

$$v := \max_{1 \leq i \leq n} \{x_i\} - \min_{1 \leq i \leq n} \{x_i\} = x_{(n)} - x_{(1)}$$

- (ii) Die Zahl

$$s_x^2 := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

heißt *Varianz* der Stichprobe, auch *empirische Varianz*.

- (iii) Die Zahl

$$s_x := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

heißt *Stichprobenstreuung* oder *Standardabweichung* der Stichprobe.

#### Bemerkung

- (i) Die eben genannten Parameter geben Auskunft darüber, wie sehr die Stichprobenwerte um ihr Zentrum „streuen“, sie heißen daher auch *Streuungsparameter*.
- (ii) Die Stichprobenstreuung  $s_x$  hat gegenüber der Stichprobenvarianz  $s_x^2$  den Vorteil, daß sie dieselbe Einheit wie die beobachtete Größe selbst hat.
- (iii) Die Stichprobenvarianz  $s_x^2$  ist bei festem Stichprobenmittel  $\bar{x}$  eine Funktion von  $n-1$  Variablen: z. B. läßt sich  $x_n$  aus der Gleichung

$$\sum_{i=1}^n x_i = n \cdot \bar{x}$$

ausrechnen und in den Ausdruck für  $s_x^2$  einsetzen. Dies ist *eine* (die schwächere) Erklärung für den Faktor  $\frac{1}{n-1}$  anstelle von  $\frac{1}{n}$ , die andere ist die Erwartungstreue der Varianz.

(iv) Alle Werte der Stichprobe liegen im Intervall

$$\mathcal{I} := [\bar{x} - \sqrt{n-1} \cdot s_x, \bar{x} + \sqrt{n-1} \cdot s_x],$$

denn für jedes  $k \in \{1, \dots, n\}$  folgt

$$(x_k - \bar{x})^2 \leq \sum_{i=1}^n (x_i - \bar{x})^2 = (n-1) \cdot s_x^2$$

$$\longrightarrow \bar{x} - \sqrt{n-1} \cdot s_x \leq x_k \leq \bar{x} + \sqrt{n-1} \cdot s_x.$$

Dies zeigt deutlich, wie die Stichprobenstreuung die „Ausdehnung“ der Stichprobenwerte  $x_1, \dots, x_n$  auf der Zahlengerade beschreibt.

Dieses mit Hilfe von  $s_x$  angegebene Intervall ist jedoch i.a. viel größer als die Stichprobenspannweite  $v$  angibt. Das kommt daher, daß obige Intervallabschätzung für alle Stichproben mit dem Mittelwert  $\bar{x}$  und der Standardabweichung  $s_x$  gilt und daher sehr großzügig ausfällt, während die Spannweite selbst an Hand einer speziellen Stichprobe ermittelt wird.

(v) Zur praktischen Berechnung der (empirischen) Varianz ist, wie bei Zufallsvariablen auch, der Satz von Steiner oft besser geeignet:

$$s_x^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right).$$

Stichprobenmittelwert und Stichprobenvarianz sind Näherungswerte für die ihnen zugeordneten theoretischen Analogie Erwartungswert und Varianz einer Wahrscheinlichkeitsverteilung. Es ist eine zentrale Aufgabe der induktiven Statistik zu berechnen, *wie gut* diese Näherungswerte (Schätzwerte) quantitativ sind.

# Kapitel 10

## Induktive Statistik

### 10.1 Statistische Schätzverfahren

#### 10.1.1 Problemstellung

Betrachtet wird eine Grundgesamtheit  $\Omega$  hinsichtlich eines bestimmten Merkmals  $X$ . Diese Grundgesamtheit ist durch die Dichtefunktion  $f_X$  oder die Verteilungsfunktion  $F_X$  der entsprechenden Zufallsvariable  $X$  vollständig charakterisiert.

Nun sei die Verteilungsfunktion von der Art her zwar bekannt, enthalte jedoch noch unbekannte Parameter. Beispielsweise liege eine Normalverteilung vor, deren bestimmende Parameter  $\mu$  und  $\sigma$  noch unbekannt seien.

Dann ergeben sich die folgenden Fragen:

1. Wie erhält man auf der Basis einer konkreten Stichprobe Schätz- oder Näherungswerte für diese unbekannt Parameter?
2. Wie genau und wie sicher sind solche Schätzwerte?

Die Beantwortung der ersten Frage führt auf sog. *Schätzfunktionen*, mit deren Hilfe sich *Schätzwerte* für die unbekannt Parameter einer Verteilung ermitteln lassen. Daher bezeichnet man diese Art der Schätzung auch als eine *Punktschätzung*.

Die Beantwortung der zweiten Frage führt zu sog. *Konfidenz-* oder *Vertrauensintervallen*. Diese sind durch Stichprobenuntersuchungen bestimmte Intervalle, die einen oder mehrere unbekannt Parameter der Verteilung mit einer vorgegebenen Wahrscheinlichkeit überdecken. Man spricht daher hier von einer *Intervall-* oder *Bereichsschätzung*.

### 10.1.2 Schätz- und Stichprobenfunktionen

#### Beispiel 10.1

Beim radioaktiven Zerfall eines bestimmten chemischen Elementes zerfallen die einzelnen Atomkerne ungeordnet und unabhängig voneinander, eine definitive Aussage darüber, wann ein bestimmter Atomkern zerfällt, ist nicht möglich.

Die diskrete Zufallsvariable

$X$   $\iff$  Anzahl der Atomkerne, die in einem bestimmten Zeitintervall zerfallen

genügt einer Poisson-Verteilung mit der Dichtefunktion

$$f(x) = P(X = x) = \frac{x^\mu}{x!} \cdot e^{-\mu} \quad (x \in \mathbb{N}).$$

Der dabei unbekannt Parameter  $\mu$  in dieser Verteilung ist der Erwartungswert von  $X$  und gibt an, wieviele Atomkerne „im Mittel“ (Was versteht man darunter?) in dem gewählten Zeitintervall zerfallen. Es liegt dann nahe, den arithmetischen Mittelwert  $\bar{x}$  einer konkreten Stichprobe als einen Schätzwert für  $\mu$  heranzuziehen:

$$\mu \approx \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

#### Problem

Warum den arithmetischen Mittelwert und nicht einen anderen, beispielsweise den Median oder den Modalwert? Weiterhin muß ein Kriterium definiert werden, wann ein Verfahren zur Gewinnung eines Schätzwertes für einen Parameter als „optimal“ angesehen werden kann.

#### Situation

$X$  sei eine (diskrete oder kontinuierliche) Zufallsvariable, deren Art der Verteilung (Normalverteilung, Poisson-Verteilung, etc.) bekannt sei, etwa dadurch, daß man die zugehörige Dichtefunktion  $f_X(\cdot, \theta)$  kennt, welche von einem ein- oder mehrdimensionalen Parameter  $\theta$  abhängt.

#### Ziel

Durch Analyse einer Stichprobe für  $X$  will man Aufschluß über den konkreten Wert von  $\theta$  erhalten.

Das aktuelle „Experiment“, welches auf die Zufallsvariable  $X$  führt, werde  $n$ -mal unabhängig voneinander wiederholt; beispielsweise das  $n$ -malige Werfen einer Münze. Die Stichprobe für  $X$  bestehe somit aus einer Meßreihe  $x_1, \dots, x_n$  von Daten.

Jeder der Werte  $x_i$  läßt sich auffassen als Beobachtung (Realisierung) einer Zufallsvariable  $X_i$  ( $i \in \{1, \dots, n\}$ ), welche den Ausgang der  $i$ -ten Beobachtung beschreibt. Somit ist jedes  $X_i$  eine Art „Kopie“ von  $X$  und besitzt dieselbe Verteilung. Die  $X_i$  heißen *Stichprobenvariable* für  $X$  und bilden ein  $n$ -Tupel von unabhängigen Zufallsvariablen, also einen  $n$ -dimensionalen *Zufallsvektor*

$$\vec{X} := \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}.$$

Diejenigen  $n$ -Tupel, die man beobachten kann, also beispielsweise das Tupel  $\langle x_1, \dots, x_n \rangle$ , bilden die Elementarereignisse desjenigen Versuchs, der in der Entnahme der aktuellen Stichprobe besteht. Ein Schätzwert  $\hat{\theta}$  für den unbekannt Parameter  $\theta$  wird somit dadurch gewonnen, daß man die jeweiligen Realisierungen  $x_i$  der Zufallsvariable  $X_i$ , welche sich aufgrund einer konkreten Stichprobe ergeben, als Variable in eine geeignete Funktion einsetzt.

### Definition 10.1

Seien  $n \in \mathbb{N}$  (Stichprobenumfang) und  $\vec{X} := (X_1 \dots X_n)^T$  ein  $n$ -dimensionaler Zufallsvektor, dessen Komponenten aus Zufallsvariablen mit derselben Verteilung bestehen.

- (i) Ein beliebiger Term  $T(X_1, \dots, X_n)$  der Variablen  $X_1, \dots, X_n$  heißt eine *Statistik*.
- (ii) Betrachtet man die  $X_i$  als reelle Variable in einer geeigneten Menge  $\mathcal{D} \subseteq \mathbb{R}^n$ , so liefert der Term  $T(X_1, \dots, X_n)$  eine Funktion

$$T : \mathbb{R}^n \supseteq \mathcal{D} \longrightarrow \mathbb{R}^d \quad (d \geq 1 \text{ geeignet}),$$

welche als Funktion von Zufallsvariablen selbst eine Zufallsvariable ist und *Stichprobenfunktion* heißt.

- (iii) Sei  $\mathcal{W} \subseteq \mathbb{R}$  der gemeinsame Wertebereich aller „identischen“ Zufallsvariablen  $X_i$  und sei  $\mathcal{D} := \mathcal{W}^n$ . Ein *Schätzwert*  $\hat{\theta}$  für den unbekannt Parameter  $\theta$  ist die Auswertung der Stichprobenfunktion  $T$  auf den Werten einer konkreten Stichprobe:

$$\begin{array}{lll} T : & \mathcal{D} := \mathcal{W}^n & \longrightarrow \quad \Theta \subseteq \mathbb{R}^d \\ & \langle x_1, \dots, x_n \rangle & \longmapsto \quad T(x_1, \dots, x_n) =: \hat{\theta}. \end{array}$$

- (iv) Eine Stichprobenfunktion zur Ermittlung des Schätzwertes  $\hat{\theta}$  für einen Parameter  $\theta$  heißt *Schätzfunktion*, *Schätzvariable* oder *Schätzer*.

Im folgenden sollen zwei Schätzfunktionen untersucht werden, mit Hilfe derer sich die unbekannt Parameter  $\mu = E(X)$  (Erwartungswert) und  $\sigma^2 = V(X)$

(Varianz) einer Verteilung mit Hilfe einer Stichprobe schätzen lassen. Insbesondere für eine Normalverteilung sind dieses die beiden wichtigsten Schätzfunktionen.

### Beispiel 10.2

Sei  $X$  eine Zufallsvariable mit dem unbekanntem Erwartungswert  $\mu := E(X)$  und der unbekanntem Varianz  $\sigma^2 := V(X)$ .

- (i) Schätzwert für den Erwartungswert  $\mu = E(X)$ :

$$\text{Stichprobenmittelwert} \quad \bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$$

- (ii) Schätzwert für die Varianz  $\sigma^2 = V(X)$ :

$$\text{Stichprobenvarianz} \quad \hat{s}^2 := \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

### Definition 10.2

- (i) Jeder Wert  $x_i$  ist ein beobachteter Wert der Zufallsvariable  $X_i$ , daher ist  $\bar{x}$  ein beobachteter Wert der aus den  $X_i$  abgeleiteten Zufallsvariable

$$\bar{X} := T_1(X_1, \dots, X_n) := \frac{1}{n} \sum_{i=1}^n X_i,$$

welche ebenfalls als *Stichprobenmittel* bezeichnet wird.

- (ii) Die Stichprobenvarianz

$$\hat{s}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

ist ein beobachteter Wert der Zufallsvariable

$$\hat{S}^2 := T_2(X_1, \dots, X_n) := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2,$$

welche ebenfalls als *Stichprobenvarianz* bezeichnet wird.

### Problem

Diese beiden Schätzfunktionen sind anhand von Schätzwerten, welche man schließlich aus ihnen erzeugen will, konstruiert worden. Es ist noch nicht bekannt, wie gut die Werte prinzipiell sind, welche sie liefern, wenn jeweils eine

konkrete Stichprobe als Argument eingesetzt wird. Gibt es beispielsweise systematische Abweichungen zu den wahren, hier aber noch unbekanntem Parameter  $\mu = E(X)$  und  $\sigma^2 = V(X)$ ?

**Definition 10.3**

Sei  $T$  eine Schätzfunktion (Schätzer) für den unbekanntem Parameter  $\theta$  einer Zufallsvariable  $X$ .

- (i)  $T$  heißt *erwartungstreu*  $:\iff E(T) = \theta$ .

Eine erwartungstreue Schätzung enthält keinen systematischen Fehler, d.h. ihr Erwartungswert – der Schwerpunkt der Verteilung – stimmt mit dem zu schätzenden Parameter überein.

- (ii)  $T$  heißt *asymptotisch erwartungstreu*  $:\iff \lim_{n \rightarrow \infty} E(T) = \theta$ ;  
hierbei bedeutet  $n$  den Stichprobenumfang.

- (iii)  $T$  heißt *konsistent*, wenn  $T$  für  $n \rightarrow \infty$  in Wahrscheinlichkeit gegen  $\theta$  konvergiert oder die Folge  $\{T(X)\}_{n \in \mathbb{N}}$  dem schwachen Gesetz der großen Zahlen unterliegt, genauer:

$$\bigwedge_{\varepsilon > 0} P(\{\omega : |T(X)(\omega) - \theta| > \varepsilon\}) \xrightarrow{(n \rightarrow \infty)} 0;$$

hierbei bedeutet  $n$  wieder den Stichprobenumfang. Kurzschreibweise:

$$\bigwedge_{\varepsilon > 0} P(|T - \theta| > \varepsilon) \xrightarrow{(n \rightarrow \infty)} 0$$

- (iv)  $T$  heißt *effizient*, wenn gilt:

- (a)  $T$  ist erwartungstreu.  
(b) Für  $\theta$  gibt es keine andere erwartungstreue Schätzfunktion mit kleinerer Varianz, d.h. jede andere erwartungstreue Schätzfunktion  $\tilde{T}$  erfüllt die Bedingung  $V(T) \leq V(\tilde{T})$ .

Die Schätzung eines unbekanntem Parameters  $\theta$  erfolgt also mit einer geeigneten Stichprobenfunktion, die in diesem Zusammenhang als *Schätzfunktion für den Parameter  $\theta$*  fungiert. Sie wird als „optimal“ betrachtet, wenn sie die oben genannten Kriterien erfüllt.

**Satz 10.1**

Sei  $X$  eine Zufallsvariable mit dem Erwartungswert  $\mu = E(X)$  und der Varianz  $\sigma^2 = V(X)$ .

- (i) Die Schätzfunktion

$$\bar{X} = T_1(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$$

für den Erwartungswert ist erwartungstreu, konsistent und effizient.



(ii) Die Schätzfunktion

$$\hat{S}^2 = T_2(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

für die Varianz ist *nicht* erwartungstreu.

(iii) Die Schätzfunktion

$$S^2 := T_3(X_1, \dots, X_n) := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

für die Varianz ist erwartungstreu.

*Beweis*

(i) (a) *Erwartungstreue*

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \stackrel{(E \text{ linear})}{=} \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \cdot n \cdot E(X) = E(X)$$

(b) *Konsistenz*

Die Behauptung ist gerade das schwache Gesetz der großen Zahlen für identisch verteilte Zufallsvariablen, welches in Kapitel 6 für den Spezialfall von binomialverteilten Zufallsvariablen bewiesen worden ist (Beispiel 6.3):

$$\bigwedge_{\varepsilon > 0} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - E(X)\right| > \varepsilon\right) \longrightarrow 0 \quad (n \rightarrow \infty),$$

d.h. Konvergenz in Wahrscheinlichkeit von

$$\frac{1}{n} \sum_{i=1}^n X_i \longrightarrow E(X) \quad (n \rightarrow \infty).$$

(c) *Effizienz*

Cramér-Rao Ungleichung

(ii) (1) Mit  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  folgt

$$\begin{aligned} V(\bar{X}) &= V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) \\ &= \frac{1}{n^2} \cdot n \cdot V(X) = \frac{1}{n} V(X), \end{aligned}$$

da sich die Varianzen voneinander unabhängiger Zufallsvariablen addieren<sup>1</sup> und sich bei der Varianz ein konstanter Faktor als Quadrat herausziehen läßt.

<sup>1</sup>Für beliebige Zufallsvariablen  $X$  und  $Y$  gilt

$$V(X + Y) = V(X) + V(Y) + 2 \cdot Cov(X, Y),$$

(2) Sei  $\mu = E(X)$ , dann ist mit  $\sum_{i=1}^n X_i = n\bar{X}$ :

$$\begin{aligned} (X_i - \bar{X})^2 &= [(X_i - \mu) - (\bar{X} - \mu)]^2 \\ &= (X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2, \\ \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i - \mu)^2 - 2 \sum_{i=1}^n (X_i - \mu)(\bar{X} - \mu) + \sum_{i=1}^n (\bar{X} - \mu)^2, \\ &= \sum_{i=1}^n (X_i - \mu)^2 - 2n(\bar{X} - \mu)^2 + n(\bar{X} - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2. \end{aligned}$$

(3) Wegen  $V(X_i) = E((X_i - \mu)^2) = E((X - \mu)^2) = V(X)$  ( $i \in \{1, \dots, n\}$ )

---

mit der sog. *Kovarianz* von  $X$  und  $Y$ . Sie ist definiert durch

$$Cov(X, Y) := E [X - E(X)] \cdot [Y - E(Y)] = E(X \cdot Y) - E(X) \cdot E(Y).$$

Zum Beweis der letzten Gleichung multipliziere man aus und wende die Regeln für das Rechnen mit Erwartungswerten an.

Aus der letzten Beziehung erkennt man, daß die Kovarianz für unabhängige Zufallsvariablen verschwindet, denn für die Erwartungswerte unabhängiger Zufallsvariablen gilt die Beziehung

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

Damit wird die Additionsregel für die Varianz zu

$$V(X + Y) = V(X) + V(Y).$$

Die Kovarianz zweier Zufallsvariablen kann als ein Maß für die „Stärke ihrer Abhängigkeit“ bezeichnet werden. Da der Wert von  $Cov(X, Y)$  jedoch vom verwendeten Maßstab abhängt, verwendet man als solches Maß besser den normierten Ausdruck

$$\rho(X, Y) := \frac{Cov(X, Y)}{\sqrt{V(X)} \cdot \sqrt{V(Y)}},$$

den man als *Korrelationskoeffizienten* von  $X$  und  $Y$  bezeichnet.

folgt daraus

$$\begin{aligned}
 E(\hat{S}^2) &= E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) \\
 &= \frac{1}{n} E\left(\sum_{i=1}^n (X_i - \bar{X})^2\right) \\
 &\stackrel{(2)}{=} \frac{1}{n} \left[ E\left(\sum_{i=1}^n (X_i - \mu)^2\right) - n \cdot E((\bar{X} - \mu)^2) \right] \\
 &= \frac{1}{n} \sum_{i=1}^n \underbrace{E((X_i - \mu)^2)}_{V(X_i)} - \underbrace{E((\bar{X} - \mu)^2)}_{V(\bar{X})} \\
 [V(X_i) = V(X)] &= \frac{1}{n} \cdot n \cdot V(X) - V(\bar{X}) \\
 &\stackrel{(1)}{=} V(X) - \frac{1}{n} V(X) \\
 &= \frac{n-1}{n} V(X) < V(X)
 \end{aligned}$$

- (iii) Damit ist  $\hat{S}^2$  *keine* erwartungstreue Schätzfunktion für die Varianz  $V(X)$ . Aus dem letzten Ergebnis folgt jedoch wegen der Linearität des Erwartungswertes die Aussage

$$\begin{aligned}
 E(S^2) &= E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) \\
 &= \frac{n}{n-1} E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) \\
 &= \frac{n}{n-1} E(\hat{S}^2) \\
 &\stackrel{(ii)(3)}{=} \frac{n}{n-1} \cdot \frac{n-1}{n} V(X) \\
 &= V(X),
 \end{aligned}$$

also die Erwartungstreue der Schätzfunktion  $S^2$ . Diese Schätzfunktion ist im übrigen auch konsistent.

### Ergebnis

Für den Erwartungswert  $\mu = E(X)$  und die Varianz  $\sigma^2 = V(X)$  einer Zufallsvariable  $X$  liegen damit jeweils eine erwartungstreue und konsistente Schätzfunktion vor:

Schätzfunktion für  $\mu = E(X)$ :

$$\bar{X} := T_1(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$$

$x_i$	$x_1$	$x_2$	$\dots$	$x_n$
$P(X = x_i) = f(x_i; \theta)$	$f(x_1; \theta)$	$f(x_2; \theta)$	$\dots$	$f(x_n; \theta)$

Tabelle 10.1: Wahrscheinlichkeitstabelle für  $n$  Stichprobenwerte

Schätzfunktion für  $\sigma^2 = V(X)$ :

$$S^2 := T_3(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

### 10.1.3 Maximum-Likelihood-Methode

Es gibt eine Reihe von Verfahren zur Gewinnung von Schätzfunktionen für die unbekannt Parameter einer Wahrscheinlichkeitsverteilung. Das wohl bedeutendste ist die sog. *Maximum-Likelihood-Methode*.

#### 10.1.3.1 Beschreibung des Verfahrens

- (1)  $X$  sei eine Zufallsvariable, deren Wahrscheinlichkeitsfunktion  $f(\cdot)$  noch einen unbekannt Parameter  $\theta$  enthalte. Dieser Parameter soll aus einer Zufallsstichprobe mit  $n$  voneinander unabhängigen Stichprobenwerten  $x_1, \dots, x_n$  geschätzt werden; es liegt somit der in Tabelle 10.1 beschriebene Sachverhalt vor.
- (2) Die Stichprobenwerte sind voneinander unabhängig. Damit ist die Wahrscheinlichkeit dafür, gerade diese Stichprobe zu erhalten, also eine Stichprobe zu erhalten, die gerade aus den vorliegenden Werten  $x_1, \dots, x_n$  besteht, gegeben durch

$$L(\vec{x}; \theta) := f(x_1; \theta) \cdot \dots \cdot f(x_n; \theta) = \prod_{i=1}^n f(x_i; \theta). \quad (10.1)$$

Diese Funktion  $L$  ist bei festem  $\theta$  eine Funktion der  $n$  Variablen  $x_i$  oder, die Rollen von  $x_i$  und  $\theta$  vertauscht, bei festen Stichprobenwerten  $x_1, \dots, x_n$  eine Funktion von  $\theta$ .

#### Definition 10.4

Die Funktion  $L$ , bei festem  $\vec{x} = \langle x_1, \dots, x_n \rangle$  als Funktion von  $\theta$  aufgefaßt, heißt *Likelihood-Funktion* der Stichprobenwerte.

- (3) Die Maximum-Likelihood-Methode ist nun durch die etwas vage Aussicht motiviert, daß ein Ereignis, welches beobachtet wird, d.h. für welches die Zufallsvariable  $X$  die Werte  $x_1, \dots, x_n$  der vorliegenden Stichprobe annimmt,

gerade durch einen solchen Wert für den Parameter  $\theta$  beschrieben wird, der die Wahrscheinlichkeit  $L(\vec{x}; \theta)$  für sein Eintreten möglichst groß macht.

**Definition 10.5**

- (i) Ein *Maximum-Likelihood-Schätzwert* (kurz: *ML-Schätzwert*)  $\hat{\theta}$  ist ein Wert für den unbekanntem Parameter  $\theta$  einer Wahrscheinlichkeitsverteilung, für den die Likelihood-Funktion in der Parametermenge  $\Theta$  ihr Maximum annimmt:

$$\bigwedge_{\theta \in \Theta} L(\vec{x}; \theta) \leq L(\vec{x}; \hat{\theta}).$$

- (ii) Der Schätzwert  $\hat{\theta}$  ist natürlich von  $\vec{x}$  abhängig. Die zugehörige Zufallsvariable

$$\hat{X} \longmapsto T(\hat{X}) = T(X_1, \dots, X_n)$$

heißt *Maximum-Likelihood-Schätzer* für  $\theta$ .

**Bemerkung**

- (i) Die Verwendung der ML-Methode beruht weniger auf der oben beschriebenen vagen Motivation als vielmehr auf der guten Qualität der durch sie gewonnenen Schätzfunktionen.
- (ii) Im Falle einer stetigen Zufallsvariable  $X$  muß die Likelihood-Funktion  $L(\vec{x}; \cdot)$  mit der entsprechenden stetigen Dichtefunktion  $f(\vec{x}; \cdot)$  gebildet werden.
- (iii) Man hat zur Bestimmung eines ML-Schätzwertes  $\hat{\theta}$  bei gegebener Stichprobe  $\vec{x} = x_1, \dots, x_n$  eine Extremwertaufgabe zu lösen. Das ist oft mit den Hilfsmitteln der Analysis möglich.
- (iv) Aufgrund des in (10.1) auftretenden Produktes ist es technisch oft einfacher, statt  $L(\vec{x}; \cdot)$  die Funktion  $\ln L(\vec{x}; \cdot)$  zu untersuchen, da der Logarithmus eines Produktes gleich der Summe der Logarithmen ist.

**Aufgabe**

Man begründe, warum  $L(\vec{x}, \cdot)$  genau dort ein relatives Minimum bzw. Maximum hat, wo die Funktion  $\ln L(\vec{x}, \cdot)$  ein relatives Minimum bzw. Maximum hat.

- (v) Das übliche Procedere ist daher wie folgt:
- (a) Man bilde  $L(\vec{x}; \cdot)$ , danach  $\ln L(\vec{x}; \cdot)$ .
  - (b) Man differenziere  $\ln L(\vec{x}; \cdot)$  zweimal nach der Variablen  $\theta$ .
  - (c) Man setze die erste Ableitung gleich Null (Extremwertaufgabe).
  - (d) Man überprüfe die erhaltene(n) Nullstelle(n) mit Hilfe der zweiten Ableitung auf Maximum oder Minimum (Hesse-Matrix bei mehreren Parametern).

### 10.1.3.2 Anwendung des ML-Verfahrens auf spezielle Wahrscheinlichkeitsverteilungen

**Binomialverteilung** Eine  $n$ -fache Ausführung eines Bernoulli-Experimentes führt zu einer Binomialverteilung mit der Dichtefunktion

$$f(x) := b(x; n, p) = \begin{cases} \binom{n}{x} p^x q^{n-x} & , \quad x \in \{0, \dots, n\} \\ 0 & , \quad \text{sonst} \end{cases} .$$

- (a) Um den Wert des unbekanntem Parameters  $p$  – die Erfolgswahrscheinlichkeit – abzuschätzen, wird der Grundgesamtheit eine Stichprobe  $\{x_1, \dots, x_n\}$  vom Umfang  $n$  entnommen. Die Gesamtwahrscheinlichkeit dieser Stichprobe ist gegeben durch<sup>2</sup> – es wird  $L(p)$  geschrieben anstelle von  $L(\vec{x}, p)$  –

$$\begin{aligned} L(p) &:= \prod_{i=1}^n f(x_i; p) = \prod_{i=1}^n p^{x_i} \cdot q^{1-x_i} \cdot \chi_{\{0,1\}}(x_i) \\ &= p^{\left(\sum_{i=1}^n x_i\right)} \cdot q^{\left(n - \sum_{i=1}^n x_i\right)} \cdot \chi_{\mathcal{Q}}(\vec{x}) \\ &= p^{n\bar{x}} \cdot q^{n-n\bar{x}} \cdot \chi_{\mathcal{Q}}(\vec{x}) \quad (p \in [0, 1]). \end{aligned}$$

Dieser Ausdruck ist die Likelihood-Funktion.

- (b) Logarithmieren (aus technischen Gründen):

$$\begin{aligned} L^*(p) &:= \ln[L(p)] = \ln[p^{n\bar{x}} \cdot (1-p)^{n-n\bar{x}} \cdot \chi_{\mathcal{Q}}(\vec{x})] \\ &= \ln p^{n\bar{x}} + \ln(1-p)^{n-n\bar{x}} + \ln \chi_{\mathcal{Q}}(\vec{x}) \\ &= n\bar{x} \cdot \ln p + (n - n\bar{x}) \cdot \ln(1-p) + \ln \chi_{\mathcal{Q}}(\vec{x}) \end{aligned}$$

- (c) Ableitung nach  $p$ :

$$\frac{dL^*}{dp}(p) = n\bar{x} \cdot \frac{1}{p} + (n - n\bar{x}) \cdot \frac{1}{1-p} \cdot (-1)$$

- (d) Notwendige Bedingung für ein lokales Extremum:

$$\begin{aligned} \frac{dL^*}{dp}(p) \stackrel{!}{=} 0 &\quad \longrightarrow \quad \frac{n\bar{x}}{p} = \frac{n - n\bar{x}}{1-p} \\ \longrightarrow \quad \hat{p} := \max_{[0,1]} \{p\} &= \frac{n\bar{x}}{n} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \end{aligned}$$

<sup>2</sup> $\chi_{\mathcal{S}}$  bezeichne die sog. charakteristische Funktion einer Menge  $\mathcal{S}$ , d.h.

$$\chi_{\mathcal{S}}(x) := \begin{cases} 1 & , \quad x \in \mathcal{S} \\ 0 & , \quad \text{sonst} \end{cases} ;$$

$\mathcal{Q}$  bezeichne die Eckenmenge des  $n$ -dimensionalen Einheitswürfels, also  $\mathcal{Q} = \{0, 1\}^n$ .

(e) Hinreichende Bedingung für ein lokales Extremum:

$$\frac{d^2 L^*}{dp^2}(\hat{p}) = \left( -\frac{n\bar{x}}{\hat{p}^2} - \frac{n - n\bar{x}}{(1 - \hat{p})^2} \right) \Big|_{\hat{p}=\bar{x}} = -\frac{n}{\bar{x}(1 - \bar{x})} < 0,$$

also liegt bei  $p = \hat{p}$  ein lokales Maximum vor.

Der so erhaltene Schätzwert  $\hat{p}$  für  $p$  ist nichts Anderes als die empirisch bestimmte relative Häufigkeit  $p(A)$  des Ereignisses  $A$ , welches in der  $n$ -fachen Ausführung des Bernoulli-Experimentes besteht. Die zugehörige Schätzfunktion ist die Stichprobenfunktion

$$T(X) := \frac{X}{n}$$

für die binomialverteilte Zufallsvariable

$X \iff$  Anzahl der Erfolge bei einer  $n$ -fachen  
Ausführung des Bernoulli-Experimentes

**Exponentialverteilung** Die Ermittlung einer Schätzfunktion für den Parameter  $\lambda$  einer Exponentialverteilung mit Hilfe des ML-Verfahrens ist Inhalt einer Hausaufgabe.

**Normalverteilung** Hier müssen die unbekannt Parameter  $\mu$  und  $\sigma$  einer Normalverteilung mit der Dichtefunktion

$$f(x; \mu, \sigma) := \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

anhand einer Stichprobe  $\{x_1, \dots, x_n\}$  geschätzt werden.

(a) Likelihood-Funktion:

$$\begin{aligned} L(\mu, \sigma) &:= \prod_{i=1}^n f(x_i; \mu, \sigma) \\ &= \left[ \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \right] \cdots \left[ \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(x_n-\mu)^2}{2\sigma^2}} \right] \\ &= \left( \frac{1}{\sqrt{2\pi} \cdot \sigma} \right)^n \cdot e^{-\frac{1}{2\sigma^2} \left[ \sum_{i=1}^n (x_i - \mu)^2 \right]} \end{aligned}$$

Dieser Ausdruck ist die Likelihood-Funktion, aufgefaßt als Funktion von  $\mu$  und  $\sigma$ . Zur Abkürzung werde im folgenden  $\alpha := \sum_{i=1}^n (x_i - \mu)^2$  gesetzt, so daß die Likelihood-Funktion gegeben ist durch

$$L(\mu, \sigma) = \left( \frac{1}{\sqrt{2\pi} \cdot \sigma} \right)^n \cdot e^{-\frac{\alpha}{2\sigma^2}}.$$

(b) Logarithmieren (aus technischen Gründen):

$$\begin{aligned}
 L^*(\mu, \sigma) &:= \ln[L(\mu, \sigma)] \\
 &= \ln\left[\left(\frac{1}{\sqrt{2\pi} \cdot \sigma}\right)^n \cdot e^{-\frac{\alpha}{2\sigma^2}}\right] \\
 &= \ln\left(\frac{1}{\sqrt{2\pi} \cdot \sigma}\right)^n + \ln\left(e^{-\frac{\alpha}{2\sigma^2}}\right) \\
 &= \ln 1 - \ln(\sqrt{2\pi} \cdot \sigma)^n - \frac{\alpha}{2\sigma^2} \\
 &= -n \cdot \ln(\sqrt{2\pi} \cdot \sigma) - \frac{\alpha}{2\sigma^2} \\
 &= -n \cdot (\ln \sqrt{2\pi} + \ln \sigma) - \frac{1}{2\sigma^2} \cdot \sum_{i=1}^n (x_i - \mu)^2
 \end{aligned}$$

(c) Partielle Ableitungen nach  $\mu$  und  $\sigma$ :

$$\begin{aligned}
 \frac{\partial L^*}{\partial \mu}(\mu, \sigma) &= -\frac{1}{2\sigma^2} \cdot \sum_{i=1}^n 2(x_i - \mu) \cdot (-1) \\
 &= \frac{1}{\sigma^2} \cdot \sum_{i=1}^n (x_i - \mu) \\
 \frac{\partial L^*}{\partial \sigma}(\mu, \sigma) &= -n \cdot \frac{1}{\sigma} - \frac{1}{2}(-2\sigma^{-3}) \cdot \sum_{i=1}^n (x_i - \mu)^2 \\
 &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \cdot \sum_{i=1}^n (x_i - \mu)^2 \\
 \frac{\partial^2 L^*}{\partial \mu^2}(\mu, \sigma) &= -\frac{n}{\sigma^2} \\
 \frac{\partial^2 L^*}{\partial \mu \partial \sigma}(\mu, \sigma) &= \frac{\partial^2 L^*}{\partial \sigma \partial \mu}(\mu, \sigma) = -\frac{2}{\sigma^3} \cdot \sum_{i=1}^n (x_i - \mu) \\
 \frac{\partial^2 L^*}{\partial \sigma^2}(\mu, \sigma) &= \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \cdot \sum_{i=1}^n (x_i - \mu)^2
 \end{aligned}$$

(d) Notwendige Bedingung für ein lokales Extremum:

$$\begin{aligned}
 \frac{\partial L^*}{\partial \mu}(\mu, \sigma) \stackrel{!}{=} 0 &\quad \longrightarrow \quad \hat{\mu} := \frac{1}{n} \cdot \sum_{i=1}^n x_i = \bar{x} \\
 \frac{\partial L^*}{\partial \sigma}(\mu, \sigma) \stackrel{!}{=} 0 &\quad \longrightarrow \quad \hat{\sigma}^2 := \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \hat{\mu})^2
 \end{aligned}$$

Die Maximum-Likelihood-Schätzung für den Erwartungswert  $\mu$  führt auf den arithmetischen Mittelwert  $\bar{x}$  der Stichprobe als Schätzwert, die zu-



gehörige Schätzfunktion ist die erwartungstreue Stichprobenfunktion

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

Unter Berücksichtigung des Ergebnisses  $\hat{\mu} = \bar{x}$  der ersten Gleichung der notwendigen Bedingung führt die ML-Methode für die Varianz  $\sigma^2$  mittels der zweiten der obigen Gleichung auf den Schätzwert  $\hat{\sigma}^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$ . Die dieser Schätzung zugrundeliegende Maximum-Likelihood-Schätzfunktion

$$(\hat{S})^2 = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^2$$

ist jedoch *nicht* erwartungstreu. Sie unterscheidet sich von der erwartungstreuen Schätzfunktion

$$(S)^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - \bar{X})^2$$

um den konstanten Faktor  $\frac{n-1}{n}$ . Es gilt

$$(\hat{S})^2 = \frac{n-1}{n} \cdot S^2.$$

(e) Hinreichende Bedingung für ein lokales Extremum:

$$\begin{aligned} H_{L^*}(\mu, \sigma) &:= \begin{pmatrix} \frac{\partial^2 L^*}{\partial \mu^2}(\mu, \sigma) & \frac{\partial^2 L^*}{\partial \sigma \partial \mu}(\mu, \sigma) \\ \frac{\partial^2 L^*}{\partial \mu \partial \sigma}(\mu, \sigma) & \frac{\partial^2 L^*}{\partial \sigma^2}(\mu, \sigma) \end{pmatrix} \Big|_{\mu=\hat{\mu}=\bar{x}, \sigma=\hat{\sigma}^2} \\ &= \begin{pmatrix} -\frac{n}{\hat{\sigma}^2} & -\frac{2}{\hat{\sigma}^3} \cdot \sum_{i=1}^n (x_i - \bar{x}) \\ -\frac{2}{\hat{\sigma}^3} \cdot \sum_{i=1}^n (x_i - \bar{x}) & \frac{n}{\hat{\sigma}^2} - \frac{3}{\hat{\sigma}^4} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 \end{pmatrix} \\ &= \begin{pmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{2n}{\hat{\sigma}^2} \end{pmatrix}, \end{aligned}$$

also liegt bei

$$\hat{\mu} = \bar{x} \quad , \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

ein lokales Maximum vor.

## 10.1.4 Konfidenzbereiche

### 10.1.4.1 Vertrauensintervalle und statistische Sicherheit

In den letzten beiden Abschnitten wurde einerseits untersucht, wie man aus einer konkreten Stichprobe mit Hilfe geeigneter Schätzfunktionen – etwa  $\bar{X}$

oder  $S^2$  – Näherungs- oder Schätzwerte für die betreffenden Parameter erhält, andererseits, wie man solche Schätzfunktionen denn auch erhalten kann (ML-Verfahren). Diese sog. *Punktschätzungen* ermöglichen jedoch keinerlei Aussagen über die Genauigkeit oder Sicherheit einer solchen Schätzung.

Es liegt daher nahe, anhand einer konkreten Stichprobe ein Intervall  $[c_u, c_o]$  zu bestimmen, welches den unbekannt Parameter  $\theta$  enthält, entweder mit Sicherheit ( $\gamma = 1$ ) oder mit einer Wahrscheinlichkeit  $\gamma \in (0, 1)$ , s. Abb. 10.1.



Abbildung 10.1: „Fixes Vertrauensintervall“ für eine Punktschätzung

### Beh. 1

Es existiert kein nichttriviales Intervall  $\mathcal{I}$ , welches den unbekannt Parameter mit Sicherheit, d.h. mit Wahrscheinlichkeit  $\gamma = 1$  enthält.

#### Beweis

Absolut sichere Rückschlüsse von einer Zufallsstichprobe auf die Grundgesamtheit sind grundsätzlich nicht möglich, von Trivialfällen abgesehen (etwa: mit Wahrscheinlichkeit  $\gamma = 1$  ist der unbekannt Parameter im Intervall  $(-\infty, \infty)$ ).

### Beh. 2

Es existiert kein nichttriviales Intervall  $\mathcal{I}$ , welches den unbekannt Parameter mit einer Wahrscheinlichkeit  $\gamma$  ( $\gamma \in (0, 1)$  geeignet) enthält.

#### Beweis

Die Aussage  $\theta \in [c_u, c_o]$  ist

$$\begin{aligned} \text{entweder richtig} & \iff P(c_u \leq \theta \leq c_o) = 1 \\ \text{oder falsch} & \iff P(c_u \leq \theta \leq c_o) = 0, \end{aligned}$$

man weiß es nur nicht. Hier darf man nicht den Grad der Unwissenheit mit einem Maß versehen und dieses als Wahrscheinlichkeit – der sog. *subjektiven Wahrscheinlichkeit* – interpretieren, mit der die Schätzung „sicher“ ist, sondern muß von den Daten einer konkreten Stichprobe ausgehen. Der Fall  $P(c_u \leq \theta \leq c_o) = \gamma$  mit von der jeweiligen Stichprobe festen unabhängigen Intervallgrenzen  $c_u$  und  $c_o$  ist also nicht möglich, falls man  $\gamma \in (0, 1)$  wählt.

**Ausweg**

- (i) Man gibt eine Wahrscheinlichkeit  $\gamma$  vor (die sog. *Vertrauenswahrscheinlichkeit*) und bestimmt dazu zwei *variable* Intervallgrenzen  $c_u$  und  $c_o$ , welche – außer von  $\gamma$  – nur noch von der vorliegenden Stichprobe abhängen. Diese beiden Intervallgrenzen schließen dann den wahren, aber unbekannt Parameter  $\theta$  mit der Wahrscheinlichkeit  $\gamma$  ein:

$$P(c_u \leq \theta \leq c_o) = \gamma.$$

- (ii) Diese beiden variablen Intervallgrenzen erhält man als Realisierungen

$$\begin{aligned} c_u &:= g_u(x_1, \dots, x_n) \\ c_o &:= g_o(x_1, \dots, x_n) \end{aligned}$$

der beiden Stichprobenfunktionen

$$\begin{aligned} \Theta_u &:= g_u(X_1, \dots, X_n) \\ \Theta_o &:= g_o(X_1, \dots, X_n) \end{aligned}$$

als deren Auswertung auf einer konkreten Stichprobe vom Umfang  $n$ . Diese Stichprobenfunktionen sind Schätzfunktionen zur Bestimmung der beiden Intervallgrenzen. Zugleich sind sie natürlich auch Schätzfunktionen des Parameters  $\theta$ , wenn auch keine sehr guten!

**Definition 10.6**

- (i) Das Intervall  $[c_u, c_o]$  heißt *Vertrauens- oder Konfidenzintervall*. Die Intervallrandpunkte  $c_u$  und  $c_o$  heißen *Vertrauens- oder Konfidenzgrenzen*.
- (ii)  $\gamma$  heißt *statistische Sicherheit, Vertrauens- oder Konfidenzniveau*.
- (iii)  $\alpha := 1 - \gamma$  heißt *Irrtumswahrscheinlichkeit*.

**Bemerkung**

- (i) Das *vor* einer Stichprobenuntersuchung gewählte Konfidenzniveau  $\gamma = 1 - \alpha$  ist die Wahrscheinlichkeit dafür, aus einer konkreten Stichprobe ein Konfidenzintervall zu erhalten, welches den wahren, aber unbekannt Parameter  $\theta$  enthält. In der Praxis wählt man meistens  $\gamma = 0.95$  oder  $\gamma = 0.99$ .
- (ii) Die Konfidenzgrenzen  $c_u$  und  $c_o$  hängen von den Stichprobenwerten  $x_1, \dots, x_n$  ab. I. a. verändern sie sich also mit jeder neuen Stichprobe. Von 100 Konfidenzintervallen enthalten dann ungefähr  $\gamma \cdot 100$  Intervalle den wahren Wert, und ungefähr  $\alpha \cdot 100$  Intervalle enthalten ihn nicht. M. a. W., in rund  $\gamma \cdot 100$  Fällen wird eine richtige, in rund  $\alpha \cdot 100$  Fällen eine falsche Entscheidung getroffen.

- (iii) Die Konfidenzgrenzen hängen nicht nur von den Stichprobenwerten  $x_1, \dots, x_n$  ab sondern auch noch vom gewählten Konfidenzniveau  $\gamma = 1 - \alpha$ . Dabei gilt:

Je größer  $\gamma$ , um so länger werden die Konfidenzintervalle. Eine große statistische Sicherheit wird erkaufte durch ein großes Konfidenzintervall. Es besteht also ein Gegensatz zwischen der Schärfe einer Aussage und der Sicherheit, mit der diese Aussage getroffen werden kann:

Sichere Aussagen (Aussagen mit einer hohen Wahrscheinlichkeit) sind unscharf (große Bereiche); scharfe Aussagen sind unsicher.

Die Länge  $l := c_o - c_u$  des Konfidenzintervalles, also das Maß für die Genauigkeit der Parameterschätzung, verringert sich allerdings mit zunehmendem Stichprobenumfang  $n$ .

#### 10.1.4.2 Erwartungswert einer Normalverteilung bei bekannter Varianz

$X$  sei eine  $N(\mu, \sigma^2)$ -verteilte Zufallsvariable, also eine normalverteilte Zufallsvariable mit dem Erwartungswert  $E(X) = \mu$  und der Varianz  $V(X) = \sigma^2$ . In diesem Abschnitt wird vorausgesetzt, daß die Varianz bekannt, der Erwartungswert aber unbekannt ist und anhand einer vorliegenden Stichprobe geschätzt werden muß.

##### Ziel

Bestimmung eines Konfidenzintervalles für  $\mu$

##### Praktische Bedeutung

Bei vielen Meßinstrumenten, Maschinen oder Automaten läßt sich die Varianz  $\sigma^2$  als Gerätekonstante ansehen, als eine Art Güte des Gerätes. Diese ändert sich nicht, wenn beispielsweise eine Produktionsmaschine auf eine andere Größennorm für die zu produzierenden Teile umgestellt wird. Der Erwartungswert  $\mu$  wird als neuer Sollwert zwar ebenfalls eingestellt, es ist aber unbekannt, wie gut er dann tatsächlich eingehalten wird.

Als Schätzfunktion für den Erwartungswert wird die erwartungstreue Stichprobenfunktion

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

verwendet.

##### Beh. 1

$\bar{X}$  ist  $N(\mu, \frac{\sigma^2}{n})$  - verteilt.

*Beweis* (s. auch den Beweis von Satz 10.1, p. 164)

$X$  und die Kopien  $X_i$  sind unabhängig und  $N(\mu, \sigma^2)$ -verteilt. Eine nichttriviale Linearkombination von unabhängigen normalverteilten Zufallsvariablen ist wieder normalverteilt<sup>3</sup>. Für den Erwartungswert und die Varianz unabhängiger Zufallsvariablen gilt darüberhinaus (s. Fußnote auf p. 165)

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) \\ &= \frac{1}{n} \sum_{i=1}^n E(X) = \frac{1}{n} n E(X) = E(X), \\ V(\bar{X}) &= V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) \\ &= \frac{1}{n^2} \sum_{i=1}^n V(X) = \frac{1}{n^2} n V(X) = \frac{1}{n} V(X). \end{aligned}$$

## Beh. 2

Die zugehörige standardisierte Zufallsvariable

$$Z := \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

ist dann standardnormalverteilt, d.h. es gilt

$$E(Z) = 0 \quad , \quad V(Z) = 1.$$

*Beweis*

Lineare Funktionen von normalverteilten Zufallsvariablen sind normalverteilt, und die Behauptungen über den Erwartungswert und die Varianz ergeben sich wie folgt, s. auch Satz 6.3, p. 125:

$$\begin{aligned} E(Z) &= \frac{1}{\sigma/\sqrt{n}} \cdot (E(\bar{X}) - \mu) = \frac{1}{\sigma/\sqrt{n}} \cdot (\mu - \mu) = 0 \\ V(Z) &= \left(\frac{1}{\sigma/\sqrt{n}}\right)^2 \cdot V(\bar{X}) = \left(\frac{1}{\sigma/\sqrt{n}}\right)^2 \cdot \frac{\sigma^2}{n} = 1 \end{aligned}$$

<sup>3</sup>Greiner/Tinhofer: Stochastik für Studienanfänger der Informatik, Hanser-Verlag 1996, Satz 5.4.2, p. 123

**Konstruktion eines Konfidenzintervalles für  $\mu$** 

- (i) Wahl eines Vertrauensniveaus
- $\gamma$
- . In der Praxis übliche Werte sind

$$\gamma = 0.95 = 95\% \quad \text{oder} \quad \gamma = 0.99 = 99\%$$

- (ii) Die standardisierte Zufallsvariable
- $Z$
- soll dann mit der gewählten Wahrscheinlichkeit
- $\gamma = 1 - \alpha$
- einen Wert aus dem symmetrischen Intervall
- $[-c, c]$
- annehmen, es muß also gelten:

$$P(-c \leq Z \leq c) = \gamma = 1 - \alpha.$$

- (iii) Unter Verwendung der Definition von
- $Z$
- läßt sich das Intervall darstellen in der Form

$$-c \leq Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq c$$

$$\iff$$

$$\bar{X} - c \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + c \cdot \frac{\sigma}{\sqrt{n}}.$$

Die Bedingung (ii) lautet also

$$P\left(\underbrace{\bar{X} - c \cdot \frac{\sigma}{\sqrt{n}}}_{\Theta_u} \leq \mu \leq \underbrace{\bar{X} + c \cdot \frac{\sigma}{\sqrt{n}}}_{\Theta_o}\right) = 1 - \alpha.$$

Die beiden Zufallsvariablen

$$\Theta_u := g_u(X_1, \dots, X_n) := \bar{X} - c \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Theta_o := g_o(X_1, \dots, X_n) := \bar{X} + c \cdot \frac{\sigma}{\sqrt{n}}$$

sind dabei die noch offen gebliebenen Schätzfunktionen für die gesuchten Vertrauensgrenzen  $c_u$  und  $c_o$ .

- (iv) Die Größe
- $\frac{\sigma}{\sqrt{n}}$
- heißt
- Standardfehler*
- oder
- Stichprobenfehler*
- ; dieser ist ein Streuungsmaß für die Stichprobenverteilung und demzufolge ein Gütemaß für den berechneten Mittelwert.

- (v) Die Berechnung dieser Vertrauensgrenzen erfolgt dann anhand einer konkreten Stichprobe
- $\{x_1, \dots, x_n\}$
- dadurch, daß man den aus ihr berechneten empirischen Mittelwert
- $\bar{x}$
- als Argument in diese Schätzfunktionen einsetzt. Man erhält

$$c_u := \bar{x} - c \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + c \cdot \frac{\sigma}{\sqrt{n}} =: c_o.$$

Das Intervall  $[c_u, c_o]$  der Länge

$$l := \frac{2c\sigma}{\sqrt{n}},$$

welches symmetrisch bzgl. des arithmetischen Mittelwertes  $\bar{x}$  der aktuellen Stichprobe ist, enthält den wahren, aber unbekanntem Erwartungswert  $\mu = E(X)$  mit der Wahrscheinlichkeit (dem „Vertrauen“)  $\gamma \cdot 100\%$ .

Wie man sieht, läßt sich eine Verkürzung der Länge  $l$  des Vertrauensintervalles bei Aufrechterhaltung der Vertrauenswahrscheinlichkeit  $\gamma$  stets auch dadurch erreichen, daß man den Stichprobenumfang  $n$  vergrößert.

(vi) Zur Konstante  $c$ :

$$P(c_u \leq \mu \leq c_o) = P\left(\bar{x} - c \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + c \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Das Problem, ein Konfidenzintervall für  $\mu$  zu bestimmen, ist damit zurückgeführt auf das Problem, die Lösung  $c$  der Gleichung

$$P(-c \leq Z \leq c) = \Phi(c) - \Phi(-c) = 1 - \alpha$$

zu bestimmen, wobei  $\Phi$  die Verteilungsfunktion der Standardnormalverteilung ist:

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

Wegen  $\Phi(x) = 1 - \Phi(-x)$  ( $x \in \mathbb{R}$ ) ist obige Bedingung äquivalent zu

$$\Phi(c) = 1 - \frac{\alpha}{2} \iff c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right),$$

$c$  ist also das  $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung, s. Abb. 10.2.

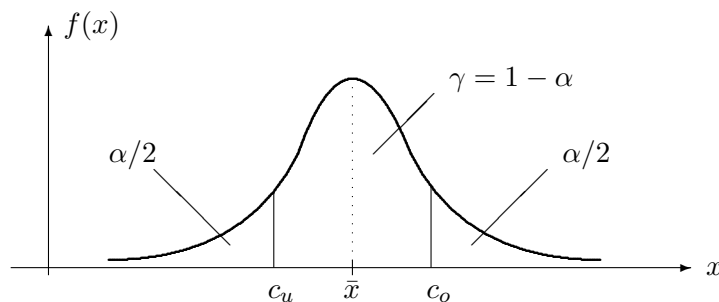


Abbildung 10.2: Vertrauensintervall bei einer normalverteilten Zufallsvariable

$$\bar{x} - c \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + c \cdot \frac{\sigma}{\sqrt{n}} \quad (10.2)$$

$i$	1	2	3	4	5	6	7	8	9	10	11	12
$x_i$ [mm]	8	10	9	10	11	12	9	8	11	10	12	11

Tabelle 10.2: Längenmessung von 12 Verbindungsnieten

**Beispiel 10.3**

Zwölf Verbindungsniete werden zufällig aus einem Karton ausgewählt und ihre Länge bestimmt. Es ergebe sich das Ergebnis aus Tabelle 10.2.

Vorausgesetzt wird, daß diese Stichprobe aus einer bzgl. der Längenmessung normalverteilten Grundgesamtheit mit der Varianz  $\sigma^2 = (2 \text{ mm})^2$  stammt. Für den unbekanntem Erwartungswert  $\mu$  der Verteilung soll ein Vertrauensintervall zum Vertrauensniveau  $\gamma = 0.95 = 95\%$  bestimmt werden.

**Konstruktion eines Konfidenzintervalles für  $\mu$** 

- (i) Das Vertrauensniveau  $\gamma$  ist bereits vorgegeben:  $\gamma = 0.95$ .
- (ii) Bestimmung des Quantils  $c$ :

$$\begin{aligned}
 P(-c \leq Z \leq c) &= \Phi(c) - \Phi(-c) = \Phi(c) - [1 - \Phi(c)] \\
 &= 2\Phi(c) - 1 = 0.95 \\
 \Phi(c) = 0.975 &\longrightarrow c = \Phi^{-1}(0.975) = 1.960
 \end{aligned}$$

- (iii) Mittelwert der Stichprobe:

$$\bar{x} = \frac{1}{12} \sum_{i=1}^{12} x_i = \frac{121}{12} \text{ mm} = 10.083 \text{ mm}$$

- (iv) Berechnung der Intervallgrenzen:

$$k := c \cdot \frac{\sigma}{\sqrt{n}} = 1.960 \cdot \frac{2 \text{ mm}}{\sqrt{12}} = 1.132 \text{ mm}$$

Das Vertrauensintervall lautet dann

$$\mathcal{I} := [\bar{x} - k, \bar{x} + k] = [10.083 - 1.132, 10.083 + 1.132] = [8.951, 11.215],$$

und der unbekanntem Erwartungswert  $\mu$  ist mit einer Wahrscheinlichkeit von  $\gamma = 0.95$  darin enthalten.



**Beispiel 10.4**

Einer normalverteilten Grundgesamtheit mit der Varianz  $\sigma^2 = 25$  soll eine Stichprobe vom Umfang  $n$  entnommen werden. Wie groß muß der Stichprobenumfang gewählt werden, damit das Vertrauensintervall für den unbekanntem Erwartungswert  $\mu$  bei einem Vertrauensniveau von  $\gamma = 0.99$  höchstens die Länge  $l_{\max} = 4$  besitzt?

**Lösung**

Die Länge des Vertrauensintervalles ist gegeben durch  $l := 2c \frac{\sigma}{\sqrt{n}}$ , also gilt

$$\sqrt{n} = \frac{2c\sigma}{l} \quad \Leftrightarrow \quad n = \left(\frac{2c\sigma}{l}\right)^2 \geq \left(\frac{2c\sigma}{l_{\max}}\right)^2.$$

Die noch unbekannte Konstante  $c$  ist das  $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung, also gilt mit  $\alpha = 1 - \gamma = 0.01$ :

$$c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = \Phi^{-1}(0.995) = 2.576.$$

Damit muß der gesuchte Stichprobenumfang  $n$  mindestens

$$n \geq \left(\frac{2c\sigma}{l_{\max}}\right)^2 = \left(\frac{2 \cdot 2.576 \cdot 5}{4}\right)^2 = 41.47$$

betragen, also  $n \geq 42$ .

Häufig ist man nicht an einer zweiseitigen Abgrenzung für den unbekanntem Erwartungswert interessiert sondern nur an einer einseitigen Abgrenzung nach unten oder nach oben. Diese Betrachtungen führen auf sog. *einseitige Konfidenzintervalle*, und zwar

$$\begin{aligned} \text{einseitiges unteres Konfidenzintervall} \quad \mathcal{I}_u &= (-\infty, c_o], \\ \text{einseitiges oberes Konfidenzintervall} \quad \mathcal{I}_o &= [c_u, \infty). \end{aligned}$$

**Beispiel 10.5**

Das Gewicht eines Waschmittels sei eine  $N(\mu, \sigma^2)$ -verteilte Zufallsvariable mit der Varianz  $V(X) = \sigma^2 = (0.2 \text{ kg})^2$ . Von einer Tagesproduktion von 20 000 Paketen werde eine Stichprobe vom Umfang  $n = 15$  genommen und das Gewicht eines jeden einzelnen Waschmittelpaketes dieser Stichprobe in kg gemessen. Man erhalte als Mittelwert dieser Stichprobe den Wert

$$\bar{x} = \frac{1}{15} \sum_{i=1}^{15} x_i = \dots = 10.05 \text{ kg.}$$

Man ermittle jeweils ein

- (i) zweiseitiges,
- (ii) einseitiges unteres
- (iii) einseitiges oberes

95 %-Konfidenzintervall für den unbekanntem Erwartungswert  $\mu$ .

### Lösung

$$\bar{x} = 10.05 \text{ kg} \quad , \quad \sigma = 0.2 \text{ kg} \quad , \quad n = 15 \quad , \quad \alpha = 0.05$$

- (i) zweiseitiges 95 %-Konfidenzintervall:

$$\begin{aligned} P\left(-c \leq Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq c\right) &= 1 - \alpha \\ &\iff \\ P\left(\bar{X} - c \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + c \cdot \frac{\sigma}{\sqrt{n}}\right) &= 1 - \alpha \end{aligned}$$

$(1 - \alpha/2)$ -Quantil  $c$  der Standardnormalverteilung:

$$\begin{aligned} \Phi(c) = 1 - \frac{\alpha}{2} &\iff c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \\ c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) &= \Phi^{-1}(0.975) = 1.960 \end{aligned}$$

Man erhält

$$\begin{aligned} k &:= c \cdot \frac{\sigma}{\sqrt{n}} = 1.960 \cdot \frac{0.2}{\sqrt{15}} = 0.1012 \\ \mathcal{I}_{0.95}^z &= [10.05 - 0.1012, 10.05 + 0.1012] = [9.95, 10.15] . \end{aligned}$$

- (ii) einseitiges unteres 95 %-Konfidenzintervall:

Einseitige Konfidenzintervalle der Form  $(-\infty, c_o]$  oder  $[c_u, \infty)$  erhält man, indem man den Flächenanteil nur auf einer Seite der Normalverteilung abschneidet, s. Abb. 10.3.

$$\begin{aligned} P\left(-\infty < Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq c\right) &= 1 - \alpha \\ \iff P\left(\bar{X} - c \cdot \frac{\sigma}{\sqrt{n}} \leq \mu < \infty\right) &= 1 - \alpha \end{aligned}$$

$(1 - \alpha)$ -Quantil  $c$  der Standardnormalverteilung:

$$\begin{aligned} P(Z \leq c) &= \Phi(c) = 1 - \alpha \\ \iff c &= \Phi^{-1}(1 - \alpha) = \Phi^{-1}(0.95) = 1.645 \end{aligned}$$

Man erhält

$$k := c \cdot \frac{\sigma}{\sqrt{n}} = 1.645 \cdot \frac{0.2}{\sqrt{15}} = 0.085$$

$$\mathcal{I}_{0.95}^u = (-\infty, 10.05 + 0.085] = (-\infty, 10.135].$$

(iii) einseitiges oberes 95 %-Konfidenzintervall:

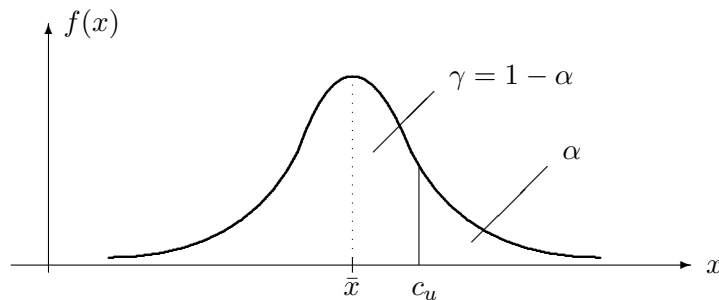


Abbildung 10.3: Unteres Vertrauensintervall bei einer normalverteilten Zufallsvariable

$$P\left(-c \leq Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \infty\right) = 1 - \alpha$$

$$\Leftrightarrow P\left(-\infty < \mu \leq \bar{X} + c \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Wieder muß jetzt das  $(1 - \alpha)$ -Quantil  $c$  der Standardnormalverteilung bestimmt werden, was aber schon unter (ii) gemacht worden ist:

$$\begin{aligned} P(-c \leq Z) &= 1 - P(Z \leq -c) \\ &= 1 - \Phi(-c) \\ &= 1 - [1 - \Phi(c)] \\ &= \Phi(c) = 1 - \alpha \end{aligned}$$

$$\Leftrightarrow c = \Phi^{-1}(1 - \alpha) = \Phi^{-1}(0.95) = 1.645$$

Man erhält somit hier

$$k := c \cdot \frac{\sigma}{\sqrt{n}} = 1.645 \cdot \frac{0.2}{\sqrt{15}} = 0.085$$

$$\mathcal{I}_{0.95}^o = [10.05 - 0.085, \infty) = [9.965, \infty).$$

**Bemerkung**

Die obere Konfidenzgrenze  $c_o$  des *unteren einseitigen* Konfidenzintervalles ist kleiner als die obere Konfidenzgrenze  $c_o$  des *zweiseitigen* Konfidenzintervalles. Analog ist die untere Konfidenzgrenze  $c_u$  des *oberen einseitigen* Konfidenzintervalles größer als die untere Konfidenzgrenze  $c_u$  des *zweiseitigen* Konfidenzintervalles, s. Abb. 10.4. Woran liegt das?

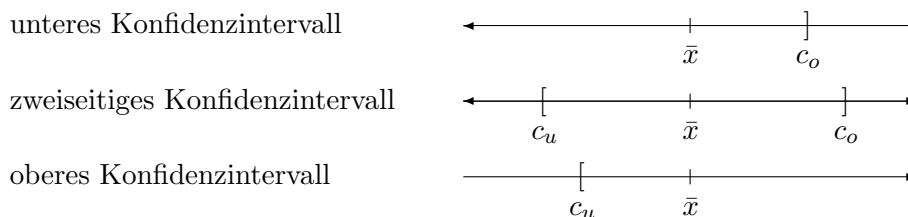


Abbildung 10.4: Vertrauensintervalle im Vergleich

**10.1.4.3 Erwartungswert einer Normalverteilung bei unbekannter Varianz**

$X$  sei eine  $N(\mu, \sigma^2)$ -verteilte Zufallsvariable, für deren unbekanntes Erwartungswert  $\mu$  anhand einer konkreten Stichprobe ein Vertrauensintervall angegeben werden soll.

**Problem**

Im Gegensatz zum letzten Abschnitt ist hier die Varianz  $\sigma^2 = V(X)$  und damit die Standardabweichung  $\sigma$  unbekannt. Auch diese Größen müssen daher anhand der vorliegenden konkreten Stichprobe geschätzt werden. Wie?

**Procedere**

- (i) Bei der Herleitung eines Vertrauensintervalles für den Erwartungswert  $\mu$  muß die bisher verwendete standardnormalverteilte Zufallsvariable

$$Z := \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

ersetzt werden durch die Zufallsvariable

$$T := \frac{\bar{X} - \mu}{S/\sqrt{n}};$$

hierbei ist  $S$  die erwartungstreue Schätzfunktion für die unbekannte Standardabweichung der Grundgesamtheit:

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}.$$

Welcher Verteilung genügt die so modifizierte Zufallsvariable  $T$ ?

(ii) Es läßt sich zeigen, daß die Zufallsvariable

$$Y := \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{\sigma^2} \cdot S^2$$

einer  $\chi^2$ -Verteilung genügt. Diese hat allerdings nur  $n-1$  Freiheitsgrade, da bei festem  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  durch die Abhängigkeit der  $n$  Summanden  $X_i$  ein Freiheitsgrad gebunden wird und somit nur  $n-1$  Zufallsvariablen frei wählbar sind. Dann jedoch besitzt die Zufallsvariable

$$T := \frac{Z}{\sqrt{Y/(n-1)}}$$

definitionsgemäß eine  $t$ -Verteilung mit  $n-1$  Freiheitsgraden. Nun gilt aber

$$T = \frac{Z}{\sqrt{Y/(n-1)}} = \sqrt{n-1} \cdot \frac{Z}{\sqrt{Y}} = \sqrt{n-1} \cdot \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\frac{\sqrt{n-1}}{\sigma} \cdot S} = \frac{\bar{X}-\mu}{S/\sqrt{n}}.$$

(iii) Die weitere Vorgehensweise bei der Konstruktion eines Konfidenzintervalles entspricht der Methode, wie sie im letzten Abschnitt bei bekannter Varianz verwendet wurde. Hier sind im Unterschied dazu lediglich die Quantile der entsprechenden  $t$ -Verteilung zu verwenden:

$$\begin{aligned} P(-c \leq T \leq c) &= \gamma = 1 - \alpha \\ \longrightarrow F_t(c) - F_t(-c) &= F_t(c) - (1 - F_t(c)) = 2F_t(c) - 1 = 1 - \alpha \\ \longrightarrow F_t(c) &= 1 - \frac{\alpha}{2}, \end{aligned}$$

und die gesuchte Konstante

$$c := F_{t;n-1}^{-1}\left(1 - \frac{\alpha}{2}\right)$$

ist das  $(1 - \frac{\alpha}{2})$ -Quantil der  $t$ -Verteilung mit  $n-1$  Freiheitsgraden. Sie heißt auch *Student-Faktor*. Das Konfidenzintervall für den unbekanntem Erwartungswert  $\mu$  hat dann die Gestalt

$$\boxed{\bar{x} - c \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + c \cdot \frac{s}{\sqrt{n}}} \quad (10.3)$$

### Beispiel 10.6

Die Inhalte von sieben Milchkannen seien wie folgt verteilt, s. Tabelle 10.3.

Unter der Annahme, daß die Verteilung eine Normalverteilung ist, finde man ein 95 %-Konfidenzintervall für den unbekanntem Erwartungswert  $\mu$ .

Kanne	1	2	3	4	5	6	7
$V[1]$	9.8	10.2	10.4	9.8	10.0	10.2	9.6

Tabelle 10.3: Inhalte von sieben Milchkannen

**Lösung**

$$\bar{x} = \frac{1}{7} \sum_{i=1}^7 x_i = 10.0$$

$$s = \sqrt{\frac{1}{7-1} \sum_{i=1}^7 (x_i - 10.0)^2} = \sqrt{0.08} = 0.283$$

$$c = F_{t;7-1}^{-1}\left(1 - \frac{0.05}{2}\right) = F_{t;6}^{-1}(0.975) = 2.447$$

$$k := c \cdot \frac{s}{\sqrt{n}} = 2.447 \cdot \frac{0.283}{\sqrt{7}} = 0.2617$$

Damit ist das 95 %-Konfidenzintervall für den unbekanntem Erwartungswert gegeben durch

$$\mathcal{I}_{(0.95)} = [(10.00 - 0.26)1, (10.00 + 0.26)1] = [9.741, 10.261].$$

**10.1.4.4 Varianz einer Normalverteilung**

Ähnlich wie für den Erwartungswert  $\mu$  läßt sich auch für die Varianz  $\sigma^2 = V(X)$  einer normalverteilten Grundgesamtheit anhand einer konkreten Stichprobe ein Konfidenzintervall bestimmen. Als Schätzfunktion für die Varianz wird die erwartungstreue Stichprobenfunktion

$$S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

für den Schätzwert

$$s^2 := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

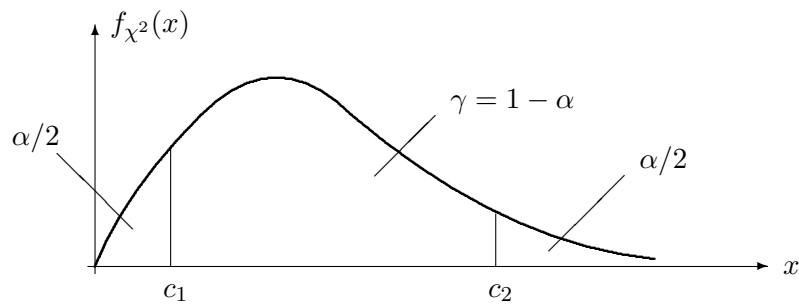
verwendet. Damit wird dann die Zufallsvariable

$$\chi^2 := Y := (n-1) \frac{S^2}{\sigma^2}$$

gebildet, von der man zeigen kann, daß sie einer  $\chi^2$ -Verteilung mit  $n-1$  Freiheitsgraden genügt.

Die Berechnung eines Konfidenzintervalles erfolgt dabei aus der Bedingung

$$P(c_1 \leq \chi^2 \leq c_2) = \gamma = 1 - \alpha.$$

Abbildung 10.5: Vertrauensintervall einer  $\chi^2$ -verteilten Zufallsvariable

Hierbei wird der Flächenanteil der Irrtumswahrscheinlichkeit  $\alpha$  anteilmäßig gleich auf beide Seiten der Dichtefunktion verteilt, s. Abb. 10.5.

Für die Intervallgrenzen  $c_1$  und  $c_2$  erhält man dann

$$\begin{aligned} P(\chi^2 \leq c_1) &= F_{\chi^2; n-1}(c_1) = \frac{\alpha}{2} \\ P(\chi^2 \leq c_2) &= F_{\chi^2; n-1}(c_2) = 1 - \frac{\alpha}{2}. \end{aligned}$$

Aus diesen beiden Gleichungen lassen sich  $c_1$  und  $c_2$  mit Hilfe der tabellierten Verteilungsfunktion  $F(\cdot)_{\chi^2}$  der  $\chi^2$ -Verteilung mit  $f = n - 1$  Freiheitsgraden leicht berechnen. Aus der Ungleichung

$$c_1 \leq \chi^2 = (n-1) \frac{S^2}{\sigma^2} \leq c_2$$

folgt dann eine entsprechende Ungleichung für die unbekannte Varianz  $\sigma^2$ :

$$(n-1) \frac{S^2}{c_2} \leq \sigma^2 \leq (n-1) \frac{S^2}{c_1} \quad (10.4)$$

### Bemerkung

- (i) Das Vertrauensintervall besitzt die Länge

$$l := (n-1) s^2 \frac{c_2 - c_1}{c_1 \cdot c_2}.$$

- (ii) Aus diesem Vertrauensintervall erhält man durch Radizieren ein Vertrauensintervall für die Standardabweichung  $\sigma$ .

### Beispiel 10.7

Es wird noch einmal Beispiel 10.6 aufgegriffen und dafür ein Vertrauensintervall für die unbekannte Varianz  $\sigma^2$  berechnet.

1. Das Vertrauensniveau ist  $\gamma = 0.95$ , die Irrtumswahrscheinlichkeit  $\alpha = 0.05$ .
2. Ermittlung der Konstanten  $c_1$  und  $c_2$ :

$$\begin{aligned} P(\chi^2 \leq c_1) &= F_{\chi^2;6}(c_1) = \frac{\alpha}{2} = 0.025 \\ P(\chi^2 \leq c_2) &= F_{\chi^2;6}(c_2) = 1 - \frac{\alpha}{2} = 0.975 \end{aligned}$$

Unter Verwendung der Tabelle für die Quantile der  $\chi^2$ -Verteilung erhält man für die  $f = n - 1 = 7 - 1 = 6$  Freiheitsgrade die folgenden Werte:

$$\begin{aligned} c_1 &= F_{\chi^2;6}^{-1}(0.025) = 1.24 \\ c_2 &= F_{\chi^2;6}^{-1}(0.975) = 14.45 \end{aligned}$$

3. Die empirische Standardabweichung der Stichprobe wurde in Beispiel 10.6 zu  $s = 0.283$  ermittelt mit der empirischen Varianz  $s^2 = 0.08$ .
4. Berechnung der beiden Vertrauensgrenzen:

$$\begin{aligned} k_1 &:= (n-1) \frac{s^2}{c_2} = 6 \cdot \frac{0.08}{14.45} = 0.033 \\ k_2 &:= (n-1) \frac{s^2}{c_1} = 6 \cdot \frac{0.08}{1.24} = 0.387 \end{aligned}$$

Das Vertrauensintervall für die unbekannte Varianz  $\sigma^2$  der Grundgesamtheit lautet damit

$$k_1 = 0.033 \text{ l}^2 \leq \sigma^2 \leq 0.387 \text{ l}^2 = k_2,$$

und der wahre, aber hier unbekannte Wert der Varianz liegt mit einem Vertrauen (= der Wahrscheinlichkeit) von 95 % in diesem Intervall.

5. Für die Standardabweichung  $\sigma$  erhält man bei derselben Irrtumswahrscheinlichkeit  $\alpha = 0.05$  % das Vertrauensintervall

$$\sqrt{0.033} \text{ l} = 0.1821 \leq \sigma \leq 0.6221 = \sqrt{0.387} \text{ l}$$

### Aufgabe

Man berechne das Vertrauensintervall für die unbekannte Varianz, indem man dieses Mal  $n = 14$  Werte der Stichprobe nimmt, ansonsten aber dieselben Größen verwendet.

### Bemerkung

Will man ein Vertrauensintervall für die unbekannte Varianz  $\sigma^2$  angeben, so ist für gewöhnlich der Erwartungswert  $\mu = E(X)$  ebenfalls unbekannt und muß



auch geschätzt werden, so wie im letzten Beispiel durchgeführt. Manchmal ist er jedoch bekannt, und dann läßt sich diese Kenntnis natürlich ausnutzen.

Das geschieht dadurch, daß man anstelle der Schätzfunktion

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

die Schätzfunktion

$$\hat{S}^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

verwendet und damit die Zufallsvariable

$$\hat{\chi}^2 := n \cdot \frac{\hat{S}^2}{\sigma^2}$$

bildet, welche einer  $\chi^2$ -Verteilung mit  $n$  Freiheitsgraden genügt.

### 10.1.5 Toleranzintervalle

Konfidenzintervalle sind Intervalle, welche mit einer vorgegebenen Wahrscheinlichkeit  $\gamma$  einen Parameter einer unbekanntem Verteilung enthalten. Häufig ist man jedoch weniger an der Abschätzung von Parametern interessiert als vielmehr daran, Grenzen zu bestimmen, innerhalb derer einzelne Werte der Zufallsvariable fallen und die von (möglichst) wenigen Einzelwerten unter- oder überschritten werden.

#### Beispiel 10.8

Eine Zufallsstichprobe von Scherbolzen wird in bezug auf die Härte des verwendeten Materials hin untersucht. Der Käufer dieser Scherbolzen wird wahrscheinlich eine untere Grenze für die zulässige Härte vorgeben, unterhalb derer er überhaupt keine Bolzen kauft.

Hier sind sog. *Toleranzgrenzen* für die einzelnen Bolzen viel wichtiger als irgendwelche Mittelwerte.

Eine Methode, Toleranzgrenzen für Einzelwerte einer Grundgesamtheit anzugeben besteht darin, für einen festen Anteil dieser Grundgesamtheit – etwa 95 % – ein Konfidenzintervall anzugeben. Man betrachte beispielsweise eine Zufallsstichprobe einer Normalverteilung mit bekanntem Erwartungswert  $\mu$  und bekannter Varianz  $\sigma^2$ . Das Intervall, welches die mittleren (= um den Erwartungswert  $\mu$  gruppierten) 95 % der Werte der Grundgesamtheit enthält, ist dann gegeben durch

$$\mathcal{I}_{(0.95)} = [\mu - c \cdot \sigma, \mu + c \cdot \sigma]$$

mit  $c := \Phi^{-1}(1 - \frac{0.05}{2}) = 1.96$ , dem  $(1 - \frac{0.05}{2})$ -Quantil der Standardnormalverteilung.

Dieses Intervall heißt *Toleranzintervall*, und es werden exakt 95 % der Meßwerte von ihm überdeckt.

### Problem

In der Praxis sind jedoch  $\mu$  und/oder  $\sigma$  selten bekannt, und man muß die zugehörigen Schätzwerte  $\bar{x}$  (Stichprobenmittel) und  $s$  (empirische Standardabweichung) einer Stichprobe verwenden, um Toleranzgrenzen zu berechnen.

Das Intervall  $[\bar{x} - k \cdot s, \bar{x} + k \cdot s]$  ist nun selbst Wert einer Zufallsvariable, und die Überdeckung eines festen Anteils der Grundgesamtheit ist nicht mehr exakt. Daher muß auch hier mit dem Begriff des Konfidenzintervalles gearbeitet werden.

### Definition 10.7

Gegeben sei eine Normalverteilung mit unbekanntem Erwartungswert  $\mu$  und unbekannter Standardabweichung  $\sigma$ . Das Intervall  $[\bar{x} - k \cdot s, \bar{x} + k \cdot s]$  heißt *Toleranzintervall mit minimaler Trefferwahrscheinlichkeit  $p$  und Sicherheitswahrscheinlichkeit  $\gamma = 1 - \alpha$* , wenn  $k$  so bestimmt wird, daß gilt:

Mit einer statistischen Sicherheit von  $\gamma \cdot 100\%$  liegen mindestens  $p \cdot 100\%$  der zukünftigen Beobachtungswerte in diesem Intervall.

### Beispiel 10.9

Eine Maschine produziere zylindrische Metallstifte.  $n = 9$  dieser Metallstifte besitzen die Werte  $\bar{x} = 1.0056$  und  $s = 0.0245$ . Eine entsprechende Toleranzta-  
belle<sup>4</sup> liefert für  $n = 9, p = 0.99, \gamma = 0.95$  den Wert  $k = 4.550$ .

Damit sind die Toleranzgrenzen gegeben durch  $1.0056 \pm 4.550 \cdot 0.0245$ , und das 99 %-Toleranzintervall lautet

$$\mathcal{T}_{(0.99)} := [0.894, 1.117].$$

### Ergebnis

Mit einer statistischen Sicherheit von 99 % kann man annehmen, daß das Toleranzintervall  $\mathcal{T}_{(0.99)}$  95 % der von der Maschine produzierten Metallstifte enthält.

<sup>4</sup>C. Eisenhart, M.W.Hastay und W. A. Wallis *Techniques of Statistical Analysis*, Chapter 2, McGraw-Hill Book Company, New York 1947

R. E. Walpole, R. H. Myers *Probability and Statistics for Engineers and Scientists* Table A.7, Prentice Hall, Englewood Cliffs, 5<sup>th</sup> Edition

**Bemerkung**

Berechnet man mit den angegebenen Werten das 99%-Konfidenzintervall für den unbekanntem Erwartungswert  $\mu$ , so ergibt sich

$$\begin{aligned}\bar{x} \pm t_{n-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{s}{\sqrt{n}} &= 1.0056 \pm 3.355 \cdot \frac{0.0245}{\sqrt{9}} \\ &= 1.0056 \pm 0.0274 \\ &\stackrel{\wedge}{=} [0.978, 1.033].\end{aligned}$$

Zur selben statistischen Sicherheitswahrscheinlichkeit  $\gamma = 1 - \alpha$  ist das Toleranzintervall für die einzelnen Werte der Stichprobe immer größer als das Konfidenzintervall für den unbekanntem Erwartungswert. Beide Intervalle sind von der Form

$$\bar{x} \pm k \cdot s$$

mit

$$\begin{aligned}k_{\text{Toleranz}} &= 4.550 \\ k_{\text{Konfidenz}} &= \frac{t_8(0.995)}{\sqrt{9}} = 1.118\end{aligned}$$

**10.2 Statistische Prüfverfahren**

Gegenstand der induktiven Statistik ist das Schließen von Stichproben auf Grundgesamtheiten, die Grundgesamtheit repräsentiert dabei den zu untersuchenden Aspekt der Realität. Eine Grundgesamtheit wird vollständig beschrieben durch ihre Dichtefunktion  $f$  oder durch ihre Verteilungsfunktion  $F$ , welche ihrerseits durch bestimmte Parameter charakterisiert werden, wie z.B. Erwartungswert und Varianz bzw. Standardabweichung.

Im letzten Kapitel wurden Schätzverfahren für solche Parameter einer Verteilungsfunktion untersucht und für konkrete Werte von Schätzfunktionen – die Schätzwerte – Konfidenzintervalle angegeben.

In diesem Kapitel sollen Verfahren behandelt werden, die es gestatten, bestimmte *Hypothesen* zu überprüfen. Eine Hypothese ist dabei eine Aussage über die Verteilungsfunktion einer Grundgesamtheit oder deren Parameter. Solche Verfahren heißen *statistische Testverfahren* oder kurz *Tests*.

**10.2.1 Statistische Hypothesen und statistische Tests****Definition 10.8**

- (i) Eine *statistische Hypothese* ist eine Aussage über die Wahrscheinlichkeitsverteilung einer Zufallsvariable oder über eine Grundgesamtheit und deren Parameter.

- (ii) Ein *statistischer Test* ist ein statistisches Prüfverfahren zur Prüfung einer statistischen Hypothese.

Eine statistische Hypothese kann wahr oder falsch sein. Ihre Richtigkeit ist nicht mit Sicherheit bekannt, falls nicht die gesamte Grundgesamtheit untersucht wird. Anhand einer konkreten Stichprobe möchte man nun Entscheidungshilfen für das Für und Wider einer (statistischen) Hypothese bekommen. Eine solche Entscheidungshilfe läßt sich mit einem statistischen Test erzielen.

Mittels statistischer Tests lassen sich etwa die folgenden Hypothesen prüfen:

- (i) *Parametertests*  
Hypothesen über die (unbekannten) Parameter einer (als bekannt vorausgesetzten) Wahrscheinlichkeitsverteilung
- (ii) *Nichtparametrische Tests*  
Hypothesen über die Lokalisation (Lage) oder die Streuung einer Wahrscheinlichkeitsverteilung bei unbekanntem Verteilungstyp
- (iii) *Anpassungstests*  
Hypothesen über die Art einer Verteilung
- (iv) *Unabhängigkeitstests*  
Hypothesen über die Abhängigkeit von Merkmalen

Ein statistischer Test besteht aus einem Paar  $\langle H_0, H_1 \rangle$ , wobei gilt:

- $H_0$  : *Nullhypothese* oder kurz: Hypothese  
 $H_1$  : *Alternativhypothese* oder kurz: Alternative

Man betrachte eine adäquate *Prüfgröße* oder *Teststatistik*. Darunter versteht man eine Stichprobenfunktion

$$\langle X_1, \dots, X_n \rangle \longmapsto T := g(X_1, \dots, X_n),$$

die auf  $n$  unabhängig verteilten Zufallsvariablen  $X_1, \dots, X_n$  definiert ist, deren Verteilung unter Zugrundelegung der Hypothese  $H_0$  – d.h. bei Vorliegen von  $H_0$  – bekannt oder abschätzbar ist.

Die Prüfgröße wird anhand der Daten  $x_1, \dots, x_n$  der vorliegenden Stichprobe ausgewertet, und der erhaltene Schätzwert liefert dann eine Entscheidung zugunsten von  $H_0$  oder zugunsten von  $H_1$ .

### Bemerkung

Häufig werden die Begriffe

- Akzeptieren einer Hypothese      oder

- Verwerfen einer Hypothese

benutzt. Das Verwerfen einer Hypothese – etwa  $H_0$  – bedeutet den (statistischen) Schluß, daß sie falsch ist. Die aus der konkreten Stichprobe gewonnenen Daten liefern dann Anlaß zu der Annahme, daß  $H_1$  nicht nur zufallsbedingt richtig ist. Die Akzeptanz der Hypothese  $H_0$  hingegen bedeutet lediglich, daß keine Anhaltspunkte vorliegen, etwas anderes anzunehmen. Mehr nicht!  $H_0$  kann aufgrund der Datenlage oder einfach nur zufallsbedingt richtig sein.

Aus diesem Grunde geht man bei Tests in der Regel so vor, daß man die zu behauptende Hypothese als Alternativhypothese  $H_1$  formuliert und dann aus statistischen Überlegungen anhand der vorliegenden Stichprobe die Nullhypothese  $H_0$  ablehnt.

Absolut sichere Rückschlüsse von einer Stichprobe auf die Grundgesamtheit sind dennoch grundsätzlich nicht möglich. Es besteht immer die Möglichkeit eines Irrtums, d.h. einer falschen Entscheidung.

### 10.2.1.1 Parametertests

Ein *Parametertest* ist ein statistisches Prüfverfahren für eine Hypothese über einen unbekanntem Parameter in der Wahrscheinlichkeitsverteilung einer Zufallsvariable oder Grundgesamtheit, wobei die *Art* der Verteilung als bekannt vorausgesetzt wird.

#### Beispiel 10.10

- (i) Hypothese  $H_0$ :

Der Erwartungswert  $\mu$  der normalverteilten Zufallsvariable  $X$  liegt im Intervall  $[-2, 2]$ .

- (ii) Hypothese  $H_0$ :

Die Erwartungswerte  $\mu_X$  und  $\mu_Y$  der beiden normalverteilten Zufallsvariablen  $X$  und  $Y$  sind gleich.

#### Ziel eines Parametertests

Herbeiführung einer Entscheidung darüber, ob man die Nullhypothese  $H_0$  *annehmen* oder sie zugunsten der Alternativhypothese  $H_1$  *verwerfen* muß.

#### Beispiel 10.11

Im Zufallsexperiment „Wurf eines homogenen Würfels“ sei das Experiment  $A$  definiert durch

$A \iff$  der Wurf liefert eine 6.

$$\begin{array}{ll} \text{Nullhypothese } H_0 & : \quad p(A) = \frac{1}{6} \\ \text{Alternativhypothese } H_1 & : \quad p(A) \neq \frac{1}{6} \end{array}$$

Dieses Experiment ist ein sog. *zweiseitiger Parametertest*, da die Alternative  $H_1$  Parameterwerte nach beiden Seiten hin zuläßt:

$$H_1 : \quad p(A) < \frac{1}{6} \quad \text{oder} \quad p(A) > \frac{1}{6} .$$

### Standardsituation

Die Art der Verteilung einer Zufallsvariable  $X$  sei bekannt; etwa dadurch, daß man die Dichtefunktion  $f(\cdot; \theta)$  in Abhängigkeit von einem ein- oder mehrdimensionalen Parameter  $\theta$  kennt. Die zu testende Hypothese sei eine Aussage über  $\theta$ .

### Definition 10.9

- (i)  $\Theta$  sei die Menge aller in Frage kommenden Werte für  $\theta$ ;  $\Theta$  heißt *Parametermenge* oder *Parameterraum*. Man betrachtet dann eine disjunkte Zerlegung

$$\Theta = \Theta_0 \cup \Theta_1$$

des Parameterraumes und entscheidet anhand der Daten einer konkreten Stichprobe, ob

$$\theta \in \Theta_0 \quad \text{oder} \quad \theta \in \Theta_1$$

gilt.

- (ii) Jede Hypothese über  $\theta$  läßt sich auf die Form

$$H_0 : \quad \theta \in \Theta_0$$

bringen.  $H_0$  heißt *Nullhypothese*,  $\Theta_0$  heißt *Hypothesenmenge*.

- (iii) Jede zu  $\Theta_0$  disjunkte Teilmenge  $\Theta_1$  von  $\Theta$  kann zur Formulierung einer Alternative

$$H_1 : \quad \theta \in \Theta_1$$

dienen.  $\Theta_1$  heißt *Alternativenmenge*.

### Beispiel 10.12

Sei  $X$  eine normalverteilte Zufallsvariable mit unbekanntem Erwartungswert  $\mu = E(X)$  und bekannter Varianz  $\sigma^2 = V(X)$ . Man vermutet, daß der unbekannte Erwartungswert durch  $\mu_0$  gegeben ist und interessiert sich infolgedessen für die Irrtumswahrscheinlichkeit  $\alpha$  der Nullhypothese

$$H_0 : \quad \mu = \mu_0,$$

d.h. man untersucht die Nullhypothese  $H_0$  auf dem *Signifikanzniveau*  $\alpha$ .

Sei  $\langle x_1, \dots, x_n \rangle$  eine Stichprobe vom Umfang  $n$ . Dann gilt für die Wahrscheinlichkeit, daß der unbekannte Erwartungswert  $\mu$  im Konfidenzintervall

$$\left[ \bar{x} - c \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + c \cdot \frac{\sigma}{\sqrt{n}} \right]$$

liegt, die Beziehung

$$P\left(\left\{\mu \in \left[\bar{x} - c \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + c \cdot \frac{\sigma}{\sqrt{n}}\right]\right\}\right) = 2\Phi(c) - 1,$$

bzw. mit vertauschten Rollen von  $\bar{x}$  und  $\mu$  (warum ist das zulässig?):

$$P\left(\left\{\bar{x} \in \left[\mu - c \cdot \frac{\sigma}{\sqrt{n}}, \mu + c \cdot \frac{\sigma}{\sqrt{n}}\right]\right\}\right) = 2\Phi(c) - 1;$$

hierbei ist  $\Phi$  die Verteilungsfunktion der Standardnormalverteilung. Jetzt wählt man  $c$  als  $(1 - \frac{\alpha}{2})$ -Quantil dieser Verteilung:

$$\begin{aligned} 2\Phi(c) - 1 &= \gamma = 1 - \alpha \\ \longrightarrow \Phi(c) &= 1 - \frac{\alpha}{2}, \quad c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right). \end{aligned}$$

Wenn also für unsere Stichprobe die Bedingung

$$\bar{x} \in \left[\mu - c \cdot \frac{\sigma}{\sqrt{n}}, \mu + c \cdot \frac{\sigma}{\sqrt{n}}\right]$$

mit der Wahrscheinlichkeit  $\gamma = 1 - \alpha$  erfüllt ist, dann wird  $H_0$  akzeptiert – zumindest provisorisch, d.h. bei alleiniger Untersuchung der aktuellen Stichprobe; ist dies nicht erfüllt, so wird  $H_0$  abgelehnt.

### Konkrete Rechnung

$X$  sei normalverteilt mit  $\sigma^2 = V(X) = 100$ . Die Vermutung sei etwa  $\mu = E(X) = 800$ . Daraus resultiert die

$$\text{Nullhypothese} \quad H_0 : \mu = \mu_0 := 800.$$

Verwendet wird das

$$\text{Signifikanzniveau} \quad \alpha := 0.05 = 5\%.$$

$$\Phi(c) = 1 - \frac{\alpha}{2} = 0.975 \quad \longrightarrow \quad c = \Phi^{-1}(0.975) = 1.960.$$

Damit lautet die zu erfüllende Bedingung

$$\bar{x} \in \left[\mu - c \cdot \frac{\sigma}{\sqrt{n}}, \mu + c \cdot \frac{\sigma}{\sqrt{n}}\right] = \left[800 - \frac{19.6}{\sqrt{n}}, 800 + \frac{19.6}{\sqrt{n}}\right].$$

- (i) Bei einer Stichprobe vom Umfang  $n = 10$  und arithmetischem Mittelwert  $\bar{x} = 805$  lautet das Intervall

$$\mathcal{I}_1 := [800 - 6.20, 800 + 6.20] = [793.8, 806.2],$$

und wegen  $\bar{x} \in \mathcal{I}_1$  man hat keinen Grund, die Nullhypothese  $H_0$  aufgrund dieser Stichprobe zu verwerfen. Man sagt, die aktuelle Abweichung  $\bar{x} = 805$  vom vermuteten Wert  $\mu_0 = 800$  sei (auf diesem Niveau) nicht *signifikant*.

- (ii) Bei einer zweiten Untersuchung liege eine Stichprobe vom Umfang  $n = 100$  vor, es ergebe sich ebenfalls  $\bar{x} = 805$ . Das zu betrachtende Intervall ist dann hier gegeben durch

$$\mathcal{I}_2 := [800 - 1.96, 800 + 1.96] = [798.04, 801.96].$$

Nun gilt  $\bar{x} \notin \mathcal{I}_2$ , daher muß die Nullhypothese  $H_0$  jetzt abgelehnt werden: die Abweichung von 805 zu 800 ist auf diesem Niveau nun *signifikant*.

Natürlich kann man trotzdem nicht ausschließen, daß  $H_0$  dennoch zutrifft! Die momentane Datenlage spricht allerdings zu 95 % dagegen.

### Bemerkung

Beim Testen von Hypothesen schließt man von Stichproben auf Grundgesamtheiten, d.h. aufgrund der beobachteten Meßreihe  $x_1, \dots, x_n$  wird entschieden, ob die betreffende Hypothese  $H_0$  anzunehmen oder abzulehnen ist.

Ein Test  $T$  ordnet also einem  $n$ -Tupel  $\vec{x} := \langle x_1, \dots, x_n \rangle$  aus dem Stichprobenraum  $X$  einen der beiden Zahlenwerte 1 (Annahme) oder 0 (Ablehnung) zu und läßt sich damit als eine Abbildung des Stichprobenraumes  $X$  in die Menge  $\{0, 1\}$  auffassen:

$$T : X \longrightarrow \{0, 1\}.$$

Andererseits kann im Prinzip jede solche Abbildung als Test dienen. Es fragt sich nur, was solch ein Test leistet. Es werden einige Forderungen an die Abbildung  $T$  zu stellen sein, damit man auch „vernünftige“ Tests erhält.

### Definition 10.10

Sei  $T$  ein Test, also eine Abbildung

$$T : X \longrightarrow \{0, 1\}.$$

- (i) Die Menge

$$S := T^{-1}(1) = \{x : x \in X, T(x) = 1\}$$

heißt *Annahmereich* oder *nicht-kritischer Bereich* des Tests, auch *Akzeptanzbereich*.

- (ii) Die Menge

$$S_C := T^{-1}(0) = \{x : x \in X, T(x) = 0\}$$

heißt *Ablehnungsbereich* oder *kritischer Bereich* des Tests.

Ein Test  $T$  zerlegt also den Stichprobenraum disjunkt in einen Annahme- und einen Ablehnungsbereich.



	$H_0$	$H_1$
(i)	$\theta = \theta_0$	$\theta \neq \theta_0$
(ii)	$\theta \leq \theta_0$	$\theta > \theta_0$
(iii)	$\theta \geq \theta_0$	$\theta < \theta_0$

Tabelle 10.4: Beispiel eines Parametertests

### 10.2.1.2 Aufbau eines Parametertests

In Beispiel 10.12 des letzten Abschnittes klang an, wie ein Parametertest denn aufgebaut ist. In diesem Abschnitt soll der Aufbau eines solchen Tests systematisch beschrieben werden. Dessen Struktur besteht im wesentlichen aus den folgenden sechs Schritten:

(1.) Aufstellen von Null- und Alternativhypothese

Nullhypothese  $H_0$  und Alternativhypothese  $H_1$  werden anhand einer konkreten Fragestellung formuliert, s. etwa das Beispiel in Tabelle 10.4. Bei (i) spricht man von einem *zweiseitigen*-, bei (ii) oder (iii) von *einseitigen Tests*.

(2.) Vorgabe der Irrtumswahrscheinlichkeit  $\alpha$

Man wählt jetzt eine Zahl  $\alpha$ , genannt

- *Signifikanzzahl*,
- *Signifikanzniveau*      oder
- *Irrtumswahrscheinlichkeit*.

Sie ist die Wahrscheinlichkeit dafür, daß die Nullhypothese  $H_0$  abgelehnt wird, obwohl sie richtig ist (sog. *Fehler erster Art* oder  $\alpha$ -*Fehler*). In der Praxis wird oft  $\alpha = 0.05 = 5\%$  oder  $\alpha = 0.01 = 1\%$  gewählt.

(3.) Wahl einer geeigneten Prüfgröße  $T$

Dieser Schritt ist bei einem konkreten Test wohl der schwierigste. Bei der Prüfgröße  $T$  handelt es sich um eine dem konkreten Problem angepaßte Stichprobenfunktion, also um eine Abbildung

$$g : \begin{array}{ccc} \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ \langle X_1, \dots, X_n \rangle & \longmapsto & T := g(X_1, \dots, X_n). \end{array}$$

Die Zufallsvariablen  $X_1, \dots, X_n$  besitzen alle dieselbe Verteilung wie  $X$ , sie sind „Kopien“ von  $X$ . Die Prüfgröße  $T$  heißt auch *Prüfstatistik*; es wird hier vorausgesetzt, daß ihre Verteilung bekannt ist.

(4.) Bestimmung des kritischen Bereiches  $\mathcal{K}$ 

Auf der Basis der gewählten Signifikanzzahl  $\alpha$  werden zwei kritische Grenzen  $c_u$  und  $c_o$  derart bestimmt, daß die Prüfgröße  $T$  mit der Wahrscheinlichkeit  $\gamma = 1 - \alpha$  Werte aus dem Intervall  $c_u \leq T \leq c_o$  annimmt (Abb. 10.6):

$$P(c_u \leq T \leq c_o)_{H_0} = \gamma = 1 - \alpha$$

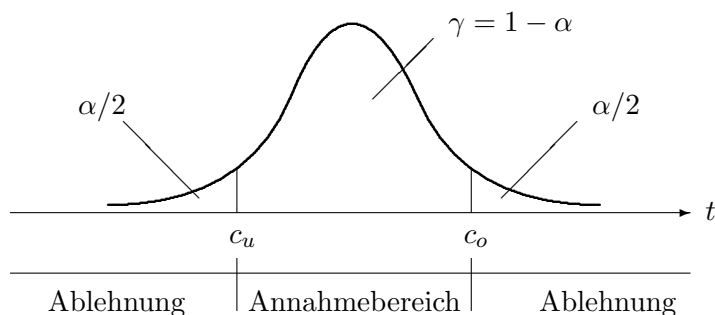


Abbildung 10.6: Anahme- und Ablehnungsbereich bei einer normalverteilten Zufallsvariable

Die Berechnung der kritischen Grenzen erfolgt dabei unter der Voraussetzung, daß die Nullhypothese  $H_0$  tatsächlich zutrifft; aus diesem Grunde ist das Symbol  $P$  mit dem Index „ $H_0$ “ versehen.

(5.) Berechnung einer Realisierung der Prüfgröße  $T$ 

Jetzt wird der Wert  $t$  der Testvariable  $T$  aus einer vorgegebenen Stichprobe  $x_1, \dots, x_n$  vom Umfang  $n$  berechnet. Das geschieht dadurch, daß diese Werte der Reihe nach für die unabhängigen Zufallsvariablen  $X_1, \dots, X_n$  in  $T$  eingesetzt werden:

$$\begin{aligned} T &= g(X_1, \dots, X_n) && \text{Prüfvariable} \\ t &:= g(x_1, \dots, x_n) && \text{Prüfwert} \end{aligned}$$

## (6.) Testentscheidung

Danach wird die Entscheidung über die Nullhypothese  $H_0$  einfach wie folgt vorgenommen:

$$\begin{aligned} t \in [c_u, c_o] &\implies H_0 \text{ wird nicht abgelehnt} \\ t \notin [c_u, c_o] &\implies H_0 \text{ wird abgelehnt} \end{aligned}$$

Wird  $H_0$  abgelehnt, so sagt man, der Schätzwert  $\hat{\theta}$  des unbekanntes Parameters  $\theta$  weiche *signifikant* vom vermuteten Wert  $\theta_0$  ab (etwa:  $\hat{\theta} = \bar{x} \neq \theta_0$ ).

**Bemerkung**

Der Begriff „signifikant“ meint hier also etwas Anderes als in der Umgangssprache gebräuchlich. Dort hat sich der Begriff signifikant als Synonym für alles und jedes eingebürgert, was lediglich einer Hervorhebung bedarf, ein anderes Wort für wichtig sozusagen.

In der Statistik dagegen bezeichnet man einen Sachverhalt dann als signifikant, wenn die vorhandene Datenlage darauf hindeutet, daß man die gemachte Hypothese mit großer Wahrscheinlichkeit nicht dem Zufall in die Schuhe schieben kann. Beträgt die Wahrscheinlichkeit dafür mehr als 95 %, so bezeichnet man diesen Sachverhalt als *signifikant* schlechthin, beträgt sie sogar mehr als 99 %, so bezeichnet man ihn als *hochsignifikant*.

Nachdem in Beispiel 10.12 des letzten Abschnittes eine eine sog. *einfache Hypothese*:  $\mu = \mu_0$  behandelt wurde, soll jetzt ein Beispiel zu einer sog. *zusammengesetzten Hypothese*:  $\mu \geq \mu_0$  untersucht werden.

**Beispiel 10.13**

Die Druckfestigkeit einer Betonsorte sei normalverteilt mit einer Standardabweichung von  $\sigma = 2.6$  [MPa] (Megapascal). Aus einer konkreten Stichprobe vom Umfang  $n = 10$  wird für die Druckfestigkeit das arithmetische Mittel  $\bar{x} = 26.23$  [MPa] ermittelt.

**Frage**

Entstammt diese Stichprobe einer (normalverteilten) Grundgesamtheit mit dem Sollwert  $\mu_0 = 28$  [MPa]?

Im vorliegenden Fall ist eine Unterschreitung des Sollwertes kritisch. Deshalb wird die Hypothese

$$H_0 : E(X) = 28$$

für den Fall einer einseitigen Fragestellung geprüft. Dabei orientiere ich mich an dem vor diesem Beispiel aufgestellten Procedere.

**Lösung**

(1.) Aufstellen von Null- und Alternativhypothese

$$\begin{array}{ll} H_0 : E(X) \geq \mu_0 = 28 & \text{Nullhypothese} \\ H_1 : E(X) < \mu_0 = 28 & \text{Alternativhypothese} \end{array}$$

(2.) Irrtumswahrscheinlichkeit  $\alpha = 0.05 = 5\%$

(3.) Wahl einer geeigneten Prüfgröße  $T$

Zu einer Stichprobe  $\langle x_1, \dots, x_n \rangle$  vom Umfang  $n$  aus der  $N(\mu, \sigma^2)$ -verteilten Grundgesamtheit wird eine Prüfgröße gewählt, deren Verteilung man kennt. Welche?

Da  $X$  eine  $N(\mu, \sigma^2)$ -Verteilung besitzt, besitzt die Zufallsvariable

$$Z := \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \sqrt{n} \cdot \frac{\bar{X} - \mu_0}{\sigma}$$

eine  $N(0, 1)$ -Verteilung. Beachte: Diese Aussage ist nur dann richtig, wenn  $\mu_0 = E(X) = E(\bar{X})$  gilt. Das jedoch ist gerade die Nullhypothese.

(4.) Bestimmung des kritischen Bereiches  $\mathcal{K}$

Bei einer einseitigen Fragestellung wird der kritische Bereich  $\mathcal{K}$  aus der Bedingung

$$P\left(T = \sqrt{n} \cdot \frac{\bar{X} - \mu_0}{\sigma} < -z_\alpha\right)_{H_0} = \alpha$$

$$\iff P\left(\bar{X} < \mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right) = \alpha$$

ermittelt.

(5.) Berechnung einer Realisierung der Prüfgröße  $T$

Anhand der konkret vorliegenden Stichprobe wird eine Realisierung der Prüfgröße berechnet:

$$t = \sqrt{n} \cdot \frac{\bar{x} - \mu_0}{\sigma} = \sqrt{10} \cdot \frac{26.23 - 28}{2.6} = -2.15.$$

(6.) Testentscheidung

Aus der Tafel für die Standardnormalverteilung ermittelt man für das Quantil  $u_\alpha$

$$z_\alpha = z_{0.05} = 1.645,$$

und der kritische Bereich  $\mathcal{K}$  ist gegeben durch

$$\mathcal{K} = S_C = (-\infty, -1.645).$$

Da  $t = -2.15 < -1.645 = z_\alpha$  ist, wird die Nullhypothese  $H_0$  abgelehnt, d.h. die (in der vorliegenden Stichprobe realisierte) Abweichung der ermittelten Druckfestigkeit vom Sollwert ist signifikant.

## 10.2.1.3 Beispiele von Parametertests

**Einstichproben-Gauß-Test**

1. Verteilungsannahme :

$X_1, \dots, X_n$  seien unabhängig, identisch  $N(\mu, \sigma_0^2)$ -verteilt,  
 $\sigma_0^2$  sei bekannt.

2. Nullhypothese :
- $H_0 : \mu = \mu_0$
- .

3. Testgröße :
- $T(X_1, \dots, X_n) = \frac{\sqrt{n}}{\sigma_0} \cdot (\bar{X} - \mu_0)$
- .

4. Kritischer Bereich :
- $\mathcal{K} = \{\vec{x} : \vec{x} \in \mathbb{R}^n, |T(\vec{x})| > u_{1-\alpha/2}\}$
- .

5. Entscheidung : Wird ein
- $\vec{x}$
- mit

$$\frac{\sqrt{n}}{\sigma_0} \cdot |\bar{x} - \mu_0| > u_{1-\alpha/2}$$

beobachtet, so wird  $H_0$  abgelehnt; ansonsten besteht  
kein Grund,  $H_0$  abzulehnen.

Beispiele 10.12 und 10.13 sind Beispiele für einseitige bzw. zweiseitige sog. *Einstichproben-Gauß-Tests*. Hierbei bezieht sich die Nullhypothese auf den unbekanntem Erwartungswert einer Normalverteilung, über den anhand einer Stichprobe eine statistische Aussage getroffen wird.

Oft hat eine Hypothese einen Vergleich zum Inhalt. Beispielsweise will man die Erwartungserte  $\mu_X$  und  $\mu_Y$  zweier normalverteilter (unabhängiger) Zufallsvariablen  $X$  und  $Y$  vergleichen. Das kann aufgrund zweier Meßreihen  $x_1, \dots, x_m$  und  $y_1, \dots, y_n$  geschehen, zu denen die Stichprobenvariablen  $X_1, \dots, X_m$  und  $Y_1, \dots, Y_n$  gehören. Sind die Varianzen  $\sigma_X^2$  von  $X$  und  $\sigma_Y^2$  von  $Y$  bekannt, so kann man die Testgröße

$$T(X_1, \dots, X_m, Y_1, \dots, Y_n) := \frac{1}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \cdot (\bar{Y} - \bar{X}) = \sqrt{\frac{mn}{n\sigma_1^2 + m\sigma_2^2}} \cdot (\bar{Y} - \bar{X})$$

verwenden. Diese besitzt<sup>5</sup> im Falle  $\mu_X = \mu_Y$  eine Standardnormalverteilung, und man kann einen *Zweistichproben-Gauß-Test* formulieren:

<sup>5</sup>Sind  $X_1, \dots, X_n$  unabhängige normalverteilte Zufallsvariablen mit Erwartungswerten  $\mu_1, \dots, \mu_n$  und Varianzen  $\sigma_1^2, \dots, \sigma_n^2$ , dann ist die Zufallsvariable

$$Y := a_1 X_1 + \dots + a_n X_n$$

normalverteilt mit dem Erwartungswert

$$\mu_Y := a_1 \mu_1 + \dots + a_n \mu_n$$

und der Varianz

$$\sigma_Y^2 := a_1^2 \sigma_1^2 + \dots + a_n^2 \sigma_n^2.$$

**Zweistichproben-Gauß-Test**

1. Verteilungsannahme :

$X_1, \dots, X_m$  seien identisch  $N(\mu_1, \sigma_1^2)$ -verteilt,

$Y_1, \dots, Y_n$  seien identisch  $N(\mu_2, \sigma_2^2)$ -verteilt,

alle Variable seien unabhängig,

$\mu_1, \mu_2$  seien unbekannt,  $\sigma_1^2, \sigma_2^2$  seien bekannt.

2. Nullhypothese :
- $H_0 : \mu_1 = \mu_2$
- .

3. Testgröße :
- $T(X_1, \dots, X_m, Y_1, \dots, Y_n) = \frac{\sqrt{m n}}{n\sigma_1^2 + m\sigma_2^2} \cdot (\bar{Y} - \bar{X})$
- .

4. Kritischer Bereich :

$\mathcal{K} = \{ \langle \vec{x}, \vec{y} \rangle : \langle \vec{x}, \vec{y} \rangle \in \mathbb{R}^{m+n}, |T(\vec{x}, \vec{y})| > u_{1-\alpha/2} \}$ .

5. Entscheidung : Werden Stichprobenmittelwerte
- $\bar{x}$
- und
- $\bar{y}$
- mit

$$\frac{1}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \cdot |\bar{y} - \bar{x}| = \sqrt{\frac{m n}{n\sigma_1^2 + m\sigma_2^2}} \cdot |\bar{y} - \bar{x}| > u_{1-\alpha/2}$$

beobachtet, so wird  $H_0$  abgelehnt; ansonsten besteht

kein Grund,  $H_0$  abzulehnen.

**Beispiel 10.14**

An zwei Fertigungsstraßen werden Widerstände hergestellt. Geprüft werden soll, ob die an jeder der beiden Fertigungsstraßen produzierten Widerstände im Mittel den gleichen Widerstandswert besitzen. Bekannt sei, daß beide Fertigungsstraßen mit der Varianz  $\sigma_X^2 = \sigma_Y^2 = 2.25$  produzieren.

Die Nullhypothese

$$H_0 : E(X) = E(Y)$$

soll für die Irrtumswahrscheinlichkeit  $\alpha = 0.05$  und eine zweiseitige Fragestellung geprüft werden. Zur Prüfung werden den Grundgesamtheiten die konkreten Stichproben  $x_1, \dots, x_{15}$  bzw.  $y_1, \dots, y_{12}$  vom Umfang  $m = 15$  bzw.  $n = 12$  entnommen. Aus ihnen werden die beiden empirischen Mittelwerte

$$\bar{x} = 152.5 \quad \text{und} \quad \bar{y} = 151.4$$

entnommen und damit eine Realisierung der Prüfgröße ermittelt:

$$u := \sqrt{\frac{15 \cdot 12}{12 \cdot 2.25 + 15 \cdot 2.25}} \cdot |151.4 - 152.5| = 1.893.$$

Das  $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung besitzt für die Irrtumswahrscheinlichkeit  $\alpha = 0.05$  den kritischen Wert  $u_{1-\alpha/2} = 1.960$ . Da  $u = 1.893 > 1.960$  ist, wird die Nullhypothese  $H_0$  abgelehnt.

Beim Gauß-Test wird die Kenntnis der Varianz vorausgesetzt. Ist diese Kenntnis nicht vorhanden, so muß man auch dafür einen Schätzwert verwenden, und zwar den Schätzwert

$$s^2 := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

welcher mit Hilfe der zugehörigen erwartungstreuen Schätzfunktion

$$S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

aus einer Stichprobe vom Umfang  $n$  ermittelt wird. Mit ihrer Hilfe wird die Stichprobenvariable

$$T(X_1, \dots, X_n) := \sqrt{\frac{n}{S^2}} \cdot (\bar{X} - \mu_0)$$

gebildet, welche nach Abschnitt 7.1.2 eine  $t$ -Verteilung mit  $n-1$  Freiheitsgraden besitzt. Damit verläuft der  $t$ -Test analog zum Gauß-Test, nur daß dieses Mal das  $(1 - \alpha/2)$ -Quantil  $t_{n-1;1-\alpha/2}$  verwendet wird.

### **$t$ -Test**

1. Verteilungsannahme :

$X_1, \dots, X_n$  seien unabhängig, identisch  $N(\mu, \sigma^2)$ -verteilt,  
 $\mu$  und  $\sigma^2$  seien unbekannt.

2. Nullhypothese :  $H_0 : \mu = \mu_0$ .

3. Testgröße :  $T(X_1, \dots, X_n) = \sqrt{\frac{n}{S^2}} \cdot (\bar{X} - \mu_0)$ .

4. Kritischer Bereich :  $\mathcal{K} = \{\vec{x} : \vec{x} \in \mathbb{R}^n, |T(\vec{x})| > t_{n-1;1-\alpha/2}\}$ .

5. Entscheidung: Werden  $\bar{x}$  und  $s^2$  mit

$$\sqrt{\frac{n}{s^2}} \cdot |\bar{x} - \mu_0| > t_{n-1;1-\alpha/2}$$

beobachtet, so wird  $H_0$  abgelehnt; ansonsten besteht kein Grund,  $H_0$  abzulehnen.

### **Beispiel 10.15**

In Beispiel 10.12 sei etwa die Varianz nun nicht bekannt und werde anhand einer Stichprobe vom Umfang  $n = 20$  zu  $s^2 = 103$  geschätzt. Auf welchem Niveau beginnt die Abweichung  $\bar{x} = 805$  vom vermuteten Wert  $\mu_0 = 800$  signifikant zu werden?

**Lösung**

Es ist dasjenige Niveau  $\alpha$  zu bestimmen, für welches

$$\sqrt{\frac{n}{s^2}} \cdot |\bar{x} - \mu_0| = t_{n-1;1-\alpha/2}$$

ist, also

$$t_{19;1-\alpha/2} = \sqrt{\frac{20}{103}} \cdot |805 - 800| = 2.203.$$

Aus der  $t$ -Tabelle liest man für  $n = 19$  Freiheitsgrade den Wert  $\alpha = 0.02 = 2\%$  ab. Auf diesem Niveau beginnt die Abweichung  $\bar{x} = 805$  vom vermuteten Wert  $\mu_0 = 800$  signifikant zu werden.

**Bemerkung**

Auch beim  $t$ -Test gibt es wie beim Gauß-Test eine Fassung zum Vergleich zweier Normalverteilungen, allerdings unter der Einschränkung, daß die beiden Varianzen  $\sigma_X^2$  und  $\sigma_Y^2$  identisch sein müssen. Als Testgröße verwendet man dabei die Stichprobenvariable

$$T(X_1, \dots, X_m, Y_1, \dots, Y_n) = \frac{\sqrt{n+m-2}}{\sqrt{\frac{1}{m} + \frac{1}{n}}} \cdot \frac{\bar{Y} - \bar{X}}{\sqrt{(m-1) \cdot S_X^2 + (n-1) \cdot S_Y^2}},$$

von der gezeigt werden kann, daß sie einer  $t$ -Verteilung mit  $m+n-2$  Freiheitsgraden genügt; dabei sind  $S_X^2$  und  $S_Y^2$  die Stichprobenvarianzen von  $X$  und  $Y$ . Die Vorgehensweise beim Test ist analog zum geschilderten Fall bei einer Stichprobe (für  $X$ ).

**Beispiel 10.16**

Mit diesem Beispiel soll an Beispiel 10.14 angeknüpft werden; hier sollen allerdings die (als gleich vorausgesetzten) Varianzen  $\sigma_X^2$  und  $\sigma_Y^2$  unbekannt sein und müssen daher empirisch geschätzt werden:

$$\begin{array}{lll} \text{Empirische Mittelwerte} & : & \bar{x} = 152.5 \quad \text{und} \quad \bar{y} = 151.4 \\ \text{Empirische Varianzen} & : & s_x^2 = 2.56 \quad \text{und} \quad s_y^2 = 1.44 \end{array}$$

Die Realisierung der Prüfgröße lautet hier

$$u = \frac{\sqrt{12+15-2}}{\sqrt{\frac{1}{15} + \frac{1}{12}}} \cdot \frac{|151.4 - 152.5|}{\sqrt{(15-1) \cdot 2.56 + (12-1) \cdot 1.44}} = 1.975.$$

Das  $(1 - \alpha/2)$ -Quantil der  $t$ -Verteilung mit  $15 + 12 - 2 = 25$  Freiheitsgraden besitzt für die Irrtumswahrscheinlichkeit  $\alpha = 0.05$  den kritischen Wert  $t_{0.975;25} = 2.059$ . Da  $u = 1.975 < 2.059$  ist, kann man dieses Mal die Nullhypothese  $H_0$  auf dem 5%-Niveau nicht ablehnen.



$\chi^2$ -Varianztest

1. Verteilungsannahme :  
 $X_1, \dots, X_n$  seien unabhängig und identisch  $N(\mu, \sigma^2)$ -verteilt,  
 $\mu$  und  $\sigma^2$  seien unbekannt.
2. Nullhypothese :  $H_0 : \sigma^2 = \sigma_0^2$ .
3. Testgröße :  $T(X_1, \dots, X_n) = \frac{n-1}{\sigma_0^2} \cdot S^2$ .
4. Kritischer Bereich :  
 $\mathcal{K} = \{\vec{x} : \vec{x} \in \mathbb{R}^n, T(\vec{x}) < \chi_{n-1; \alpha/2}^2 \vee T(\vec{x}) > \chi_{n-1; 1-\alpha/2}^2\}$
5. Entscheidung: Wird ein Wert  $s^2$  mit  
 $\frac{n-1}{\sigma_0^2} \cdot s^2 < \chi_{n-1; \alpha/2}^2$  oder  $\frac{n-1}{\sigma_0^2} \cdot s^2 > \chi_{n-1; 1-\alpha/2}^2$   
beobachtet, so wird  $H_0$  abgelehnt; ansonsten besteht  
kein Grund,  $H_0$  abzulehnen.

**Beispiel 10.17**

Ein empfindliches Meßgerät eines medizinischen Labors, von dem man weiß, daß sein Meßfehler einer  $N(\mu = 0, \sigma^2 = 0.1)$ -Verteilung genügt, ist versehentlich von einem Mitarbeiter beschädigt worden. Es wird daraufhin repariert und neu geeicht. Durch  $n = 30$  Probemessungen soll geprüft werden, ob nach der Reparatur wieder die ursprüngliche Meßgenauigkeit vorliegt; dabei ergibt sich  $s^2 = 0.13$ . Deutet dieses Ergebnis auf einen signifikanten Unterschied zur Genauigkeit vor der Reparatur hin?

**Lösung**

Es wird gesetzt

$$H_0 : \sigma^2 = \sigma_0^2 \quad , \quad H_1 : \sigma^2 \neq \sigma_0^2 \quad , \quad n = 30 \quad , \quad \alpha = 0.05.$$

Aus der Tabelle für die Quantile der  $\chi^2$ -Verteilung entnimmt man die beiden Werte

$$\chi_{29; 0.025}^2 = 16.047 \quad \text{und} \quad \chi_{29; 0.975}^2 = 45.722.$$

Die Testgröße  $T(X_1, \dots, X_n) = \frac{n-1}{\sigma_0^2} \cdot S^2$  erhält den Wert

$$t = \frac{30-1}{0.1} \cdot 0.13 = 37.7,$$

und es besteht kein Grund,  $H_0$  abzulehnen, d.h. der durch den Mitarbeiter verursachte Schaden scheint behoben zu sein.

**Bemerkung**

Bei der Beschreibung der Testverfahren werden manchmal sog. *einfache Nullhypothesen* verwendet, etwa  $\mu = \mu_0$ . Die Alternative ist die Negation der Nullhypothese. Wie im letzten Beispiel gezeigt, kann man auch Kombinationen davon behandeln. In dem vorstehenden Varianztest beispielsweise wird die Qualität eines Meßgerätes durch die Varianz seines Meßfehlers beurteilt. Da das Gerät umso besser erscheint, je kleiner diese Varianz ist, kann diese also gar nicht klein genug sein, d.h. eine „zu kleine“ Varianz macht hier keinen Sinn. Man hätte daher anstelle von

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{und} \quad H_1 : \sigma^2 \neq \sigma_0^2$$

besser die Hypothesen

$$H_0 : \sigma^2 \leq \sigma_0^2 \quad \text{und} \quad H_1 : \sigma^2 > \sigma_0^2$$

verwenden sollen.

**Aufgabe**

Man formuliere den Varianztest aus Beispiel 10.17 als zweiseitigen Test.

Die Behandlung zusammengesetzter Hypothesen – z.B. Nullhypothese  $\mu \leq \mu_0$  gegen Alternativhypothese  $\mu > \mu_0$  – unterscheidet sich von den Verfahren zur Behandlung einseitiger Hypothesen jedoch nur durch die Wahl des kritischen Bereiches, die jeweiligen Testgrößen bleiben in analogen Situationen gleich, und auch der Testablauf ist derselbe. Der in Beispiel 10.13 durchgeführte Test ist ein Beispiel für einen zweiseitigen Test.

### 10.2.2 Äquivalenz von statistischen Schätzverfahren und statistischen Prüfverfahren

In der induktiven Statistik wurden bisher zwei Arten von Problemen besprochen:

- (i) Berechnung von Konfidenzintervallen für unbekannte Parameter einer Wahrscheinlichkeitsverteilung
- (ii) Testen von Hypothesen

Beide Methoden hängen miteinander zusammen, ja sie sind sogar zueinander äquivalent, denn:

Der Test

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

auf dem Signifikanzniveau  $\alpha$  besteht darin, für  $\mu$  ein  $100(1 - \alpha)$ -Konfidenzintervall  $\mathcal{K}$  zu berechnen und dann

$$\begin{aligned} H_0 \text{ abzulehnen} &\iff \mu_0 \notin \mathcal{K}, \\ H_0 \text{ nicht abzulehnen} &\iff \mu_0 \in \mathcal{K}. \end{aligned}$$

Umgekehrt definieren Konfidenzintervalle für einen reellen Parameter  $\theta$  einen Test auf dem Signifikanzniveau  $\alpha$ .

### Satz 10.2

Seien  $\alpha \in (0, 1)$  und

$$\mathcal{K} := [c_u, c_o] = [g_u(x), g_o(x)]$$

das  $100(1 - \alpha)$ -Konfidenzintervall für einen reellen Parameter  $\theta$  von  $X$ .

Dann definiert das Konfidenzintervall  $\mathcal{K}$  einen Test  $T$  für die Hypothese  $H_0 : \theta = \theta_0$  mit dem Annahmebereich

$$S := \{\vec{x} : \vec{x} \in \mathbb{R}^n, \theta_0 \in \mathcal{K}\}$$

und dem Ablehnungsbereich

$$S_C := \{\vec{x} : \vec{x} \in \mathbb{R}^n, \theta_0 \notin \mathcal{K}\}.$$

*Beweis*

Sei  $S := \{\vec{x} : \vec{x} \in \mathbb{R}^n, \theta_0 \in \mathcal{K}\}$  und definiere

$$\begin{aligned} T : \quad X &\longrightarrow \{0, 1\} \\ \vec{x} &\longmapsto T(x_1, \dots, x_n) =: \begin{cases} 0 & , \vec{x} \notin S \\ 1 & , \vec{x} \in S \end{cases}. \end{aligned}$$

Dann gilt für alle Stichproben vom Umfang  $n$ :

$$\begin{aligned} \iff \mathcal{K} \text{ ist Konfidenzintervall für } \theta \\ \iff \theta_0 \in \mathcal{K} \\ \iff \vec{x} \in S \\ \iff T(\vec{x}) = 1 \\ \iff H_0 : \theta = \theta_0 \end{aligned}$$

sowie die Alternative

$$\begin{aligned} \iff \theta \text{ liegt außerhalb des Konfidenzintervalles} \\ \iff \theta_0 \notin \mathcal{K} \\ \iff \vec{x} \notin S \\ \iff T(\vec{x}) = 0 \\ \iff H_1 : \theta \neq \theta_0 \end{aligned}$$

**Beispiel 10.18**

$\bar{x}$  sei empirischer Mittelwert einer Stichprobe, welche einer Normalverteilung mit der bekannten Varianz  $V(X) = \sigma^2$  entstammt. Betrachte den Test

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu \neq \mu_0 \end{aligned}$$

Dann gilt:

$$\begin{aligned} \Leftrightarrow & H_0 \text{ wird auf dem Niveau } \alpha \text{ nicht abgelehnt} \\ \Leftrightarrow & -u_{\alpha/2} \leq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \leq u_{\alpha/2} \\ \Leftrightarrow & \bar{x} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \\ \Leftrightarrow & \mu_0 \in \mathcal{K} := \left[ \bar{x} - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right] \end{aligned}$$

**Bemerkung**

Die hier dargestellte Äquivalenz von Konfidenzintervallen und Hypothesentests gilt auch für

- Tests bzgl. unbekannter Varianzen,
- Tests bzgl. der Differenz zweier Erwartungswerte,
- Tests bzgl. des Verhältnisses zweier Varianzen,
- etc.,

jedesmal natürlich mit anderen Testverteilungen.

Es gibt allerdings auch Tests, welche sich nicht durch ein Konfidenz-Schätzverfahren erledigen lassen. Dazu gehören alle nichtparametrischen Tests, also etwa

- Anpassungstests,
- Verteilungstests,
- Unabhängigkeitstests.

	$H_0$ ist wahr	$H_0$ ist falsch
$H_0$ wird akzeptiert	Korrekte Entscheidung	Fehler 2. Art
$H_0$ wird verworfen	Fehler 1. Art	Korrekte Entscheidung

Tabelle 10.5: Fehler erster und zweiter Art

### 10.2.3 Fehler bei statistischen Prüfverfahren

Am Ende eines Parametertests ist stets eine Entscheidung zu fällen, welche entweder zugunsten der Nullhypothese  $H_0$  oder zugunsten der Alternative  $H_1$  ausfallen kann. In beiden Fällen werden Rückschlüsse von einer Zufallsstichprobe auf die entsprechende Grundgesamtheit gezogen.

Bei einer Testentscheidung besteht somit immer die Möglichkeit eines Irrtums, d.h. einer Fehlentscheidung. M.a.W., es besteht eine bestimmte Wahrscheinlichkeit dafür, daß die getroffene Entscheidung falsch ist.

#### Definition 10.11

- (i) Ein *Fehler erster Art* oder  $\alpha$ -*Fehler* liegt vor, wenn eine Nullhypothese abgelehnt wird, obwohl sie wahr ist. Die Wahrscheinlichkeit eines Fehlers erster Art wird mit  $\alpha$  bezeichnet.
- (ii) Ein *Fehler zweiter Art* oder  $\beta$ -*Fehler* liegt vor, wenn eine Nullhypothese nicht abgelehnt wird, obwohl sie falsch ist. Die Wahrscheinlichkeit eines Fehlers zweiter Art wird mit  $\beta$  bezeichnet.

Beim Prüfen einer statistischen Hypothese anhand einer Stichprobe können somit die in Tabelle 10.5 beschriebenen Fälle eintreten.

#### 10.2.3.1 Fehler erster Art

##### Beispiel 10.19

In einer Grundgesamtheit von Kondensatoren sollen alle Elemente dieselbe Kapazität aufweisen, nämlich  $C := 68 \mu\text{F}$ . Ihr wird eine Stichprobe von  $n = 36$  Kondensatoren entnommen, um diese Hypothese zu testen.

Zum Nachweis dieser Hypothese wird ein zweiseitiger Parameter-test vorgenommen.

(1.) Nullhypothese, Alternativhypothese

$$\begin{aligned} H_0 &: \quad \mu = \mu_0 := 68 \mu\text{F} \\ H_1 &: \quad \mu \neq \mu_0 = 68 \mu\text{F} \end{aligned}$$

Die Alternativhypothese läßt Parameterwerte nach beiden Seiten hin zu, daher handelt es sich um einen zweiseitigen Parameterstest:

$$\mu \neq 68 \mu\text{F} \quad \iff \quad \mu < 68 \mu\text{F} \quad \vee \quad \mu > 68 \mu\text{F}.$$

(2.) Signifikanzniveau  $\alpha$  (zunächst unbestimmt)

(3.) Wahl einer geeigneten Prüfgröße

Die Standardabweichung dieser (nicht notwendig normalverteilten) Verteilung sei bekannt:

$$\sigma := 3.6 \mu\text{F}.$$

Als Schätzfunktion für den unbekanntem Erwartungswert  $\mu$ , also als gesuchte Prüfgröße, wird das Stichprobenmittel als die am besten geeignete Schätzfunktion genommen:

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

Nach dem zentralen Grenzwertsatz ist  $\bar{X}$  approximativ normalverteilt<sup>6</sup>. Hier gilt also

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3.6}{\sqrt{36}} = 0.6.$$

(4.) Bestimmung des kritischen Bereiches  $\mathcal{K}$  (Fehler erster Art)

Ein Stichprobenwert, welcher nahe am vermuteten Wert  $\mu_0 = 68 \mu\text{F}$  liegt, wird als eine Bestätigung der Nullhypothese  $H_0$  angesehen. Wie nahe ist nun „nahe“?

Dazu wählt man sich eine Vertrauenswahrscheinlichkeit  $\gamma$  und bestimmt dazu das Vertrauensintervall  $[c_u, c_o]$  mit der Eigenschaft

$$P(c_u \leq \bar{X} \leq c_o) = \gamma = 1 - \alpha;$$

die Berechnung des Vertrauensintervalles erfolgt dabei stets unter der Voraussetzung, daß die Nullhypothese auch zutrifft; im weiteren Verlauf dieser Betrachtung darf also die Bedingung  $\mu = \mu_0$  benutzt werden.

<sup>6</sup>Dieses bemerkenswerte Resultat gilt auch dann, wenn die  $X_i$  nicht normalverteilt sind. Die Approximation von  $\bar{X}$  durch eine Normalverteilung ist i.a. gut für  $n \geq 30$ . Ist  $n < 30$ , so ist die Approximation nur dann gut, wenn die gemeinsame Verteilung der  $X_i$  nicht zu stark von einer Normalverteilung abweicht.

Ist die gemeinsame Verteilung der  $X_i$  jedoch eine Normalverteilung, so ist die Verteilung von  $\bar{X}$  bekanntlich exakt normal, unabhängig von  $n$ .

Das Signifikanzniveau unseres Tests, also die Wahrscheinlichkeit, die richtige Nullhypothese  $H_0$  abzulehnen (Fehler erster Art), beträgt (Abb. 10.7):

$$\alpha = P(\bar{X} < c_u | H_0 \text{ ist wahr}) + P(\bar{X} > c_o | H_0 \text{ ist wahr}).$$

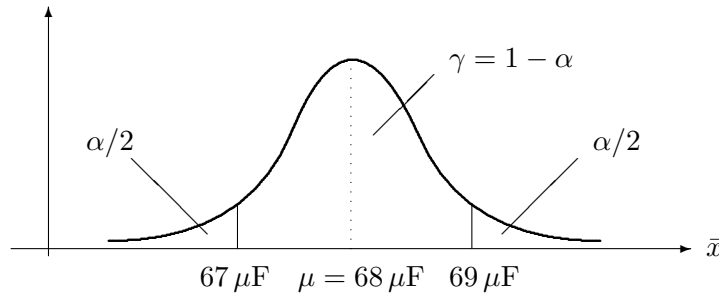


Abbildung 10.7: Kritischer Bereich für den Test  $\mu = 68 \mu\text{F}$  gegen  $\mu \neq 68 \mu\text{F}$

(5.) Berechnung einer konkreten Realisierung der Prüfgröße

In welchem Intervall  $[c_u, c_o]$  muß der Mittelwert  $\bar{x}$  der Stichprobe liegen, damit auf dem Signifikanzniveau  $\alpha = 5\%$  die Nullhypothese

$$H_0 : \quad \mu = \mu_0 = 68 \mu\text{F}$$

akzeptiert, d.h. nicht abgelehnt wird?

**Lösung**

$$\begin{aligned} \gamma &= 1 - \alpha = P(c_u \leq \bar{X} \leq c_o) \\ &= P\left(\frac{c_u - \mu_0}{\sigma/\sqrt{n}} \leq \bar{Z} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq \frac{c_o - \mu_0}{\sigma/\sqrt{n}}\right) \quad \left[\frac{\sigma}{\sqrt{n}} = 0.6\right] \\ &= \Phi\left(\frac{c_o - 68}{0.6}\right) - \Phi\left(\frac{c_u - 68}{0.6}\right). \end{aligned}$$

Es wird ein symmetrisches Vertrauensintervall um den Sollwert  $\mu_0 = 68 \mu\text{F}$  gewählt, d.h.

$$|c_o - 68| = |c_u - 68|.$$

Dann folgt mit Hilfe von

$$c := \frac{c_o - 68}{0.6} = \frac{68 - c_u}{0.6} = -\frac{c_u - 68}{0.6}$$

und der üblichen Bedingung für das  $(1 - \frac{\alpha}{2})$ -Quantil der Standardnormalverteilung

$$\begin{aligned} \gamma &= 1 - \alpha = \Phi(c) - \Phi(-c) = 2\Phi(c) - 1, \\ c &= \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = \Phi^{-1}\left(1 - \frac{0.05}{2}\right) = \Phi^{-1}(0.975) = 1.960. \end{aligned}$$

Damit ergeben sich die Vertrauensgrenzen zu

$$c_u = \mu_0 - c \cdot \frac{\sigma}{\sqrt{n}} = 68 - 1.960 \cdot \frac{3.6}{\sqrt{36}} = 66.82 \mu\text{F}$$

$$c_o = \mu_0 + c \cdot \frac{\sigma}{\sqrt{n}} = 68 + 1.960 \cdot \frac{3.6}{\sqrt{36}} = 69.18 \mu\text{F}$$

(6.) Testentscheidung

Auf dem Signifikanzniveau  $\alpha = 5\%$  kann die Nullhypothese

$$H_0 : \mu = \mu_0 := 68 \mu\text{F}$$

aufgrund der verwendeten Stichprobe dann nicht abgelehnt werden, wenn für den Stichprobenmittelwert gilt:

$$\bar{x} \in [66.82 \mu\text{F}, 69.18 \mu\text{F}].$$

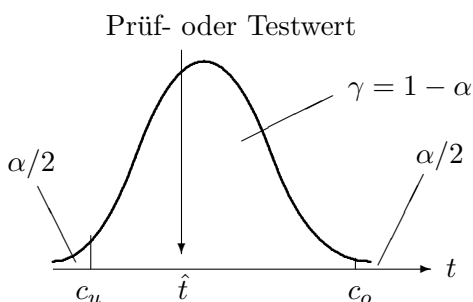


Abbildung 10.8: Der Prüf- oder Testwert  $\hat{t}$  fällt in den nicht-kritischen Bereich (Annahmebereich), die Nullhypothese  $H_0 : \theta = \theta_0$  wird daher angenommen.

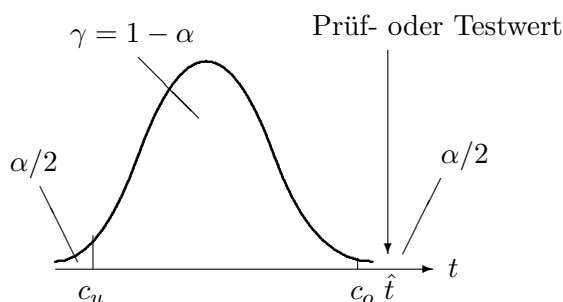


Abbildung 10.9: Der Prüf- oder Testwert  $\hat{t}$  fällt in den kritischen Bereich (Ablehnungsbereich), die Nullhypothese  $H_0 : \theta = \theta_0$  wird daher zugunsten der Alternativhypothese  $H_1 : \theta \neq \theta_0$  abgelehnt.

### Beispiel 10.20

- (i) Wie groß muß das Signifikanzniveau  $\alpha$  sein, falls die Nullhypothese dann nicht abgelehnt werden soll, falls  $\bar{x} \in [67 \mu\text{F}, 69 \mu\text{F}]$  gilt?
- (ii) Auf welchen Wert kann das Signifikanzniveau  $\alpha$  aus (i) verkleinert werden, wenn der Stichprobenumfang auf  $n = 64$  vergrößert wird?

### Lösung

$$(i) \quad c_u = \mu_0 - c \cdot \frac{\sigma}{\sqrt{n}} \quad \longrightarrow \quad c = \frac{\sqrt{n}}{\sigma} (\mu_0 - c_u)$$

$$\longrightarrow \quad c = \frac{\sqrt{36}}{3.6} (68 - 67) \mu\text{F} = 1.67 \mu\text{F}$$

$$\longrightarrow \quad \alpha = 1 - [2\Phi(c) - 1] = 2 [1 - \Phi(c)]$$

$$= 2 (1 - 0.9525) = 0.095 = 9.5\%$$



$$\begin{aligned}
 \text{(ii)} \quad c &= \frac{\sqrt{n}}{\sigma} (\mu_0 - c_u) = \frac{\sqrt{64}}{3.6} (68u - 67) = 2.22 \mu\text{F} \\
 \longrightarrow \alpha &= 1 - [2\Phi(c) - 1] = 2 [1 - \Phi(c)] \\
 &= 2 (1 - 0.9868) = 0.0264 = 2.64 \%
 \end{aligned}$$

### 10.2.3.2 Fehler zweiter Art

Ein Fehler zweiter Art liegt vor, wenn eine Nullhypothese  $H_0$  angenommen (= nicht verworfen) wird, obwohl sie falsch ist.

Liegt der aktuelle Prüfwert  $\hat{t}$  rechts von der kritischen Grenze  $c$ , so muß die angenommene Verteilung verworfen werden,  $H_0$  wird abgelehnt. Man begeht einen Fehler erster Art, falls  $H_0$  dennoch wahr ist.

Liegt der aktuelle Prüfwert  $\hat{t}$  links von der kritischen Grenze  $c$ , so muß die angenommene Verteilung (vorerst) akzeptiert werden,  $H_0$  wird angenommen. Man begeht dann einen Fehler zweiter Art, falls  $H_0$  falsch ist.

Ohne weitergehende Änderungen im Aufbau eines Versuches, beispielsweise eine Erhöhung des Umfanges  $n$  einer Stichprobe, führt eine Verkleinerung des Fehlers erster Art  $\alpha$  zu einer Vergrößerung des Fehlers zweiter Art  $\beta$  und umgekehrt, s. Abb. 10.10.

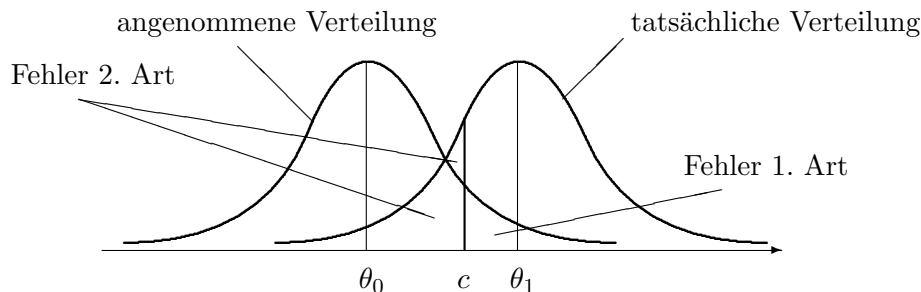


Abbildung 10.10: Fehler erster und zweiter Art

Entscheidet man sich in einem konkreten Fall z.B. für ein sehr kleines  $\alpha$  (= geringes Risiko, die Nullhypothese ablehnen zu müssen), so nimmt man gleichzeitig ein deutlich höheres Risiko für einen Fehler zweiter Art in Kauf (= hohes Risiko, eine evtl. falsche Nullhypothese zu akzeptieren).

#### Beispiel 10.21 (Wiederaufgriff von Beispiel 10.19)

Die Reduzierung des Signifikanzniveaus  $\alpha$  allein ist also keine Garantie dafür, daß der Test das gewünschte Ergebnis liefert. Man muß die Wahrscheinlichkeit  $\beta$ , also die Ablehnungswahrscheinlichkeit für diejenigen Alternativhypothesen berechnen, welche unbedingt akzeptiert werden sollten, wenn sie zutreffen. Falls es beispielsweise wichtig ist,  $H_0$  abzulehnen, falls  $\mu \leq 66 \mu\text{F}$  oder  $\mu \geq 70 \mu\text{F}$  gilt, so sollte die Wahrscheinlichkeit für einen Fehler zweiter Art klein gehalten werden.

Auf dem Signifikanzniveau  $\alpha = 5\%$  gilt

$$\beta = P(66.82 \mu\text{F} \leq \bar{X} \leq 69.18 \mu\text{F} \mid H_1 \text{ ist wahr}).$$

Aufgrund der Symmetrie der Normalverteilung braucht man nur eine der beiden Alternativen  $\mu = 66 \mu\text{F}$  oder  $\mu = 70 \mu\text{F}$  zu betrachten, o.B.d.A. die letzte, s. Abb. 10.11.

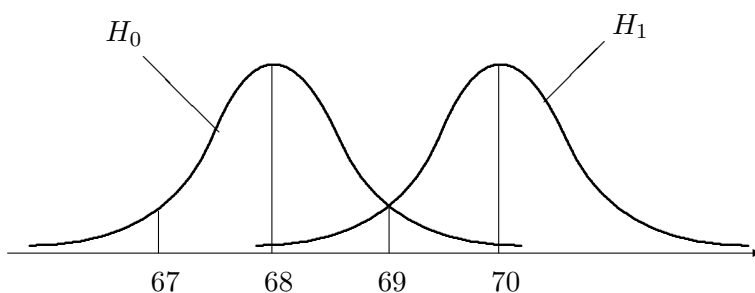


Abbildung 10.11: Fehler erster und zweiter Art

Die Stichprobe besteht aus  $n = 36$  Kondensatoren, also ist  $\sqrt{n} = 6$ . Weiterhin ist  $\sigma = 3.6 \mu\text{F}$ . Dann erhält man die Werte

$$z_1 := \frac{66.82 - 70}{0.6} = -5.30 \quad , \quad z_2 := \frac{69.18 - 70}{0.6} = -1.37$$

$$\begin{aligned} \longrightarrow \quad \beta &= P(-5.30 \leq Z \leq -1.37) \\ &= \Phi(-1.37) - \Phi(-5.30) \\ &= 1 - \Phi(1.37) - [1 - \Phi(5.30)] \\ &= \Phi(5.30) - \Phi(1.37) \\ &= 1 - 0.9147 \\ &= 0.0853 = 8.53\% \end{aligned}$$

Wäre die Alternativhypothese  $\mu = 66 \mu\text{F}$ , so würde man dasselbe Resultat erhalten.

Bei einer Stichprobe vom Umfang  $n = 64$  wäre der Fehler  $\beta$  zweiter Art kleiner:

$$n = 64 \quad \Longrightarrow \quad \beta = \Phi(7.07) - \Phi(1.82) = 0.0344 = 3.44\%,$$

und damit wäre auch die Chance kleiner, eine falsche Nullhypothese zu akzeptieren.

Die Wahrscheinlichkeit für einen Fehler zweiter Art wächst schnell, wenn sich der angenommene Wert  $\bar{x}$  für  $\mu_0$  dem wahren Wert  $\mu$  annähert. In einer solchen Situation stört es in der Regel jedoch nicht, wenn man einen Fehler zweiter Art begeht.

Ist etwa die Alternativhypothese

$$H_1 : \quad \mu = 68.5 \mu\text{F}$$

wahr, so ist der Fehler zweiter Art für den statistischen Schluß

$$\bar{x} = 68 \mu\text{F} \quad \Longrightarrow \quad \mu = \mu_0 = 68 \mu\text{F}$$

nicht weiter schlimm. Die Wahrscheinlichkeit für einen solchen Fehler ist groß, auch für große Stichproben. Beispielsweise gilt für  $n = 64$ :

$$\beta = P(66.82 \mu\text{F} \leq \bar{X} \leq 69.18 \mu\text{F} \mid H_1 : \mu = 68.5 \mu\text{F}),$$

$$z_1 := \frac{66.82 - 68.5}{0.45} = -3.73 \quad , \quad z_2 := \frac{69.18 - 68.5}{0.45} = 1.51,$$

$$\begin{aligned} \longrightarrow \quad \beta &= P(-3.73 \leq Z \leq 1.51) \\ &= \Phi(1.51) - [1 - \Phi(3.73)] \\ &= \Phi(1.51) + \Phi(3.73) - 1 \\ &= 0.9345 + 0.999 - 1 \\ &= 0.9344 = 93.44 \% \end{aligned}$$

### Bemerkung

- (i) Zwischen den Wahrscheinlichkeiten  $\alpha$  und  $\beta$  für die Fehler erster und zweiter Art besteht ein relativ komplizierter Zusammenhang. Im allgemeinen führt eine Verkleinerung der einen Größe zu einer Vergrößerung der anderen und umgekehrt.
- (ii) Unter Zuhilfenahme bedingter Wahrscheinlichkeiten lassen sich die Fehler erster und zweiter Art folgendermaßen schreiben:

$$\begin{aligned} \alpha &= P(H_0 \text{ wird verworfen} \mid H_0 \text{ ist wahr}) \\ \beta &= P(H_0 \text{ wird akzeptiert} \mid H_0 \text{ ist falsch}) \end{aligned}$$

Der  $\alpha$ -Fehler ist der falsch-positive Fehler eines Tests und bedeutet damit die *Spezifität des Tests*: Tests mit hoher Spezifität haben einen geringen Fehler erster Art.

Der  $\beta$ -Fehler ist der falsch-negative Fehler eines Tests und bedeutet damit die *Sensitivität des Tests*: Tests mit hoher Sensitivität haben einen geringen Fehler zweiter Art.

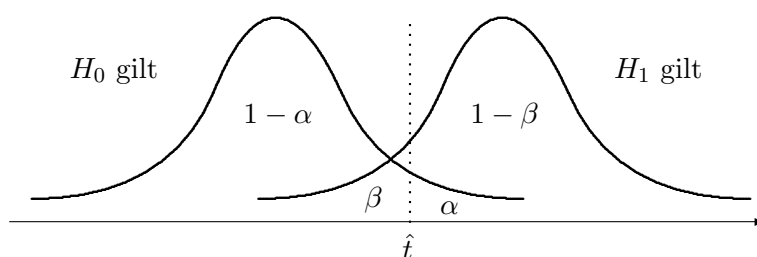


Abbildung 10.12: Kritischer Wert der Prüfgröße  $T$  in Abhängigkeit von  $\alpha$  und  $\beta$  (einseitiger Test)

### 10.2.3.3 Trennschärfe (Power) eines Tests

Ein Fehler  $\alpha$  erster Art bedeutet die Wahrscheinlichkeit, die Nullhypothese  $H_0$  abzulehnen, obwohl sie richtig ist; diesen Fehler möchte man eigentlich klein halten. Je kleiner nämlich  $\alpha$  ist, desto kleiner ist auch der kritische Bereich, also der Ablehnungsbereich eines Tests, s. Abb. 10.12.

Jeder Problemstellung ist die Irrtumswahrscheinlichkeit  $\alpha$  vorzugeben. Ist  $\alpha$  sehr klein, so wird  $H_0$  „selten“ abgelehnt, mit allen sich daraus ergebenden Konsequenzen, falls  $H_0$  doch einmal falsch ist. Beispielsweise kann dann die Wahrscheinlichkeit  $\beta$  für einen Fehler zweiter Art recht groß werden.

Ist nun  $H_0$  falsch, also  $H_1$  richtig, so soll die Alternativhypothese  $H_1$  nicht abgelehnt werden.

#### Definition 10.12

Die *Trennschärfe* eines Tests (auch: *Güte*, *Teststärke*, *Power*) ist definiert als die Wahrscheinlichkeit,  $H_0$  abzulehnen unter der Bedingung, daß  $H_1$  zutrifft:

$$\text{Power} := P(T \in S_C \mid H_1).$$

Hierbei bedeuten  $T$  die jeweilige Testgröße und  $S_C$  den kritischen Bereich (Definition 10.10).

Die Power eines Tests ist demnach die Wahrscheinlichkeit, eine richtige Alternativhypothese  $H_1$  auch als solche zu erkennen. Damit gilt also:

$$\text{Power} = 1 - \beta.$$

#### Bemerkung

- (i) Der kritische Bereich (Ablehnungsbereich) sollte demnach so gewählt werden, daß gilt:
- Irrtumswahrscheinlichkeit  $\alpha$  möglichst klein
  - Trennschärfe  $1 - \beta$  möglichst groß

	$H_0$ ist wahr	$H_0$ ist falsch
$H_0$ wird akzeptiert	Richtige Entscheidung $p_1 := 1 - \alpha$	Fehlentscheidung Fehler 2. Art $p_2 := \beta$
$H_0$ wird verworfen	Fehlentscheidung Fehler 1. Art $p_3 := \alpha$	Richtige Entscheidung $p_4 := 1 - \beta$

Tabelle 10.6: Entscheidungsmöglichkeiten bei einem Hypothesentest

- (ii) Die mit einem statistischen Test auftretenden vier möglichen Entscheidungen sollen an Hand der Tabelle 10.6 verdeutlicht werden.
- (iii) Oft wird bei einem Parametertest nur die Irrtumswahrscheinlichkeit  $\alpha$  vorgegeben und auf eine Alternative  $H_1$  verzichtet. Solch ein Test heißt *Signifikanztest*.
- (iv) Das Verwerfen einer Hypothese bedeutet den statistischen Schluß, daß sie falsch ist, während deren Akzeptanz lediglich bedeutet, daß man (im Moment noch, bei Vorliegen der aktuellen Stichprobe) keine Anhaltspunkte hat, das Gegenteil anzunehmen.  
Aus diesem Grunde sollte man solche Aussagen als Nullhypothesen aufstellen, die man zu verwerfen gedenkt. Das Ziel sollte also eher sein, eine Nullhypothese zu falsifizieren, anstatt sie zu verifizieren.
- (v) Aufgrund des Zusammenhanges zwischen Parametertests und Konfidenzintervallen erkennt man, daß man die Größe des Ablehnungsbereiches  $S_C$  im Stichprobenraum, also die Wahrscheinlichkeit  $\alpha$  für einen Fehler erster Art, durch Anpassen der Grenzen  $c_u$  und  $c_o$  des Konfidenzintervalles  $\mathcal{K}$  reduzieren kann. Geschieht dieses durch eine Vergrößerung des Stichprobenumfanges  $n$ , so bewirkt das eine gleichzeitige Verkleinerung von  $\alpha$  und  $\beta$ . Allerdings steigt damit auch der Testaufwand (Kosten).
- (vi) Sei  $\Theta := \Theta_1 \cup \Theta_2$  eine disjunkte Zerlegung des Parameterraumes eines Tests. Die Wahrscheinlichkeit für einen Fehler zweiter Art in Abhängigkeit vom Parameter  $\theta$  sei gegeben durch

$$\beta := b(\theta) := P(T \notin S_C \mid \theta \in \Theta_1).$$

Da der wahre Wert von  $\theta$  nicht bekannt ist, läßt sich diese Wahrscheinlichkeit nicht berechnen.

Es läßt sich aber  $b$  als Funktion von  $\theta$  über der Menge  $\Theta_1$  untersuchen:

$$\beta = b(\theta) \quad (\theta \in \Theta_1).$$

Diese Funktion heißt *Operationscharakteristik* (*operating characteristic curve*, *Annahmekennlinie*). Sie ist das Komplement der Trennschärfe:

$$\begin{array}{rclcl} \text{Operationscharakteristik} & + & \text{Trennschärfe} & = & 1 \\ \beta = b(\theta) & + & 1 - b(\theta) & = & 1 \end{array}$$

Die Operationscharakteristik kann dazu dienen, verschiedene Tests für das gleiche Testproblem miteinander zu vergleichen und somit möglicherweise unter allen Tests zum gleichen Niveau  $\alpha$  für das vorliegende Problem einen *besten Test* zu bestimmen.

# Stichwortverzeichnis

- $E(X)$ , 84
- $N(\cdot; 0, 1)$ , 124
- $N(\cdot; \mu, \sigma)$ , 123
- $P(A | B)$ , 38
- $V(X)$ , 92
- $X$ , 60
- $\Omega$ , 7
- $\Phi(\cdot)$ , 124
- $\alpha$ , 175
- $\alpha$ -Fehler, 209
- $\beta$ -Fehler, 209
- $\binom{n}{k}$ , 15
- $\chi^2$ -Test, 205
- $\chi^2$ -Verteilung, 144
- $\gamma$ , 175
- $\mu$ , 84, 85
- $\phi(\cdot)$ , 124
- $\sigma$ , 92, 95
- $\sigma^2$ , 92
- $t$ -Test, 145, 203
- $t$ -Verteilung von Student, 145
- Ablehnungsbereich
  - eines Tests, 196
- Additionssatz, 31
- Akzeptanzbereich, 196
- Alternativenmenge, 194
- Alternativhypothese, 192
- Annahmebereich
  - eines Tests, 196
- Anpassungstest, 192
- Anzahl der Freiheitsgrade, 145
- asymptotisch erwartungstreu, 164
- asymptotisch normalverteilt, 136
- Ausfallrate, 80
- Ausfallwahrscheinlichkeit, 80
- Ausprägung, 148
- Ausschöpfung, 42
- Baumdiagramm, 49
- Bayes
  - Formel von, 43
  - Gleichung von, 40
- bedingte Wahrscheinlichkeit, 38
- Beobachtungsmenge, 148
- Bereichsschätzung, 160
- Bernoulli-Experiment, 104
  - vom Umfang  $n$ , 106
- Bernoulli-Kette
  - vom Umfang  $n$ , 106
- Binomialkoeffizient, 15, 16
- Binomialverteilung, 108
- Binomischer Satz, 16
- charakteristische Funktion, 121, 170
- Dichtefunktion
  - diskrete, 62
  - kontinuierliche, 76
  - stetige, 76
- Differenz
  - von Mengen, 8
- disjunkt, 9
- Durchschnitt, 8
- effizient, 164
- Effizienz, 45
- Elementarereignis, 7
- elementfremd, *siehe* disjunkt
- empirische Varianz, *siehe* Varianz
- Ereignis, 7
  - günstiges, 23
  - mögliches, 23
- Ereignisbaum, 49
- Ereignisraum, 7
- Erfolgswahrscheinlichkeit, 105
- Ergebnismenge, 7
- Ergebnisraum, 7
- erwartungstreu, 164
- Erwartungswert, 84, 85

- Experiment  
  deterministisches, 5  
  stochastisches, 5  
  zufälliges, 2, 5
- Exponentialverteilung, 140
- Exzeß, 98
- Fakultät, 13
- Fehler  
  erster Art, 197, 209  
  zweiter Art, 209
- Freiheitsgrad, *siehe* Anzahl der Freiheitsgrade
- Funktion  
  meßbare, 61
- Güte  
  eines Tests, 216
- Galton'sches Brett, 139
- Gauß-Test, 201
- Gauß-Verteilung, 122
- geometrische Verteilung, 120
- Gesetz der großen Zahlen, 110
- Gleichverteilung  
  diskrete, 102  
  stetige, 121
- Grenzwertsatz  
  von Moivre-Laplace, 135  
  zentraler, 136
- Grundgesamtheit, 148, 152
- Häufigkeit  
  absolute, 26  
  relative, 26
- Histogramm, 154
- hochsignifikant, 199
- Hypergeometrische Verteilung, 112, 114
- Hypergeometrische Zufallsvariable, 113
- Hypothese, 191  
  einfache, 199  
  zusammengesetzte, 199
- Hypothesenmenge, 194
- Intervallschätzung, 160
- Irrtumswahrscheinlichkeit, 175, 197
- Kennwerte  
  statistische, 156
- Klasse  
  von Stichprobenwerten, 154
- Kombination, 14  
  mit Wiederholung, 16  
  ohne Wiederholung, 15
- Komplement, 8
- konditionale Wahrscheinlichkeit, *siehe* bedingte Wahrscheinlichkeit
- Konfidenzgrenzen, 175
- Konfidenzintervall, 175  
  einseitiges, 181
- Konfidenzniveau, 175
- konsistent, 164
- kontinuierliche Zufallsvariable, 76
- konträr, 9
- Korrelationskoeffizient, 166
- Kovarianz, 166
- Laplace-Experiment, 23
- Linearität  
  des Erwartungswertes, 90
- Logarithmische Normalverteilung, 134
- Markov-Ungleichung, 99
- Maximum-Likelihood-Schätzer, 169
- Maximum-Likelihood-Schätzwert, 169
- Median, 131, 156
- mehrstufiges Zufallsexperiment, 48
- Merkmal, 148
- Merkmalraum, 7  
  reduzierter, 37
- Merkmalsausprägung, 152
- Merkmalwert, 61
- Mißerfolgswahrscheinlichkeit, 105
- Mittelwert  
  arithmetischer, 156
- Modalwert, 156
- Moment, 98
- Nichtparametrischer Test, 192
- Normalverteilung, 122
- Null-Eins-Experiment, 104
- Nullhypothese, 192, 194
- Operationscharakteristik, 218



- Parametermenge, 194  
Parameterraum, 194  
Parametertest, 192, 193  
    zweiseitiger, 194  
Partition, 42  
Percentil, 131  
Permutation, 12  
Poisson-Dichte, 118  
Potenzmenge, 8  
Power  
    eines Tests, 216  
Prüfstatistik, 197  
Punktschätzung, 160, 174  
  
Quantil, 131  
Quartil, 131  
  
Realisierung  
    einer Zufallsvariable, 61  
Rechteckverteilung, 121  
Regeln  
    von de Morgan, 9  
  
Satz  
    von der totalen Wahrscheinlichkeit, 42  
    von Steiner, 94  
Schätzer, 162  
Schätzfunktion, 162  
Schätzvariable, 162  
Schätzwert, 162  
Schiefe, 98  
Sensitivität, 45  
    eines Tests, 215  
signifikant, 195, 199  
Signifikanzniveau, 194, 197  
Signifikanztest, 217  
Signifikanzzahl, 197  
Spannweite  
    einer Stichprobe, 158  
Spezifität, 45  
    eines Tests, 215  
Spur, 39  
Standardabweichung, 92, 158  
Standardfehler, 178  
Standardnormalverteilung, 124  
Standardtransformation, 101, 125  
Statistik, 162  
    statistische Aussage, 148  
    statistische Sicherheit, 175  
Steiner  
    Satz von, 94  
stetige Zufallsvariable, 76  
Stichprobe, 12, 150, 152  
    geordnete, 12, 156  
    repräsentative, 153  
    Spannweite, 158  
    Stichprobenwert, 152  
    ungeordnete, 12  
    unverzerrte, 152  
    verzerrte, 152  
    Zufallsstichprobe, 152  
Stichprobenfehler, 178  
Stichprobenfunktion, 162, 175  
Stichprobenmittel, *siehe* Mittelwert, 163  
Stichprobenmittelwert, 163  
Stichprobenvariable, 143, 162  
Stichprobenvarianz, 163  
stochastisch unabhängig, 52, 59  
Streuung, 92, 95  
    einer Stichprobe, 158  
Student, *siehe* *t*-Verteilung  
Student-Faktor, 185  
  
Test  
    einseitiger, 197  
    Sensitivität, 215  
    Spezifität, 215  
    Stärke, 216  
    statistischer, 192  
    zweiseitiger, 197  
Teststatistik, 192  
Testverteilung, 144  
Toleranzgrenze, 189  
Toleranzintervall, 190  
Trägheitsmomentensatz, 94  
Trennschärfe, 216  
Tschebyschev'sche Ungleichung, 95  
  
Überlebenswahrscheinlichkeit, 80  
unabhängig, 52, 59  
Unabhängigkeitstest, 192  
Ungleichung  
    von Markov, 99

- von Tschebyschev, 95
- uniforme Verteilung, 121
- unvereinbar, 9
- Urnenmodell, 11
- Varianz, 92
  - einer Stichprobe, 158
  - empirische, 158
- Variation, 18
  - mit Wiederholung, 19
  - ohne Wiederholung, 18
- Vereinigung, 8
- Verschiebungsregel, 93
- Verteilung, 65, 148
  - $\chi^2$ -Verteilung, 144
  - $t$ -Verteilung von Student, 145
  - Binomial-, 108
  - Exponential-, 140
  - geometrische, 120
  - hypergeometrische, 112, 114
  - logarithmische Normal-, 134
  - Normalverteilung, 122
  - schiefe, 133
  - Standardnormal-, 124
  - uniforme, 121
- Verteilungsfunktion, 65
- Verteilungsgesetz, 28
- Vertrauensgrenzen, 175
- Vertrauensintervall, 175
- Vertrauensniveau, 175
- Vertrauenswahrscheinlichkeit, 175
- Wahrscheinlichkeit, 23
  - bedingte, 38
  - empirische, 2
  - klassische Definition, 24
  - subjektive, 2, 174
- Wahrscheinlichkeitsmaß, 28
- Wahrscheinlichkeitsraum, 27
- Zähldichte, 62
- Zufallsexperiment, 6
  - mehrstufiges, 48
- Zufallsvariable, 60
  - diskrete, 62
  - exponentialverteilte, 140
  - hypergeometrische, 113
  - kontinuierliche, 76
  - negative binomialverteilte, 120
  - normalverteilte, 122
  - Realisierung, 61
  - standardisierte, 101
  - stetige, 76
  - Verteilung, 65
  - Zuverlässigkeit, 80
  - Zweistichproben-Gauß-Test, 202