

# Mathematik III: Stochastik für Bauingenieure

Edgar Kaufmann

12.04.2010

## Inhaltsverzeichnis

<b>1 Grundlagen der deskriptiven Statistik</b>	<b>2</b>
<b>2 Ereignisse und Wahrscheinlichkeit</b>	<b>10</b>
<b>3 Diskrete Zufallsvariable</b>	<b>16</b>
<b>4 Stetige Zufallsvariable</b>	<b>19</b>
<b>5 Kennziffern von Verteilungen</b>	<b>21</b>
<b>6 Unabhängigkeit</b>	<b>24</b>
<b>7 Wahrscheinlichkeitsverteilungen</b>	<b>25</b>
<b>8 Grenzwertsätze</b>	<b>32</b>
<b>9 Parameterschätzung</b>	<b>36</b>
<b>10 Konfidenzintervalle</b>	<b>41</b>
<b>11 Korrelation und Regression</b>	<b>44</b>
<b>12 Hypothesentests</b>	<b>48</b>

## Literatur

1. Cramer/Kamps: Grundlagen der Wahrscheinlichkeitsrechnung und Statistik, Springer-Verlag, 2008
2. Henze, N.: Stochastik für Einsteiger, Vieweg+Teubner, 2009
3. Papula, L.: Mathematik für Ingenieure und Naturwissenschaftler, Bd 3, Vieweg+Teubner, 2008

# 1 Grundlagen der deskriptiven Statistik

Die Hauptaufgabe der deskriptiven oder beschreibenden Statistik ist die Darstellung und die Aufbereitung von beobachteten Daten in Form von Tabellen und Grafiken und die Berechnung von statistischen Kenngrößen. Im Gegensatz zur beurteilenden oder schließenden Statistik wird z.B. nicht versucht, anhand einer Stichprobe Aussagen über die Grundgesamtheit zu machen.

## Untersuchungseinheiten und Merkmale

Bei einer **statistischen Erhebung** werden an **Untersuchungseinheiten** oder **Merkmalsträgern** (z.B. Studierende einer Universität) die **Ausprägungen** von einem oder mehreren **Merkmalen** (z.B. Semester, Geschlecht, Alter) beobachtet und notiert.

Bei Merkmalen wird zwischen **quantitativen** (in natürlicher Weise zahlenmäßig erfassbaren) und **qualitativen** (artmäßig erfassbaren) **Merkmalen** unterschieden.

**Qualitative Merkmale** werden weiter unterteilt in

- **nominale Merkmale** Die Klassifizierung der Ausprägungen erfolgt nach rein qualitativen Gesichtspunkten. Eine Codierung ist rein willkürlich, z.B. ledig=0, verheiratet=1, geschieden=1, verwitwet=2.
- **ordinale Merkmale** Die Ausprägungen weisen eine natürliche Rangfolge auf, z.B. erreichter Abschluss.

Bei **quantitativen Merkmalen** unterscheidet man

- **diskrete Merkmale** Es können nur endlich (oder abzählbar unendlich) viele Zahlenwerte auftreten, z.B. Alter in Jahren, Semesteranzahl, Anzahl der defekten Bauteile in einer Lieferung.
- **stetige Merkmale** Als Wert kann jede reelle Zahl in einem Intervall angenommen werden, z.B. Druckfestigkeit, Gewicht, Größe etc.

**Beispiele:**

Untersuchungseinheiten	Merkmal	Ausprägungen	Merkmalstyp
Absolvent im Jahr xyz	Semester	...,8,9,10,..	quantitativ/diskret
	Geschlecht	m, w oder 0, 1	qualitativ/nominal
	Familienstand	ledig, verheiratet,...	qualitativ/nominal
	Abschluss	Bachelor,Master,Prom.	qualitativ/ordinal
Betonwürfel	Druckfestigkeit (in N/mm <sup>2</sup> )	reelle Zahl, z.B. 39.9	quantitativ/stetig

Im Folgenden wird es sich überwiegend um quantitative Merkmale handeln. Die Übergänge zwischen diskreten und stetigen Merkmalen sind aufgrund von vereinbarter oder technisch bedingter Messgenauigkeit fließend. Wird z.B. die Druckfestigkeit mit einer Nachkommastelle gemessen und notiert, so kann sie als diskretisiertes stetiges Merkmal angesehen werden.

## Grundgesamtheit und Stichprobe

Die Menge der Untersuchungseinheiten wird als **Grundgesamtheit** oder **Population** bezeichnet. Sie kann endlich oder unendlich sein.

**Beispiel:** Die Menge aller Absolventen der Universität Siegen im Jahr xyz bilden eine Grundgesamtheit.

Eine **Stichprobe** ist eine „zufällig gewonnene“ Auswahl aus der Grundgesamtheit. Besteht die Auswahl aus  $n$  Untersuchungseinheiten, so sprechen wir von einer **Stichprobe vom Umfang  $n$** .

Im Folgenden sei  $x_1, \dots, x_n$  eine Stichprobe vom Umfang  $n$  eines Merkmals  $X$ .

**Beispiel:** Per Computer werden  $n = 20$  Absolventen des Jahrgangs xyz „zufällig“ ausgewählt und die Anzahl der benötigten Semester notiert.

**Beispiel 1.1** 24 Haushalte werden „zufällig“ ausgewählt und nach der Zahl der in diesen Haushalten lebenden Kinder befragt. Es ergibt sich folgende fiktive Stichprobe:

$$(x_1, \dots, x_{24}) = (0, 0, 1, 3, 2, 0, 1, 0, 0, 3, 0, 0, 0, 1, 1, 2, 1, 0, 0, 2, 0, 0, 1, 3)$$

## Empirische Häufigkeitsverteilung und Stabdiagramme

Besitzt ein Merkmal die möglichen Ausprägungen  $a_1, a_2, \dots, a_m$ , so bilden wir die absoluten Häufigkeiten

$$n_1 = \#\{i : x_i = a_1\}, \quad n_2 = \#\{i : x_i = a_2\} \quad \text{usw.}$$

bzw. allgemeiner

$$n_j := \#\{i : x_i = a_j\}, \quad j = 1, \dots, m.$$

**Beispiel:** In Beispiel 1.1 besitzt das Merkmal „Anzahl der Kinder“ die möglichen Ausprägungen  $0, 1, 2, 3, 4, \dots$

Man erhält die **absoluten Häufigkeiten**

Anzahl der Kinder $a_j$	0	1	2	3	mehr als 3
absolute Häufigkeit $n_j$	12	6	3	3	0

Zur Darstellung von Häufigkeiten ist der Begriff der Indikatorfunktion nützlich.

**Definition 1.2** Sei  $\Omega$  eine Menge und  $A \subset \Omega$  eine Teilmenge von  $\Omega$ . Die durch

$$1_A(\omega) := \begin{cases} 1, & \text{falls } \omega \in A \\ 0, & \text{sonst} \end{cases}, \quad \omega \in \Omega,$$

definierte Funktion  $1_A : \Omega \rightarrow \mathbb{R}$  heißt **Indikatorfunktion** von  $A$ .

**Beispiele:**

- (i) Ist  $\Omega = \mathbb{R}$ , dann ist  $1_{(-\infty, 0]}(x) = 1$ , falls  $x \leq 0$  und 0 sonst.

(ii) In Beispiel 1.1 gilt  $1_{\{0\}}(x_i) = 1$ , falls  $x_i = 0$  und 0 sonst. Die absolute Häufigkeit der kinderlosen Haushalte lässt sich wie folgt ermitteln:

$$n_1 = 1_{\{0\}}(x_1) + 1_{\{0\}}(x_2) + \dots + 1_{\{0\}}(x_{24}) = \sum_{i=1}^{24} 1_{\{0\}}(x_i).$$

Entsprechend gilt

$$n_j = \sum_{i=1}^{24} 1_{\{a_j\}}(x_i) = \sum_{i=1}^{24} 1_{\{j-1\}}(x_i), \quad j = 1, \dots, 4$$

**Definition 1.3** Sei  $X$  ein Merkmal mit den verschiedenen Ausprägungen  $a_1, a_2, \dots, a_m$  und  $x_1, \dots, x_n$  eine Stichprobe vom Umfang  $n$ . Dann heißt

$$n_j := \sum_{i=1}^n 1_{\{a_j\}}(x_i)$$

die **absolute Häufigkeit** von  $a_j$  und

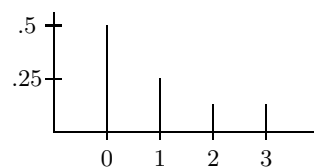
$$f_j := \frac{n_j}{n}$$

die **relative Häufigkeit** von  $a_j$

**Beispiel:** In Beispiel 1.1 gilt mit  $n = 24$

Anzahl der Kinder $a_j$	0	1	2	3	mehr als 3
absolute Häufigkeit $n_j$	12	6	3	3	0
relative Häufigkeit $f_j$	0.5	0.25	0.125	0.125	0

Zur grafischen Darstellung des Daten wählen wir ein **Stabdiagramm** für die relativen Häufigkeiten



## Histogramme

Eine Vorgehensweise wie im vorhergehenden Abschnitt ist für stetige Daten nicht sinnvoll. „Echte“ stetige Daten sind „fast immer“ verschieden (auf eine mathematisch präzise Aussage wird hier verzichtet). Selbst eine Diskretisierung durch Messungenauigkeiten kann noch zu sehr vielen Werten führen. (Beispiel: Gewichtsmessung aller Studierenden mit mehreren Nachkommastellen).

Bei Vorliegen reeller Stichprobenwerte  $x_1, \dots, x_n$  fasst man die Daten in Klassen zusammen. Hierzu werden reelle Zahlen

$$a_0 < a_1 < \dots < a_{m-1} < a_m$$

gewählt, sodass alle Daten in dem Intervall  $(a_0, a_m]$  liegen. Nun wird die **Klassen-Häufigkeit**  $n_j$  bestimmt durch

$$n_j := \sum_{i=1}^n 1_{(a_{j-1}, a_j]}(x_i)$$

Offenbar tritt hier ein Informationsverlust ein, denn aus den Klassen-Häufigkeiten  $n_j$  können die Datenwerte  $x_i$  nicht ermittelt werden.

Liegen Daten schon in der Form  $n_1, \dots, n_n$  vor, so heißen diese **gruppierte Daten**.

Die Höhe  $f_n(x)$  des Histogramms für  $x \in (a_{j-1}, a_j]$  berechnet sich aus der Länge des Intervalls gemäß der Gleichung

$$f_n(x)(a_j - a_{j-1}) = n_j/n.$$

Das **Histogramm** ist also gegeben durch die Funktion

$$f_n(x) = \frac{n_j}{n(a_j - a_{j-1})}, \quad a_{j-1} < x \leq a_j$$

und  $f_n(x) = 0$ , falls  $x \leq a_0$  oder  $x > a_m$ .

Die Formel zur Berechnung des Histogramms vereinfacht sich zu

$$f_n(x) = \frac{n_j}{n \cdot d}, \quad a_{j-1} < x \leq a_j,$$

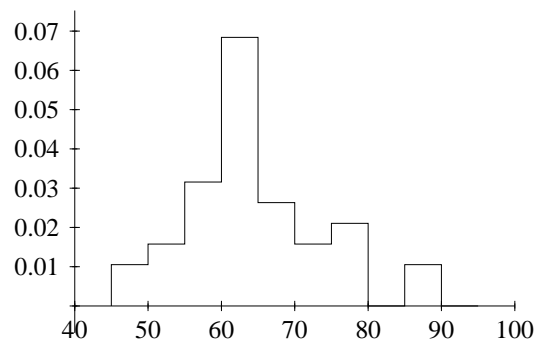
falls alle Intervalle die gleiche Länge  $d$  besitzen.

**Beispiel 1.4** Jährliche maximale Windgeschwindigkeiten in Vancouver von 1947 bis 1984

79.5 68.4 74 59.2 74 64.8 64.8 59.2 79.5 62.9 59.2 68.4 64.8 88.8 88.8 75.8 68.4 68.4 61  
51.8 62.9 64.8 61 61 70.3 68.4 55.5 64.8 77.7 57.3 48.1 53.6 55.5 62.9 61 61 51.8 48.1

Die absoluten Häufigkeiten der Windgeschwindigkeiten für die Intervalle  $(45, 50]$ ,  $(50, 55]$  usw. lautet dann

$(45, 50]$	$(50, 55]$	$(55, 60]$	$(60, 65]$	$(65, 70]$	$(70, 75]$	$(75, 80]$	$(80, 85]$	$(85, 90]$
2	3	6	13	5	3	4	0	2



### Stem-and-Leaf Diagramm

Eine dem Histogramm verwandte Darstellungsform ist das **Stem-and-Leaf Diagramm (Blatt-und Stamm Diagramm)**. Als Stamm werden für die Daten aus Beispiel 1.4

79.5 68.4 74 59.2 74 64.8 64.8 59.2 79.5 62.9 59.2 68.4 64.8 88.8 88.8 75.8 68.4 68.4 61  
51.8 62.9 64.8 61 61 70.3 68.4 55.5 64.8 77.7 57.3 48.1 53.6 55.5 62.9 61 61 51.8 48.1

die Zehner-Stellen gewählt, als Blatt die Einerstellen. Die Nachkommastellen werden vernachlässigt.

Blätter anfügen	Blätter sortieren
40 : 88	40 : 88
50 : 999157351	50 : 113557999
60 : 844284881241184211	60 : 111112224444488888
70 : 9449507	70 : 0445799
80 : 88	80 : 88

## Statistische Kenngrößen

In den vorherigen Abschnitten wurden Methoden vorgestellt, die nur einen geringen Informationsverlust (Stabdiagramm, Histogramm) aufweisen. Oft ist man daran interessiert, Daten mit wenigen Kenngrößen zu beschreiben.

### Lagemaße

Gegeben sei eine reellwertige Stichprobe  $x_1, \dots, x_n$  eines quantitativen Merkmals. Lagemaße dienen zur Beschreibung des „Zentrums“ der Stichprobe.

Das bekannteste Lagemaß ist das **arithmetische Mittel**

$$\bar{x} := \frac{x_1 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Liegt ein diskreter Datensatz in einer Häufigkeitsverteilung  $f_1, \dots, f_m$  vor, d.h.  $f_j$  ist die relative Häufigkeit für die Ausprägung  $a_j$ , dann berechnet sich  $\bar{x}$  gemäß

$$\bar{x} = \sum_{j=1}^m f_j a_j.$$

als **gewichtetes Mittel** von  $a_1, \dots, a_m$  mit den Gewichten  $f_1, \dots, f_m$ . Man beachte, dass  $\sum_{j=1}^m f_j = 1$  gilt.

Bei gruppierten Daten wählt man die Klassenmitte  $a_j^* := (a_j + a_{j-1})/2$  als Repräsentant für die Klasse und definiert

$$\bar{x}^* = \sum_{j=1}^m f_j a_j^*,$$

wobei  $f_j$  die relative Klassen-Häufigkeit bezeichnet.

**Beispiel:** Für die Kinderzahl aus Beispiel 1.1 erhält man das arithmetische Mittel

$$\bar{x} = \sum_{j=1}^m f_j a_j = \frac{12}{24} \cdot 0 + \frac{6}{24} \cdot 1 + \frac{3}{24} \cdot 2 + \frac{3}{24} \cdot 3 = \frac{21}{24} = 0.875$$

### Bemerkungen:

- (i) Für  $t \in \mathbb{R}$  sei  $g$  definiert durch  $g(t) = \sum_{i=1}^n (x_i - t)^2$ . Dann nimmt die Funktion  $g$  in  $t = \bar{x}$  ihr Minimum an.
- (ii) Setzt man  $y_i := x_i - \bar{x}$ ,  $i = 1, \dots, n$ , dann gilt  $\bar{y} = 0$ .

Ein weiteres wichtiges Lagemaß ist der empirische Median (Zentralwert). Hierzu werden die Daten der Größe nach geordnet. Wir schreiben

$x_{(j)}$  für den  $j$ -kleinsten Wert von  $x_1, \dots, x_n$ .

Es gilt dann

$$x_{(1)} = \min_{1 \leq i \leq n} x_i, \quad x_{(n)} = \max_{1 \leq i \leq n} x_i$$

und

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n-1)} \leq x_{(n)}$$

**Beispiel:** Ist  $(x_1, x_2, x_3, x_4) = (51, 3, 4, 3)$ , dann gilt  $(x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}) = (3, 3, 4, 51)$ .

Der **empirische Median (Zentralwert)**  $x_{1/2}$  ist definiert als

$$x_{1/2} := \begin{cases} x_{((n+1)/2)}, & \text{falls } n \text{ ungerade ist} \\ (x_{(n/2)} + x_{(n/2+1)})/2, & \text{falls } n \text{ gerade ist} \end{cases}$$

**Merke:** Mindestens die Hälfte der Werte der Stichprobe sind kleiner oder gleich  $x_{1/2}$  und mindestens die Hälfte der Werte der Stichprobe sind größer oder gleich  $x_{1/2}$ .

**Bemerkung:** Die Funktion  $g(t) = \sum_{i=1}^n |x_i - t|$  ist in  $t = x_{0.5}$  minimal.

**Beispiel:** Ist  $(x_1, x_2, x_3, x_4) = (51, 3, 4, 3)$ , dann ist  $n = 4$  gerade und es gilt

$$x_{1/2} = \frac{x_{(n/2)} + x_{(n/2+1)}}{2} = \frac{3 + 4}{2} = 3.5$$

Im Vergleich dazu gilt für das arithmetische Mittel  $\bar{x} = (51 + 3 + 4 + 3)/4 = 15.25$ .

Offenbar hat der im Vergleich zu den übrigen Daten große Wert  $x_1 = 51$  das arithmetische Mittel stark beeinflusst und nach rechts verschoben. Ist ein Wert relativ weit von den anderen Werten der Stichprobe entfernt, so wird dieser häufig als **Ausreißer** bezeichnet. Der Wert könnte z.B. auf einen Fehler in der Datenübertragung zurückzuführen sein, der wahre Wert könnte z.B. 5.1 statt 51 sein.

Im Gegensatz zum arithmetischen Mittel, das sehr **ausreißeranfällig** ist, würde sich der Median im Beispiel auch nicht verändern, wenn der größte Wert noch größer wäre. Der Median ist **robust** (unempfindlich) gegenüber Ausreißern.

**Achtung:** Extreme Werte können auch in der Problemstellung begründet sein (z.B. treten wenige aber sehr hohe Schäden im manchen Bereichen der Schadenversicherung auf).

In Verallgemeinerung des empirischen Medians heißt für  $0 < q < 1$  der Wert

$$x_q := \begin{cases} x_{(\lceil nq \rceil)}, & \text{falls } nq \notin \mathbb{N} \\ (x_{(nq)} + x_{(nq+1)})/2, & \text{falls } nq \in \mathbb{N} \end{cases}$$

das **empirische  $q$ -Quantil** von  $x_1, \dots, x_n$ . Dabei bezeichnet

$$\lceil nq + 1 \rceil := \max\{k \in \mathbb{Z} : k \leq nq + 1\}$$

die größte ganze Zahl, welche kleiner oder gleich  $nq + 1$  ist, z.B.  $\lceil 3.4 \rceil = 3$ ,  $\lceil 3 \rceil = 3$ .

Weiterhin heißen

$$\begin{array}{ll} x_{0.25} & \text{unteres Quartil} \\ x_{0.75} & \text{oberes Quartil.} \end{array}$$

## Streuungsmaße

Das arithmetische Mittel sagt nichts darüber aus, wie weit die Daten um den Wert  $\bar{x}$  streuen.

**Beispiel 1.5** Gegeben seien die Stichproben

i	1	2	3
$x_i$	10	10	10
$y_i$	1	10	19

Offenbar gilt  $\bar{x} = 10 = \bar{y}$  und ebenso  $x_{0.5} = 10 = y_{0.5}$ .

**Definition 1.6** Gegeben seien reelle Zahlen  $x_1, \dots, x_n$ . Dann heißt

$$s^2 := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

**empirische Varianz** oder **Stichprobenvarianz** und

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

**empirische Standardabweichung** oder **Stichprobenstandardabweichung** von  $x_1, \dots, x_n$ .

**Beispiel:** In Beispiel 1.5 erhält man eine Standardabweichung von  $s_x = 0$  bzw.  $s_y = 9$ .

**Bemerkungen:**

(i) Es gilt

$$s^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right).$$

(ii) Werden die Werte  $a_1, \dots, a_m$  mit den Häufigkeiten  $n_1, \dots, n_m$  beobachtet, dann gilt

$$s^2 = \frac{1}{n-1} \sum_{j=1}^m n_j (a_j - \bar{x})^2.$$

(iii) Im Fall gruppierter Daten verwendet man

$$(s^*)^2 = \frac{1}{n-1} \sum_{j=1}^m n_j (a_j^* - \bar{x}^*)^2.$$

## Weitere Streuungsmaße

Die Stichprobenstandardabweichung gehört zu den wichtigsten Streuungsmaßen. Sie hat jedoch den Nachteil, dass sie empfindlich gegen Ausreißer ist.

Weitere Streuungsmaße sind:

(i) die **mittlere absolute Abweichung**:  $\frac{1}{n} \sum_{i=1}^n |x_i - x_{0.5}|$

(ii) die **Stichprobenspannweite**:  $x_{(n)} - x_{(1)}$

(iii) der **Quartilsabstand**:  $x_{3/4} - x_{1/4}$

(iv) die **empirische Median-Abweichung**: Median von  $|x_1 - x_{0.5}|, \dots, |x_n - x_{0.5}|$



### Standardisierung

Gegeben sei eine Stichprobe  $x_1, \dots, x_n$ . Der Übergang zu den Werten

$$y_i := \frac{x_i - \bar{x}}{s_x}, \quad i = 1, \dots, n$$

wird als Standardisierung bezeichnet.

**Bemerkung:** Für die standardisierten Werte  $y_1, \dots, y_n$  gilt

$$\bar{y} = 0 \quad \text{und} \quad s_y = 1.$$

## 2 Ereignisse und Wahrscheinlichkeit

Eine wesentliche Aufgabe der Stochastik ist die Modellierung zufälliger Phänomene. Ein stochastischer Vorgang heißt (**ideales**) **Zufallsexperiment**, falls das Experiment

- zumindest theoretisch beliebig oft unter gleichen Bedingungen, den Versuchsbedingungen, wiederholt werden kann,
- mehrere einander ausschließende Ergebnisse haben kann, die im konkreten Fall nicht sicher vorausgesagt werden können, deren mögliche Werte aber vor der Durchführung des Experiments bekannt sind.

**Definition 2.1** Die Menge der möglichen Ergebnisse eines Zufallsexperiments heißt **Ergebnismenge** oder **Grundraum**  $\Omega$ .

**Beispiele:**

- (i) Die möglichen Ergebnisse beim Münzwurf sind Kopf oder Zahl. Codiert man diese mit 0 und 1, so erhält man als Grundraum die Menge  $\Omega = \{0, 1\}$ .
- (ii) Zur Beschreibung des Würfelwurfs wählt man als Grundraum  $\Omega = \{1, \dots, 6\}$ .
- (iii) Wird der Würfelwurf  $n$ -mal durchgeführt, dann lassen sich die Ergebnisse des Gesamtexperiments durch das  $n$ -Tupel

$$a = (a_1, \dots, a_n)$$

beschreiben. Der Grundraum ist die Menge der  $n$ -Tupel

$$\Omega = \{1, \dots, 6\}^n.$$

**Definition 2.2** (i) Eine Teilmenge  $A \subset \Omega$  heißt **Ereignis**. Tritt ein Ereignis bei einem Zufallsexperiment immer ein, dann heißt es **sicheres Ereignis**, tritt es nie ein, dann spricht man von einem **unmöglichen Ereignis**.

(ii) Eine einelementige Teilmenge  $\{\omega\} \subset \Omega$  heißt **Elementarereignis**.

**Beispiel:** Sei  $\Omega = \{1, \dots, 6\}$  der Grundraum der möglichen Ergebnisse beim Würfelwurf.

- (i)  $\{1\}, \{2\}, \dots, \{6\}$  Elementarereignisse
- (ii)  $W_g := \{2, 4, 6\}$  Würfeln einer geraden Zahl
- (iii)  $W_u := \{1, 3, 5\}$  Würfeln einer ungeraden Zahl
- (iv)  $\emptyset$  ist das unmögliche Ereignis,  $\Omega$  das sichere Ereignis

## Verknüpfungen von Ereignissen

Sind  $A$  und  $B$  Ereignisse, dann können durch Bildung von Schnitten

$$A \cap B := \{\omega \in \Omega : \omega \in A \text{ und } \omega \in B\},$$

Vereinigungen

$$A \cup B := \{\omega \in \Omega : \omega \in A \text{ oder } \omega \in B\},$$

Komplementbildung

$$A^c := \{\omega \in \Omega : \omega \notin A\}$$

und Differenzenbildung

$$A \setminus B := A \cap B^c = \{\omega \in \Omega : \omega \in A \text{ und } \omega \notin B\}$$

neue Ereignisse konstruiert werden.

**Beispiel:** Beim Würfelwurf sei  $A = \{4, 5, 6\}$  das Ereignis, dass ein Ergebnis  $\geq 4$  beobachtet wird. Dann

$$W_g \cap A = \{4, 6\} \quad W_g \cup A = \{2, 4, 5, 6\}$$

## Rechenregeln für Indikatorfunktionen

Für Indikatorfunktionen von verknüpften Ereignissen gelten folgende Rechenregeln.

**Bemerkung 2.3** Seien  $A, B$  Teilmengen von  $\Omega$  und  $\omega \in \Omega$ . Dann gilt

(i)  $1_{A \cap B}(\omega) = 1_A(\omega) \cdot 1_B(\omega)$

(ii)  $1_{A^c}(\omega) = 1 - 1_A(\omega)$

(iii)  $1_{A \cup B}(\omega) = \max(1_A(\omega), 1_B(\omega))$

(iv)  $A, B$  heißen disjunkt, falls  $A \cap B = \emptyset$ . Es gilt

$$A \cap B = \emptyset \implies 1_{A \cup B}(\omega) = 1_A(\omega) + 1_B(\omega)$$

**Beispiel:** Beim Würfelwurf gilt  $1_\Omega(\omega) = 1_{W_g}(\omega) + 1_{W_u}(\omega)$ .

## Wahrscheinlichkeit

Einem Ereignis  $A$  soll nun eine Zahl zugeordnet werden. Man spricht dann von der Wahrscheinlichkeit, dass dieses Ereignis eintritt. Zur Motivation betrachten wir zunächst Eigenschaften von relativen Häufigkeiten.

Gegeben sei eine Stichprobe  $x_1, \dots, x_n$ , die durch unabhängige Wiederholungen eines stochastischen Experiments gewonnen wurde. Die Menge der möglichen Ergebnisse eines jeden Einzelexperimentes sei  $\Omega$ .

Die **relative Häufigkeit**  $h_n(A)$  für das Eintreten des Ereignisses  $A$  ist gegeben durch

$$h_n(A) = \frac{\text{Anzahl der } x_i \text{ in } A}{n} = \frac{1}{n} \sum_{i=1}^n 1_A(x_i).$$

Offenbar besitzt die relative Häufigkeit folgende Eigenschaften:

$$0 \leq h_n(A) \leq 1 \text{ für jedes } A \subset \Omega,$$

$$h_n(\Omega) = 1,$$

$$A \cap B = \emptyset \implies h_n(A \cup B) = h_n(A) + h_n(B).$$

Die relative Häufigkeit ist eine zufallsbedingte Größe, die gewissen Schwankungen unterworfen ist. Die Erfahrung lehrt jedoch, dass mit zunehmendem Stichprobenumfang eine Stabilisierung eintritt.

**Beispiel:** Beim Würfelwurf kann man  $h_n(\{1\}) \approx 1/6$  oder  $h_n(\{W_g\}) \approx 1/2$  beobachten.

### Wahrscheinlichkeitsaxiome von Kolmogorov

Der Begriff der Wahrscheinlichkeit  $P(A)$  eines Ereignisses  $A$  kann jedoch nicht als Grenzwert einer relativen Häufigkeit definiert werden, derartige Versuche führten zu unüberbrückbaren Schwierigkeiten. Vielmehr begnügt man sich damit, grundlegende Eigenschaften (das Axiomensystem von Kolmogorov) zu fordern, die sich an den Eigenschaften von relativen Häufigkeiten orientieren.

**Definition 2.4** Ein **diskreter Wahrscheinlichkeitsraum** ist ein Paar  $(\Omega, P)$ , wobei  $\Omega \neq \emptyset$  eine endliche oder abzählbar unendliche Ergebnismenge ist und jedem Ereignis  $A$  eine reelle Zahl  $P(A)$ , die **Wahrscheinlichkeit von  $A$** , zugeordnet wird, so dass die folgenden Eigenschaften erfüllt sind.

(i)  $0 \leq P(A) \leq 1$  für  $A \subset \Omega$ ,

(ii)  $P(\Omega) = 1$ ,

(iii) Für paarweise disjunkte Ereignisse  $A_1, A_2, \dots$  gilt

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$$

**Bemerkung:** In der Definition 2.4 wird keine Aussage über den Wert  $P(A)$  in einem konkreten Zufallsexperiment getroffen.

Aus der Definition 2.4 lassen sich wesentliche Eigenschaften ableiten.

**Satz 2.5** Sei  $(\Omega, P)$  ein diskreter Wahrscheinlichkeitsraum.

(i) Für das unmögliche Ereignis  $\emptyset$  gilt:  $P(\emptyset) = 0$

(ii) Für das zu  $A$  komplementäre Ereignis  $A^c$  gilt:  $P(A^c) = 1 - P(A)$

(iii) Für endlich viele paarweise disjunkte Ereignisse  $A_1, \dots, A_k \subset \Omega$  gilt

$$P(A_1 \cup \dots \cup A_k) = P(A_1) + \dots + P(A_k)$$

### Laplace-Experimente

Ein Experiment mit endlicher Ergebnismenge  $\Omega = \{\omega_1, \dots, \omega_m\}$  mit  $m$  *gleich möglichen* Ergebnissen wird **Laplace-Wahrscheinlichkeitsraum** genannt. Einem Elementarereignis  $\{\omega_i\}$  wird dabei die Zahl

$$p(\omega_i) := P(\{\omega_i\}) := \frac{1}{m}$$

zugeordnet. Folglich erhält man wegen der Additivität 2.5 (iii)

$$P(A) = \sum_{\omega \in A} p(\omega) = \#A \cdot \frac{1}{m} = \frac{\#A}{m}.$$

Dabei werden auch  $\#A$  die Anzahl der für das Ereignis  $A$  günstigen Fälle und  $m$  die Anzahl der insgesamt möglichen Fälle genannt.

### Beispiele:

- (i) (Einfacher Würfelwurf)  $\Omega = \{1, \dots, 6\}$ ,  $p(\omega_i) = 1/6$ . Also

$$P(W_g) = \frac{\#W_g}{6} = \frac{1}{2} = \frac{\#W_u}{6} = P(W_u)$$

- (ii) In einer Urne befinden sich 40 weiße und 20 schwarze Kugeln. Eine Kugel wird zufällig aus der Urne gezogen. Für das Ereignis

$A$  : Ziehung einer schwarzen Kugel

gilt dann

$$P(A) = \frac{\#A}{40 + 20} = \frac{20}{60} = \frac{1}{3}.$$

- (iii) (Ziehen ohne Zurücklegen) Nun werden 2 Kugeln aus der Urne gezogen ohne die erste Kugel zurückzulegen. Beim zweiten Zug sind demnach noch  $60 - 1 = 59$  Kugeln in der Urne. Nummeriert man die weißen Kugeln von 1 bis 40 und die schwarzen Kugeln von 41 bis 60, dann sind die gleich möglichen Ergebnisse von der Form  $(i, j)$ , wobei  $1 \leq i < j \leq 60$ , da die Reihenfolge keine Rolle spielt. Für das Ereignis

$B$  : Ziehung von zwei schwarzen Kugeln

erhält man

$$\#B = \frac{20 \cdot 19}{2} = 190, \quad \#\Omega = \frac{60 \cdot 59}{2} = 1770$$

und damit

$$P(B) = \frac{\#B}{\#\Omega} = \frac{190}{1770} \approx 0.1073.$$

In diesem Beispiel kann jedes mögliche Ergebnis als 2-elementige Teilmenge von  $\{1, \dots, 60\}$  aufgefasst werden.

- (iv) (Ziehen mit Zurücklegen) Wird vor dem zweiten Zug die erste Kugel zurück in die Urne gelegt (vgl. (iii)) und die Urne gut durchmischt, dann lauten die gleich wahrscheinlichen Ergebnisse  $(i, j)$ , wobei  $(i, j) \in \Omega := \{1, \dots, 60\} \times \{1, \dots, 60\}$ . Für

$C$  : Ziehung von zwei schwarzen Kugeln

gilt  $C = \{41, \dots, 60\} \times \{41, \dots, 60\}$  und damit

$$P(C) = \frac{\#C}{\#\Omega} = \frac{20 \cdot 20}{60 \cdot 60} = \frac{1}{9} \approx 0.1111.$$

### Allgemeine diskrete Wahrscheinlichkeitsräume

Für die Zuordnung von Wahrscheinlichkeiten in Laplace-Wahrscheinlichkeitsräumen genügt es, die Wahrscheinlichkeiten für die Elementarereignisse zu kennen. Das gilt allgemein für diskrete Wahrscheinlichkeitsräume.

Ist  $\Omega = \{\omega_1, \omega_2, \dots\}$  ein diskreter Wahrscheinlichkeitsraum, dann gilt für jedes Ereignis  $A \subset \Omega$

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

Definieren wir  $p_i := p(\omega_i)$ , dann folgt  $p_i \geq 0$  und

$$1 = P(\Omega) = \sum_{\omega \in \Omega} p(\omega) = p_1 + p_2 + p_3 + \dots$$

Offenbar sind die Wahrscheinlichkeiten  $P(A)$  durch eine Folge von reellen Zahlen  $p_i$  mit

$$p_i \geq 0, \quad i = 1, 2, 3, \dots$$

und

$$\sum_{i=1}^{\infty} p_i = 1$$

eindeutig bestimmt.

Das gilt jedoch nur für diskrete Wahrscheinlichkeitsräume und nicht für stetige Wahrscheinlichkeitsräume.

**Beispiel:** Eine faire Münze wird solange geworfen, bis zum erstmals *Kopf* erscheint. Der Grundraum  $\Omega$  der möglichen Ergebnisse ist dann

$$\Omega = \{1, 2, 3, \dots\}$$

Wir betrachten das Elementarereignis

$\{k\}$  : Beim  $k$ -ten Wurf erscheint erstmals *Kopf*

Offenbar kann die Wahrscheinlichkeit  $p_k$  für das Ereignis  $\{k\}$  auch wie folgt berechnet werden.

Sei

$$\Omega_k = \{0, 1\} \times \dots \times \{0, 1\} = \{0, 1\}^k$$

der Ergebnisraum zu einem  $k$ -fachen Münzwurf (1 steht für *Kopf*). Da alle möglichen Ergebnisse gleich wahrscheinlich sind, liegt ein Laplace-Experiment vor. Es gilt also

$$P_k(\{a\}) = \frac{1}{\#\Omega_k} = \frac{1}{2^k}, \quad a \in \Omega_k.$$

Dabei bezeichnet  $P_k(A)$  die Wahrscheinlichkeit für ein Ereignis  $A \subset \Omega_k$ . Das Ereignis, dass genau im letzten Wurf *Kopf* erscheint, wird beschrieben durch

$$a_k = (0, \dots, 0, 1)$$

Also gilt

$$p_k = P_k(\{a_k\}) = \frac{1}{2^k}.$$

Man beachte, dass  $p_k \geq 0$  und

$$\sum_{k=1}^{\infty} \frac{1}{2^k} = \frac{1}{2} \left( 1 + \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \dots \right) = \frac{1}{2} \cdot \frac{1}{1 - 1/2} = 1$$

mit Hilfe der Formel für die geometrische Reihe. Also wird durch die Zahlen  $p_k$ ,  $k = 1, 2, 3, \dots$  ein diskreter Wahrscheinlichkeitsraum auf  $\Omega$  definiert, die geometrische Verteilung.

### Additionssatz

Für disjunkte Ereignisse  $A$  und  $B$  gilt  $P(A \cup B) = P(A) + P(B)$ . Allgemein gilt die folgende Aussage.

**Satz 2.6** Für beliebige Teilmengen  $A$  und  $B$  von  $\Omega$  gilt

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

**Beispiel:** Wie groß ist die Wahrscheinlichkeit, beim 2-fachen Würfelwurf mindestens eine „6“ zu werfen? Betrachte die Ereignisse

$A$ : Augenzahl „6“ beim ersten Wurf,

$B$ : Augenzahl „6“ beim zweiten Wurf.

Durch Verknüpfungen erhält man die Ereignisse

$A \cup B$ : In mindestens einem der beiden Würfe wird eine „6“ beobachtet,

$A \cap B$ : Augenzahl „6“ in jedem Wurf.

Da die möglichen Ergebnisse  $(i, j)$ ,  $i, j \in \{1, \dots, 6\}$ , des 2-fachen Wurfes gleich wahrscheinlich sind, liegt ein Laplace-Experiment vor. Also gilt mit Satz 2.6

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{\#A + \#B - \#(A \cap B)}{36} = \frac{6 + 6 - 1}{36} = \frac{11}{36}.$$

### 3 Diskrete Zufallsvariable

In Abschnitt 1 haben wir den Begriff des Merkmals eingeführt. Wählt man z.B. zufällig einen Studierenden aus der Gesamtheit aller Studierenden aus und notiert das Merkmal Fachsemester, dann erhält man einen zufälligen Wert. Diese Vorgehensweise wird in einem mathematischen Modell durch den Begriff der Zufallsvariablen beschrieben.

**Definition 3.1** Sei  $\Omega$  ein Grundraum. Eine Abbildung

$$X : \Omega \rightarrow \mathbb{R}$$

heißt **Zufallsvariable** (auf  $\Omega$ ).

**Bemerkung:**

- (i) Eine Zufallsvariable  $X$  ordnet jedem möglichen Ergebnis eines Zufallsexperimentes eine reelle Zahl  $X(\omega)$  zu.
- (ii) Falls  $X$  nur endlich viele oder abzählbar unendlich viele verschiedene reelle Zahlen annehmen kann, dann nennen wir  $X$  **diskret**.

**Beispiel 3.2** (i) Beim zweifachen Würfelwurf lautet die Ergebnismenge

$$\Omega := \{(i, j) : i, j \in \{1, \dots, 6\}\},$$

wobei  $i$  das Ergebnis des ersten Wurfes und  $j$  das Ergebnis des zweiten Wurfes beschreibt.

Definiert man für  $\omega = (i, j) \in \Omega$

$$X(i, j) := i + j,$$

dann beschreibt die Zufallsvariable  $X$  die Augensumme aus den beiden Würfeln. Offenbar kann  $X$  nur die Werte aus der Menge  $\{2, \dots, 12\}$  annehmen. Betrachten wir nun das Ereignis

$A_5$  : die Augensumme ist gleich 5

Offensichtlich tritt  $A_5$  genau dann ein, wenn „ $X$  den Wert 5 annimmt“. Wir schreiben dafür

$$\{X = 5\} := \{(i, j) \in \Omega : X(i, j) = 5\} = A_5$$

Da ein Laplace-Experiment vorliegt, gilt

$$P(\{X = 5\}) = P(A_5) = \frac{\#A}{\#\Omega} = \frac{\{(1, 4), (2, 3), (3, 2), (4, 1)\}}{36} = \frac{1}{9}$$

- (ii) Ebenso könnte man die Anzahl der geworfenen Einsen beschreiben durch

$$Y(i, j) := 1_{\{1\}}(i) + 1_{\{1\}}(j).$$

Dabei gilt

$k$	0	1	2
$\#\{(i, j) : Y(i, j) = k\}$	25	10	1
$P(\{Y = k\})$	25/36	10/36	1/36



## Wahrscheinlichkeitsverteilung einer diskreten Zufallsvariablen

Sei  $X : \Omega \rightarrow \mathbb{R}$  eine diskrete Zufallsvariable über und  $(\Omega, P)$  ein diskreter Wahrscheinlichkeitsraum.

Allgemein schreibt man abkürzend

$$\{X = x\} := \{\omega \in \Omega : X(\omega) = x\}$$

für das Ereignis, dass  $X$  den Wert  $x$  annimmt,

$$\{X \leq x\} := \{\omega \in \Omega : X(\omega) \leq x\}$$

für das Ereignis, dass  $X$  einen Wert  $\leq x$  annimmt, und für die entsprechenden Wahrscheinlichkeiten

$$P\{X = x\} := P(\{\omega \in \Omega : X(\omega) = x\}),$$

$$P\{X \leq x\} := P(\{\omega \in \Omega : X(\omega) \leq x\}).$$

Die **Wahrscheinlichkeitsverteilung (kurz: Verteilung)** einer diskreten Zufallsvariablen kann durch ihre **Zähldichte**

$$f(x) := P\{X = x\}$$

beschrieben werden. Entsprechend schreiben wir auch  $f_X$  um den Bezug zu  $X$  zu kennzeichnen. Falls  $X$  nur Werte  $a_1, a_2, \dots$  annehmen kann, dann gilt  $f(x) = 0$  für alle  $x \notin \{a_1, a_2, \dots\}$  und es reicht, nur die Werte

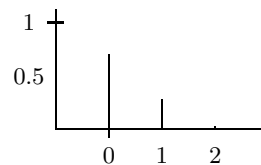
$$f(a_i) := P\{X = a_i\}, \quad i = 1, 2, \dots$$

anzugeben, z.B. in einer Tabelle der Form

$x$	$a_1$	$a_2$	$\dots$
$f(x)$	$f(a_1)$	$f(a_2)$	$\dots$

oder grafisch als Stabdiagramm.

So erhält man für die Zufallsvariable  $Y$  aus Beispiel 3.2 das Stabdiagramm



Die Zähldichte ist normiert, d.h.

$$\sum_i f(a_i) = f(a_1) + f(a_2) + \dots = 1$$

Die Verteilung einer diskreten Zufallsvariablen kann eindeutig durch ihre Verteilungsfunktion beschrieben werden.

**Definition 3.3** Sei  $X$  eine diskrete Zufallsvariable mit Zähldichte  $f$ .

Die Funktion  $F : \mathbb{R} \rightarrow [0, 1]$  mit

$$F(x) := P\{X \leq x\} = \sum_{a_i \leq x} f(a_i)$$

heißt Verteilungsfunktion  $X$ .

**Beispiel:** Die Verteilungsfunktion von  $Y$  aus Beispiel 3.2 lautet

$$F(x) = \begin{cases} 0, & x < 0 \\ 25/36, & 0 \leq x < 1 \\ 35/36, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

**Bemerkung 3.4** Seien  $f$  die Zähldichte und  $F$  die Verteilungsfunktion von  $X$ .

(i) Für die Wahrscheinlichkeit, dass  $X$  einen Wert in dem Intervall  $(a, b]$  annimmt, gilt

$$P\{a < X \leq b\} = \sum_{a < a_i \leq b} f(a_i) = F(b) - F(a).$$

Entsprechend

$$P\{a < X < b\} = \sum_{a < a_i < b} f(a_i), \quad P\{a \leq X < b\} = \sum_{a \leq a_i < b} f(a_i) \quad \text{etc.}$$

(ii)  $F$  ist eine monoton wachsende Treppenfunktion mit  $0 \leq F(x) \leq 1$ .

(iii) Es gilt

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{und} \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

## 4 Stetige Zufallsvariable

**Bemerkung 4.1** Die Verteilung einer stetigen Zufallsvariablen wird durch eine stetige und stückweise differenzierbare **Verteilungsfunktion**  $F$  vollständig beschrieben, welche in Integralform angegeben werden kann. Dabei bezeichnet  $F(x)$  die Wahrscheinlichkeit, dass die Zufallsvariable  $X$  einen Wert  $\leq x$  annimmt. Wir schreiben

$$P\{X \leq x\} = F(x) = \int_{-\infty}^x f(t) dt$$

und sprechen von einer **stetigen Zufallsvariablen** oder genauer **stetig verteilten Zufallsvariablen**  $X$  ohne auf den konkreten Grundraum  $\Omega$  etc. einzugehen. Die Funktion  $f$  heißt **Wahrscheinlichkeitsdichte (kurz: Dichte)** von  $X$ . Die Dichte  $f$  und die Verteilungsfunktion  $F$  besitzen folgende Eigenschaften:

(i)  $f(x) \geq 0$  für alle  $x \in \mathbb{R}$ .

(ii)  $f$  ist normiert, d.h. es gilt

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

(iii) Die Verteilungsfunktion  $F$  ist monoton wachsend, stetig und stückweise differenzierbar mit  $0 \leq F(x) \leq 1$ .  $F$  ist eine Stammfunktion von  $f$ . Es gilt  $F'(x) = f(x)$  bis auf endlich viele  $x$ .

(iv) Es gilt

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{und} \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

(v) Für die Wahrscheinlichkeit, dass  $X$  einen Wert  $x$  annimmt, gilt

$$P\{X = x\} = 0$$

(vi) Für die Wahrscheinlichkeit, dass  $X$  einen Wert in einem Intervall annimmt, gilt

$$P\{a < X \leq b\} = \int_a^b f(x) dx = F(b) - F(a)$$

und

$$P\{a < X < b\} = P\{a \leq X \leq b\} = \int_a^b f(x) dx = F(b) - F(a).$$

**Beispiel 4.2** (i) Eine Zufallsvariable mit der Dichte

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{falls } a \leq x \leq b \\ 0, & \text{sonst} \end{cases}$$

heißt gleichverteilt auf  $[a, b]$ . Für die Verteilungsfunktion gilt

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & \text{falls } a \leq x \leq b \\ 1, & x > b \end{cases}$$

Für jedes Intervall  $[c, c+d]$  der Länge  $d$  mit  $a \leq c \leq c+d \leq b$  gilt

$$P\{c \leq X \leq c+d\} = \int_c^{c+d} \frac{1}{b-a} dx = \frac{d}{b-a}.$$

- (ii) Die Lebensdauer eines elektronischen Bauteils werde durch eine exponentialverteilte Zufallsvariable  $X$  mit der Dichte

$$f(x) = \begin{cases} 0, & \text{falls } x < 0 \\ \lambda \cdot e^{-\lambda x}, & \text{falls } x \geq 0 \end{cases}$$

beschrieben. Dann gilt

$$F(x) = \int_{-\infty}^x f(t) dt = 0, \quad x \leq 0,$$

und

$$F(x) = \int_{-\infty}^x f(t) dt = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}, \quad x > 0.$$

Es gilt

$$P\{X > t\} = 1 - P\{X \leq t\} = 1 - F(t) = e^{-\lambda t}, \quad t \geq 0.$$

## 5 Kennziffern von Verteilungen

### Erwartungswert, Varianz und Standardabweichung für diskrete Zufallsvariable

Wir betrachten zunächst folgendes Spiel. Bei einem Würfelspiel wird ein Gewinn von einer Geldeinheit (GE) ausgezahlt, wenn eine „4“ gewürfelt wird. Wird eine „5“ oder eine „6“ gewürfelt, dann wird ein Gewinn von 3 GE ausgezahlt. Wie hoch sollte der Einsatz sein, um dieses Spiel als fair zu bezeichnen. Beschreibt die Zufallsvariable  $X$  den Gewinn, dann gilt

$$P\{X = 0\} = \frac{3}{6}, \quad P\{X = 1\} = \frac{1}{6}, \quad P\{X = 3\} = \frac{2}{6},$$

Andere Gewinne sind nicht möglich. Im „Mittel“ wird man bei diesem Spiel

$$1 \cdot \frac{1}{6} + 3 \cdot \frac{2}{6} = \frac{7}{6}$$

Geldeinheiten gewinnen, sodass ein Einsatz in dieser Höhe als fair bezeichnet werden kann. Dabei wurde der Wert

$$a_1 \cdot P\{X = a_1\} + a_2 \cdot P\{X = a_2\} + a_3 \cdot P\{X = a_3\}$$

berechnet, wobei  $a_1 = 0, a_2 = 1, a_3 = 3$  die möglichen Werte von  $X$  bezeichnen.

**Definition 5.1** Sei  $X$  eine diskrete Zufallsvariable, die nur die Werte  $a_1, a_2, \dots$  annehmen kann, und  $f$  die Zähldichte von  $X$ . Dann heißt

$$E(X) := \sum_i a_i f(a_i) \quad \text{Erwartungswert von } X,$$

$$\text{Var}(X) := \sum_i (a_i - E(X))^2 f(a_i) \quad \text{Varianz von } X,$$

$$\sigma = \sqrt{\text{Var}(X)} \quad \text{Standardabweichung von } X.$$

#### Bemerkung:

- (i) Es wird vorausgesetzt, dass

$$\sum_i |a_i| f(a_i) < \infty \quad \text{bzw.} \quad \sum_i (a_i - E(X))^2 f(a_i) < \infty.$$

Andernfalls besitzt die Verteilung von  $X$  keinen Erwartungswert bzw. keine Varianz.

- (ii) Der Erwartungswert kann physikalisch als Schwerpunkt interpretiert werden, wenn man  $f(a_i) = P\{X = a_i\}$  als Massepunkt an der Stelle  $a_i$  der ansonsten gewichtslosen reellen Zahlengeraden deutet.

#### Beispiel:

- (i) Sei  $X$  eine diskrete Zufallsvariable mit  $P\{X = 0\} = 3/6, P\{X = 1\} = 1/6$  und  $P\{X = 3\} = 2/6$ . Dann gilt für die Varianz  $\text{Var}(X) \approx 1.81$  und für die Standardabweichung  $\sigma \approx 1.34$

(ii) (Warten auf den ersten Treffer)

Sie nehmen jede Woche an einem Glücksspiel teil, bei dem Sie mit Wahrscheinlichkeit  $p$  gewinnen, wobei  $p \in (0, 1)$ . Wie lange warten Sie im Mittel bis zum ersten Gewinn?

Es bezeichne  $X$  die Anzahl der verlorenen Spiele vor dem ersten Gewinn. Dann bedeutet das Ereignis  $\{X = k\}$ , dass die ersten  $k$ -Spiele verloren sind und Gewinn genau im  $k + 1$ -ten Spiel eintritt. Analog zum Münzwurfbeispiel folgt

$$P\{X = k\} = (1 - p)^k \cdot p, \quad k = 0, 1, 2, \dots$$

Die Zufallsvariable  $X$  heißt **geometrisch verteilt zum Parameter  $p$** . Es gilt

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

Der Erwartungswert berechnet sich wie folgt.

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k(1 - p)^k p = p(1 - p) \sum_{k=1}^{\infty} k(1 - p)^{k-1} \\ &= p(1 - p) \left( -\frac{1}{1 - (1 - p)} \right)' = p(1 - p) \left( -\frac{1}{p} \right)' \\ &= \frac{1 - p}{p}. \end{aligned}$$

Für  $p = 1/2$  ist  $E(X) = 1$ , d.h. im Mittel ein verlorenes Spiel bzw. den ersten Gewinn im 2. Spiel. Bei einer Gewinnchance von  $p = 1/10$  gilt  $E(X) = (9/10)/(1/10) = 9$ , also gewinnt man im Mittel jedes 10. Spiel.

**Bemerkung 5.2** Sei  $X : \Omega \rightarrow \mathbb{R}$  eine diskrete Zufallsvariable. Dann gilt

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

falls  $E(X)$  existiert oder  $\sum_{\omega \in \Omega} |X(\omega)|P(\{\omega\}) < \infty$  gilt.

**Beispiel:** Sei  $\Omega = \{1, \dots, 6\}$ ,  $P(\{\omega\}) = 1/6$  und  $X(\omega) = 0$  für  $\omega = 1, 2, 3$ ,  $X(\omega) = 1$  für  $\omega = 4$  und  $X(\omega) = 3$  für  $\omega = 5, 6$ . Dann

$$\sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) = 1 \cdot P(\{4\}) + 3 \cdot P(\{5\}) + 3 \cdot P(\{6\}) = \frac{7}{6} = E(X).$$

**Erwartungswert, Varianz und Standardabweichung für stetige Zufallsvariable**

Für stetige Zufallsvariablen werden folgende Kennziffern definiert:

**Definition 5.3** Sei  $X$  eine stetige Zufallsvariable mit Dichte  $f$  Dann heißt

$$E(X) := \int_{-\infty}^{\infty} x \cdot f(x) dx \quad \text{Erwartungswert von } X,$$

$$Var(X) := \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x) dx \quad \text{Varianz von } X,$$

$$\sigma = \sqrt{Var(X)} \quad \text{Standardabweichung von } X.$$

**Bemerkung:** Es wird vorausgesetzt, dass

$$\int_{-\infty}^{\infty} |x| \cdot f(x) dx < \infty \quad \text{bzw.} \quad \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x) dx < \infty.$$

Andernfalls besitzt die Verteilung von  $X$  keinen Erwartungswert bzw. keine Varianz.

**Beispiel:** Die Cauchy-Verteilung mit Dichte  $g(x) = \pi^{-1}(1+x^2)^{-1}$  besitzt keinen Erwartungswert. Für die Gleichverteilung auf  $[a, b]$  erhält man

$$\int_{-\infty}^{\infty} x \cdot f(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{a+b}{2}.$$

### Rechenregeln für Erwartungswert und Varianz

**Satz 5.4** (Transformationsformel)

Sei  $X$  eine Zufallsvariable und  $g: \mathbb{R} \rightarrow \mathbb{R}$  eine Abbildung. Dann gilt für die Zufallsvariable  $Z := g(X)$ :

$$E(Z) := E(g(X)) = \sum_i g(a_i) \cdot P\{X = a_i\} = \sum_{\omega \in \Omega} g(X(\omega)) \cdot P\{\omega\},$$

falls  $X$  diskret ist bzw.

$$E(Z) = E(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f(x) dx,$$

falls  $X$  eine stetige Zufallsvariable mit Dichte  $f$  ist.

Es gelten folgende Rechenregeln:

**Satz 5.5** Seien  $X$  und  $Y$  diskrete oder stetige Zufallsvariable und  $a, b$  reelle Zahlen. Dann

- (i)  $E(1_A(X)) = P\{X \in A\}$  für Ereignisse  $A$ ,
- (ii)  $E(aX + b) = aE(X) + b$ ,
- (iii)  $E(X + Y) = E(X) + E(Y)$ ,
- (iv)  $Var(X) = E\left(\left(X - E(X)\right)^2\right) = E(X^2) - (E(X))^2$ .
- (v)  $Var(aX + b) = a^2 Var(X)$ .

**Bemerkung:** Für die **standardisierte Zufallsvariable**

$$Z := \frac{X - E(X)}{\sqrt{Var(X)}}$$

gilt  $E(Z) = 0$  und  $Var(Z) = 1$ .

**Bemerkung 5.6** Eine Zufallsvariable  $X$  heißt symmetrisch verteilt mit Symmetriezentrum  $\mu$ , falls

$$P\{X - \mu \leq -t\} = P\{X - \mu \geq t\}.$$

In diesem Fall gilt  $E(X) = \mu$ , falls  $E(X)$  existiert.

**Beispiel:**

- (i) Ist  $X$  Laplace-verteilt auf  $\{m, \dots, n\}$ , dann ist  $\mu = (n+m)/2$  das Symmetriezentrum.
- (ii) Besitzt  $X$  die Dichte  $f(x) = 1 - |x|$ ,  $-1 \leq x \leq 1$  und  $f(x) = 0$  sonst, dann ist  $\mu = 0$ .

## 6 Unabhängigkeit

Werden  $n$  Zufallsexperimente getrennt voneinander und ohne gegenseitige Beeinflussung ausgeführt, so sprechen wir von unabhängigen Zufallsexperimenten. In diesem Fall heißen die Zufallsvariablen  $X_1, \dots, X_n$ , die die Einzelexperimente beschreiben, unabhängig.

Zur Illustration betrachten wir den zweifachen Würfelwurf mit dem Grundraum  $\Omega = \{(i, j) : i, j \in \{1, \dots, 6\}\}$  und den Zufallsvariablen

$$X_1(i, j) := i \quad \text{und} \quad X_2(i, j) = j,$$

die die Ergebnisse des ersten bzw. zweiten Wurfes beschreiben. Wir nehmen an, dass die Würfe sich nicht gegenseitig beeinflussen. Dann liegt ein Laplace-Experiment auf  $\Omega$  vor und es gilt

$$\begin{aligned} P\{X_1 \leq a, X_2 \leq b\} &:= P(\{(i, j) \in \Omega : X_1(i, j) \leq a, X_2(i, j) \leq b\}) \\ &= \frac{\#\{(i, j) \in \Omega : i \leq a, j \leq b\}}{\#\Omega} \\ &= \frac{\#\{i \in \{1, \dots, 6\} : i \leq a\} \cdot \#\{j \in \{1, \dots, 6\} : j \leq b\}}{6 \cdot 6} \\ &= P\{X_1 \leq a\} \cdot P\{X_2 \leq b\} \end{aligned}$$

Für die Definition der Unabhängigkeit von  $n$  Zufallsvariablen verwenden wir Ausdrücke der Form  $P\{X_1 \leq x_1, \dots, X_n \leq x_n\}$ . Für diskrete Zufallsvariable setzen wir

$$P\{X_1 \leq x_1, \dots, X_n \leq x_n\} := P\{\omega \in \Omega : X_1(\omega) \leq x_1, \dots, X_n(\omega) \leq x_n\}$$

Für stetige Zufallsvariable wird dieser Ausdruck zunächst nur intuitiv benutzt.

**Definition 6.1** Seien  $X_1, \dots, X_n$  diskrete oder stetige Zufallsvariable.  $X_1, \dots, X_n$  heißen (**stochastisch**) **unabhängig**, falls für alle reelle Zahlen  $x_1, \dots, x_n$  gilt

$$P\{X_1 \leq x_1, \dots, X_n \leq x_n\} = P\{X_1 \leq x_1\} \cdot \dots \cdot P\{X_n \leq x_n\}.$$

**Bemerkung 6.2** Für diskrete Zufallsvariable reicht es zum Nachweis der Unabhängigkeit zu zeigen, dass gilt

$$P\{X_1 = x_1, \dots, X_n = x_n\} = P\{X_1 = x_1\} \cdot \dots \cdot P\{X_n = x_n\}$$

für alle möglichen Werte  $x_i$  von  $X_i$ .

**Rechenregeln für die Varianz von Summen von unabhängigen Zufallsvariablen**

**Satz 6.3** Sind  $X_1, \dots, X_n$  unabhängige Zufallsvariable, dann gilt

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$



## 7 Wahrscheinlichkeitsverteilungen

In diesem Abschnitt sollen wichtige Verteilungen vorgestellt werden. Dabei sei  $X$  eine Zufallsvariable und  $T_X$  die Menge der möglichen Werte, die  $X$  annehmen kann.  $T_X$  heißt der Träger von  $X$ .

Die Verteilung  $Q$  einer Zufallsvariablen  $X$  ordnet jedem Ereignis  $A$  die Wahrscheinlichkeit zu, dass  $X$  einen Wert in  $A$  annimmt. Wir schreiben dann

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

Man sagt:  $X$  ist verteilt nach  $Q$  [i.Z.  $X \sim Q$ ].

Besitzt  $X$  zudem die Verteilungsfunktion  $F$ , dann gilt

$$Q((-\infty, t]) = P\{X \leq t\} = F(t)$$

und

$$Q((a, b]) = P\{a < X \leq b\} = F(b) - F(a).$$

Ist  $X$  eine diskrete Zufallsvariable mit Verteilung  $Q$ , dann

$$Q\{k\} := Q(\{k\}) = f_X(k) = P\{X = k\}.$$

In diesem Fall ist  $(T_X, Q)$  ein Wahrscheinlichkeitsraum.

**Diskrete Gleichverteilung**  $G(a_1, \dots, a_m)$

- **Zähldichte:**  $P\{X = a_i\} = 1/m$ ,  $a_1 < \dots < a_m$ .
- **Träger:**  $T_X = \{a_1, \dots, a_m\}$ .
- **Erwartungswert:**  $E(X) = \frac{1}{m} \sum_{i=1}^m a_i = \bar{a}$
- **Varianz:**  $Var(X) = \frac{1}{m} \sum_{i=1}^m (a_i - \bar{a})^2$

**Beispiel:**  $G(1, 2, 3, 4)\{3\} = 1/4$ . Ist  $X \sim G(1, 2, 3, 4)$ , dann

$$E(X) = \frac{1 + 2 + 3 + 4}{4} = 2.5,$$

$$Var(X) = \frac{(1 - 2.5)^2 + (2 - 2.5)^2 + (3 - 2.5)^2 + (4 - 2.5)^2}{4} = 1.25.$$

**Bernoulli-Verteilung**  $B_{(1,p)}$

Das Ergebnis eines Versuches, bei dem mit Wahrscheinlichkeit  $p$  das Ereignis  $A$  (Erfolg) und mit Wahrscheinlichkeit  $1 - p$  das komplementäre Ereignis  $A^c$  (kein Erfolg) eintritt, kann mit einer Bernoulli-verteilten Zufallsvariablen modelliert werden.

- **Zähldichte:**  $P\{X = 0\} = 1 - p$  und  $P\{X = 1\} = p$ .
- **Träger:**  $T_X = \{0, 1\}$ .
- **Erwartungswert:**  $E(X) = p$
- **Varianz:**  $Var(X) = p(1 - p)$

## Binomialverteilung $B_{(n,p)}$

Wir führen einen Versuch, bei dem mit Wahrscheinlichkeit  $p$  das Ereignis  $A$  (Erfolg) und mit Wahrscheinlichkeit  $1 - p$  das komplementäre Ereignis  $A^c$  (kein Erfolg) eintritt,  $n$  mal unabhängig voneinander durch und betrachten die Anzahl der Erfolge  $X$ .

- **Zähldichte:**  $P\{X = k\} = \binom{n}{k} p^k (1 - p)^{n-k}$ ,  $k = 0, 1, \dots, n$ .
- **Träger:**  $T_X = \{0, \dots, n\}$ .
- **Erwartungswert:**  $E(X) = np$
- **Varianz:**  $Var(X) = np(1 - p)$
- **Anwendung:** Gut-Schlecht-Prüfung, Ziehen mit Zurücklegen

Dabei ist der Binomialkoeffizient  $\binom{n}{k}$  definiert durch

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

wobei

$$0! := 1, \quad \text{und} \quad n! := 1 \cdot \dots \cdot n, \quad n = 1, 2, \dots$$

Werden genau  $k$  Erfolge und  $n - k$  Misserfolge beobachtet, dann gibt es  $\binom{n}{k}$  Möglichkeiten, wie diese Erfolge und Misserfolge auftreten. Jede Möglichkeit tritt dabei mit Wahrscheinlichkeit  $p^k (1 - p)^{n-k}$  auf.

Sind  $X_1, \dots, X_n$  unabhängige,  $B_{(1,p)}$ -verteilte Zufallsvariable, dann ist  $\sum_{i=1}^n X_i$  eine  $B_{(n,p)}$ -verteilte Zufallsvariable.

**Beispiel:** Zur Qualitätskontrolle werden 10 Bauteile, die unabhängig voneinander jeweils mit Wahrscheinlichkeit 0.2 defekt sind, überprüft. Wie groß ist die Wahrscheinlichkeit, dass 4 Bauteile defekt sind?

Sei  $X$  eine  $B_{(10,0.2)}$ -verteilte Zufallsvariable. Dann gilt

$$\begin{aligned} B_{(10,0.2)}\{4\} &= \binom{10}{4} \cdot 0.2^4 \cdot (1 - 0.2)^{10-4} \\ &= \frac{10!}{4!(10-4)!} \cdot 0.2^4 \cdot (1 - 0.2)^{10-4} \\ &= \frac{1 \cdot 2 \cdot \dots \cdot 10}{1 \cdot 2 \cdot \dots \cdot 4 \cdot 1 \cdot 2 \cdot \dots \cdot 6} \cdot 0.2^4 \cdot (0.8)^6 \approx 0.088 \end{aligned}$$

## Hypergeometrische Verteilung $hyp(n, r, s)$

Aus einer Urne mit  $r$  roten und  $s$  schwarzen Kugeln werden  $n$  Kugeln ohne Zurücklegen gezogen. Die Hypergeometrische Verteilung beschreibt die Wahrscheinlichkeit, dass sich unter den  $n$  gezogenen Kugeln genau  $k$  rote Kugeln befinden.

- **Zähldichte:**  $P\{X = k\} = \frac{\binom{r}{k} \binom{s}{n-k}}{\binom{r+s}{n}}$ ,  $0 \leq k \leq r$ ,  $n - k \leq s$   
und  $P\{X = k\} = 0$  sonst
- **Träger:**  $T_X = \{\max\{0, n - s\}, \dots, \min\{r, n\}\}$ .
- **Erwartungswert:**  $E(X) = n \frac{r}{r+s}$
- **Anwendung:** Gut-Schlecht-Prüfung, Ziehen ohne Zurücklegen

### Geometrische Verteilung $Geo(p)$

- **Zähldichte:**  $P\{X = k\} = p(1 - p)^k$ ,  $p \in (0, 1)$
- **Träger:**  $T_X = \{0, 1, \dots\}$
- **Erwartungswert:**  $E(X) = \frac{1-p}{p}$
- **Varianz:**  $Var(X) = \frac{1-p}{p^2}$
- **Anwendung:** Wartezeit bei Münzwurfexperimenten

### Poisson-Verteilung $Po(\lambda)$

- **Zähldichte:**  $P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$ ,  $\lambda > 0$
- **Träger:**  $T_X = \{0, 1, \dots\}$
- **Erwartungswert:**  $E(X) = \lambda$
- **Varianz:**  $Var(X) = \lambda$
- **Anwendung:** Seltene Ereignisse

Sind  $X$  und  $Y$  unabhängige Zufallsvariable mit den Poisson-Verteilungen  $X \sim Po(\lambda_X)$  und  $Y \sim Po(\lambda_Y)$ , dann gilt  $X + Y \sim Po(\lambda_X + \lambda_Y)$ .

### Poisson-Approximation der Binomialverteilung

Zur Modellierung seltener Ereignisse wird die Poisson-Verteilung verwendet. Eine Begründung liefert der folgende Satz.

**Satz 7.1** Seien  $p_n \in (0, 1)$ ,  $n = 1, 2, \dots$  mit

$$\lim_{n \rightarrow \infty} np_n = \lambda \in (0, \infty).$$

Dann gilt

$$\lim_{n \rightarrow \infty} B_{(n, p_n)}\{k\} = Po(\lambda)\{k\}, \quad \text{und} \quad \lim_{n \rightarrow \infty} B_{(n, p_n)}\{0, \dots, k\} = Po(\lambda)\{0, \dots, k\}$$

für alle  $k \in \mathbb{N}_0$ .

**Bemerkung:** Für „großen“ Stichprobenumfang  $n$  und  $p$  „klein“ kann die Binomialverteilung durch eine Poisson-Verteilung approximiert werden

**Beispiel:** Ein radioaktives Material bestehe aus  $n = 10^8$  Teilchen, die alle unabhängig voneinander mit Wahrscheinlichkeit  $p = 10^{-6}$  in einem vorgegebenen Zeitraum zerfallen.

Da  $p$  klein ist, kann die Verteilung der Anzahl  $X$  der Zerfälle durch eine Poisson-Verteilung approximiert werden. Es gilt mit  $n \cdot p = 100$

$$P\{X = k\} = B_{(10^8, 10^{-6})}\{k\} \approx Po(100)\{k\}.$$

**Zur Erinnerung:** Stetige Zufallsvariable werden durch eine Verteilungsfunktion  $F$  bzw. eine Wahrscheinlichkeitsdichte  $f$  beschrieben. Es gilt

$$P\{X \leq t\} = F(t) = \int_{-\infty}^t f(x) dx.$$

Als **Träger**  $T_X$  einer stetigen Zufallsvariablen wird die Menge der Punkte bezeichnet, für die die Dichte  $f$  positiv ist.

**Stetige Gleichverteilung (Rechteckvert.)**  $U[a, b]$  mit den Parametern  $a < b$

- **Dichte:**

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{sonst} \end{cases}$$

- **Verteilungsfunktion:**

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

- **Träger:**  $T_X = [a, b]$
- **Erwartungswert:**  $E(X) = (a + b)/2$

**Exponentialverteilung**  $Exp(\lambda)$  mit Parameter  $\lambda > 0$

- **Dichte:**

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

- **Verteilungsfunktion:**

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

- **Träger:**  $T_X = [0, \infty)$
- **Erwartungswert:**  $E(X) = 1/\lambda$
- **Varianz:**  $Var(X) = 1/\lambda^2$

**Gamma-Verteilung**  $\Gamma(\alpha, \beta)$  mit den Parametern  $\alpha > 0$  und  $\beta > 0$

- **Dichte:**

$$f(x) = \begin{cases} 0, & x < 0 \\ \frac{\alpha^\beta}{\Gamma(\beta)} x^{\beta-1} e^{-\alpha x}, & x \geq 0 \end{cases},$$

wobei

$$\Gamma(\beta) = \int_0^\infty t^{\beta-1} e^{-t} dt.$$

- **Verteilungsfunktion:** Geschlossene Darstellung nur für  $\beta \in \mathbb{N}$
- **Träger:**  $T_X = [0, \infty]$

- **Erwartungswert:**  $E(X) = \beta/\alpha$
- **Varianz:**  $Var(X) = \beta/\alpha^2$
- **Bemerkung:**  $\Gamma(\alpha, 1) = Exp(\alpha)$

Sind  $X$  und  $Y$  unabhängige Zufallsvariable mit den Gamma-Verteilungen  $X \sim \Gamma(\alpha, \beta_1)$  und  $Y \sim \Gamma(\alpha, \beta_2)$ , dann gilt  $X + Y \sim \Gamma(\alpha, \beta_1 + \beta_2)$ .

### Chi-Quadrat-Verteilung $\chi_n^2$ mit $n \geq 1$ Freiheitsgraden

- **Dichte:**

$$f(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & x \geq 0 \end{cases},$$

- **Verteilungsfunktion:** Geschlossene Darstellung nur für gerade  $n \in \mathbb{N}$
- **Träger:**  $T_X = [0, \infty)$

Sind  $X$  und  $Y$  unabhängige Zufallsvariable mit den  $\chi^2$ -Verteilungen  $X \sim \chi_n^2$  und  $Y \sim \chi_m^2$ , dann gilt  $X + Y \sim \chi_{n+m}^2$ .

### Normalverteilung $N_{(\mu, \sigma^2)}$ mit den Parametern $\mu$ und $\sigma^2 > 0$

- **Dichte:**

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

- **Verteilungsfunktion:** Die Verteilungsfunktion ist nicht geschlossen darstellbar. Für  $\mu = 0$  und  $\sigma^2 = 1$  heißt

$$\Phi(x) := \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

Verteilungsfunktion der Standardnormalverteilung.

- **Träger:**  $T_X = (-\infty, \infty)$
- **Erwartungswert:**  $E(X) = \mu$
- **Varianz:**  $Var(X) = \sigma^2$
- **Bemerkung:** Die Normalverteilung ist symmetrisch um  $\mu$ . Für die Verteilungsfunktion  $\Phi_{\mu, \sigma}$  von  $N_{(\mu, \sigma^2)}$  gilt

$$\Phi_{\mu, \sigma}(t) = \Phi\left(\frac{t-\mu}{\sigma}\right)$$

Sind  $X$  und  $Y$  unabhängige Zufallsvariable mit den Normalverteilungen  $X \sim N_{(\mu_1, \sigma_1^2)}$  und  $Y \sim N_{(\mu_2, \sigma_2^2)}$ , dann gilt  $X + Y \sim N_{(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)}$ .

### (Student-)t-Verteilung $t_n$ mit $n$ Freiheitsgraden

- **Dichte:**

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad x \in \mathbb{R}.$$

- **Verteilungsfunktion:** Nicht geschlossen darstellbar.
- **Träger:**  $T_X = (-\infty, \infty)$
- **Erwartungswert:**  $E(X) = 0$

## Extremwertverteilungen:

### Gumbel-Verteilungsfunktion $G_0$

- **Dichte:**  $f(x) = e^{-x} \exp(-e^{-x})$
- **Verteilungsfunktion:**  $G_0(x) = \exp(-e^{-x})$
- **Träger:**  $T_X = (-\infty, \infty)$

### Fréchet-Verteilungsfunktion $G_{1,\alpha}$ mit dem Parameter $\alpha > 0$

- **Dichte:**

$$f(x) = \begin{cases} 0, & x < 0 \\ \alpha x^{-(1+\alpha)} \exp(-x^{-\alpha}), & x \geq 0 \end{cases}$$

- **Verteilungsfunktion:**

$$G_{1,\alpha}(x) = \begin{cases} 0, & x < 0 \\ \exp(-x^{-\alpha}), & x \geq 0 \end{cases}$$

- **Träger:**  $T_X = [0, \infty)$

### Weibull-Verteilungsfunktion $G_{2,\alpha}$ mit dem Parameter $\alpha > 0$

- **Dichte:**

$$f(x) = \begin{cases} \alpha(-x)^{\alpha-1} \exp(-(-x)^\alpha), & x < 0 \\ 0, & x \geq 0 \end{cases}$$

- **Verteilungsfunktion:**

$$G_{2,\alpha}(x) = \begin{cases} \exp(-(-x)^\alpha), & x < 0 \\ 1, & x \geq 0 \end{cases}$$

- **Träger:**  $T_X = (-\infty, 0]$

## Verallgemeinerte Pareto-Verteilungen

Hierzu gehören neben der Exponentialverteilung die

### Pareto-Verteilungsfunktion $W_{1,\alpha}$ mit dem Parameter $\alpha > 0$

- **Dichte:**

$$f(x) = \begin{cases} 0, & x < 1 \\ \alpha x^{-(1+\alpha)}, & x \geq 1 \end{cases}$$

- **Verteilungsfunktion:**

$$W_{1,\alpha}(x) = \begin{cases} 0, & x < 1 \\ 1 - x^{-\alpha}, & x \geq 1 \end{cases}$$

- **Träger:**  $T_X = [0, \infty]$

### Beta-Verteilungsfunktion $W_{2,\alpha}$ mit dem Parameter $\alpha > 0$

- **Dichte:**

$$f(x) = \begin{cases} \alpha(-x)^{\alpha-1}, & -1 \leq x \leq 0 \\ 0, & \text{sonst} \end{cases}$$

- **Verteilungsfunktion:**

$$W_{2,\alpha}(x) = \begin{cases} 0, & x < -1 \\ 1 - (-x)^\alpha, & -1 \leq x \leq 0 \\ 1, & x > 0 \end{cases} .$$

- **Träger:**  $T_X = [-1, 0]$

### Lokation- und Skalenparameter

Seien  $\mu$  und  $\sigma > 0$  reelle Zahlen. Ist  $X$  eine Zufallsvariable mit Verteilungsfunktion  $F$ , dann hat die Zufallsvariable

$$Y = \mu + \sigma \cdot X$$

die Verteilungsfunktion  $F_{\mu,\sigma}$  mit **Lokationsparameter**  $\mu$  und **Skalenparameter**  $\sigma$ , wobei

$$F_{\mu,\sigma}(t) = P\{\mu + \sigma \cdot X \leq t\} = P\{X \leq (t - \mu)/\sigma\} = F\left(\frac{t - \mu}{\sigma}\right).$$

## 8 Grenzwertsätze

In diesem Abschnitt sollen einige wichtige Grenzwertsätze und Ungleichungen der Wahrscheinlichkeitstheorie vorgestellt werden. Dabei sprechen wir von **identisch verteilten Zufallsvariablen**  $X_1, \dots, X_n$ , falls alle Zufallsvariablen dieselbe Verteilungsfunktion besitzen, d.h.

$$P\{X_1 \leq t\} = P\{X_2 \leq t\} = \dots = P\{X_n \leq t\}$$

für alle reellen Zahlen  $t$ . Insbesondere gilt dann auch

$$\mu := E(X_1) = \dots = E(X_n)$$

und

$$\sigma^2 := \text{Var}(X_1) = \dots = \text{Var}(X_n).$$

### Ungleichung von Tschebyschev

**Satz 8.1** Seien  $X_1, \dots, X_n$  unabhängige und identisch verteilte Zufallsvariable mit  $\mu = E(X_1)$  und  $\sigma^2 = \text{Var}(X_1)$ . Dann gilt

$$P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq t\right\} \leq \frac{\sigma^2}{n \cdot t^2}$$

für alle reellen Zahlen  $t > 0$ .

### Bemerkung:

- (i) Speziell erhält man für  $n = 1$  und  $t > 0$

$$P\{|X_1 - E(X_1)| \geq t\} \leq \frac{\text{Var}(X_1)}{t^2}.$$

- (ii) Die Ungleichung von Tschebyschev liefert eine Abschätzung für die Wahrscheinlichkeit, dass das arithmetische Mittel vom Erwartungswert  $\mu$  um mindestens  $t$  abweicht. Mit wachsendem Stichprobenumfang wird diese Wahrscheinlichkeit kleiner.
- (iii) Ein Maß für die Abweichung vom Erwartungswert ist die Varianz. Es gilt nach Satz 5.5 und Satz 6.3

$$\text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n} \text{Var}(X_1).$$

**Bemerkung 8.2** Für unabhängige und binomialverteilte Zufallsvariable  $Y_i \sim B(1, p)$  gilt

$$E(Y_1) = p \quad \text{und} \quad \text{Var}(Y_1) = p(1-p) \leq \frac{1}{4}$$

für alle  $p \in (0, 1)$ . Die Ungleichung von Tschebyscheff liefert damit folgende Abschätzung.

$$P\left\{\left|\frac{1}{n} \sum_{i=1}^n Y_i - p\right| \geq t\right\} \leq \frac{p(1-p)}{n \cdot t^2} \leq \frac{1}{4 \cdot n \cdot t^2}$$

Man beachte, dass  $Y_1 + \dots + Y_n$  eine  $B(n, p)$ -verteilte Zufallsvariable ist.



**Beispiel 8.3** Zur Überprüfung der Bruchfestigkeit sollen  $n$  gleichartige Bauteile einem Test unterzogen werden. Jedes Bauteil bricht mit Wahrscheinlichkeit  $p$  unabhängig von allen anderen Bauteilen.

- (i) Sei  $p = 0.15$  und  $n = 100$ . Gesucht ist eine untere Schranke für die Wahrscheinlichkeit, dass mindestens 11 und höchstens 19 Bauteile brechen.

Lösung: Sei  $Y_i = 1$ , falls das  $i$ -te Bauteil bricht und  $Y_i = 0$  sonst. Dann beschreibt  $S_n = \sum_{i=1}^n Y_i$  die zufällige Anzahl der Bauteile, die bei dem Test zerbrechen. Offenbar können wir Bemerkung 8.2 anwenden.

$$\begin{aligned} P\left\{\left|\sum_{i=1}^{100} Y_i - 15\right| \geq 5\right\} &= P\left\{\left|\frac{1}{100} \sum_{i=1}^{100} Y_i - \frac{15}{100}\right| \geq \frac{5}{100}\right\} \\ &\leq \frac{0.15(1 - 0.15)}{100 \cdot \left(\frac{5}{100}\right)^2} = 0.51 \end{aligned}$$

- (ii) Sei nun  $p$  unbekannt. Wieviele Bauteile müssen höchstens überprüft werden, damit mit einer Wahrscheinlichkeit von mindestens 0.95 die relative Anzahl  $S_n/n$  von dem unbekanntem Wert  $p$  um nicht mehr als 0.1 abweicht.

Lösung: Gesucht ist das kleinste  $n$  mit

$$P\left\{\left|\frac{1}{n} \sum_{i=1}^n Y_i - p\right| \leq 0.1\right\} \geq 0.95$$

Das bedeutet

$$P\left\{\left|\frac{1}{n} \sum_{i=1}^n Y_i - p\right| > 0.1\right\} \leq 0.05$$

Es gilt

$$P\left\{\left|\frac{1}{n} \sum_{i=1}^n Y_i - p\right| > 0.1\right\} \leq \frac{1}{4 \cdot n \cdot 0.1^2}$$

und

$$\frac{1}{4 \cdot n \cdot 0.1^2} \leq 0.05 \iff n \geq \frac{1}{4 \cdot 0.05 \cdot 0.1^2} = 500$$

Also müssen höchstens  $n = 500$  Bauteile überprüft werden.

## Der Zentrale Grenzwertsatz

Die große Bedeutung der Normalverteilung ist insbesondere in dem folgenden bemerkenswerten Resultat über das asymptotische Verhalten von Summen unabhängiger Zufallsvariablen begründet.

**Satz 8.4** Seien  $X_1, \dots, X_n, \dots$  unabhängige und identisch verteilte Zufallsvariable mit  $\mu = E(X_1)$  und  $\sigma := \sqrt{\text{Var}(X_1)} \in (0, \infty)$ . Setzen wir  $S_n := X_1 + \dots + X_n$ , dann konvergiert die Verteilungsfunktion  $F_n$  der standardisierten Summe  $(S_n - n\mu)/(\sqrt{n}\sigma)$  gegen die Verteilungsfunktion  $\Phi$  der Standardnormalverteilung. Es gilt

- (i)

$$\lim_{n \rightarrow \infty} P\left\{\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq t\right\} = \lim_{n \rightarrow \infty} F_n(t) = \Phi(t) = \int_{-\infty}^t \varphi(x) dx,$$

(ii)

$$\lim_{n \rightarrow \infty} P\left\{a \leq \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq b\right\} = \Phi(b) - \Phi(a) = \int_a^b \varphi(x) dx,$$

wobei

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

die Dichte der Standardnormalverteilung bezeichnet.

**Bemerkung:**

(i) Der Zentrale Grenzwertsatz besagt, dass mit wachsendem Stichprobenumfang die standardisierte Summe  $(S_n - n\mu)/(\sqrt{n} \cdot \sigma)$  annähernd normalverteilt ist. Unter der Unabhängigkeitsannahme gilt diese Näherung für alle Verteilungen, für die Erwartungswert und Varianz existieren.

(ii) Für den Erwartungswert und die Varianz erhält man

$$E\left(\frac{S_n - n\mu}{\sqrt{n}\sigma}\right) = 0$$

und

$$\text{Var}\left(\frac{S_n - n\mu}{\sqrt{n}\sigma}\right) = \frac{1}{n\sigma^2} \text{Var}(S_n) = 1.$$

Als Spezialfall des Zentralen Grenzwertsatzes erhält man für eine binomialverteilte Zufallsvariable folgende Approximation.

**Satz 8.5** (Satz von Moivre-Laplace)

Sei  $Z_n \sim B_{(n,p)}$ . Dann gilt

$$\lim_{n \rightarrow \infty} P\left\{\frac{Z_n - np}{\sqrt{np(1-p)}} \leq t\right\} = \Phi(t)$$

für alle reellen Zahlen  $t$ .

**Bemerkung:** Sind  $X_1, \dots, X_n$  unabhängige,  $B_{(1,p)}$ -verteilte Zufallsvariable, dann ist  $Z_n = X_1 + \dots + X_n$  eine  $B_{(n,p)}$ -verteilte Zufallsvariable.

**Fortsetzung von Beispiel 8.3:**

(i) Nun soll die Wahrscheinlichkeit in Beispiel 8.3 (ii) mit dem Zentralen Grenzwertsatz abgeschätzt werden. Mit  $n = 100$  und  $p = 0.15$  gilt  $n \cdot p = 15$  und

$$\begin{aligned} P\left\{\left|\sum_{i=1}^{100} Y_i - 15\right| \leq 4\right\} &= P\left\{-4 \leq \sum_{i=1}^{100} Y_i - 15 \leq 4\right\} \\ &= P\left\{a \leq \frac{\sum_{i=1}^{100} Y_i - 15}{\sqrt{100 \cdot 0.15(1-0.15)}} \leq b\right\}, \end{aligned}$$

wobei

$$a = \frac{-4}{\sqrt{100 \cdot 0.15(1-0.15)}} \approx -1.12$$

und

$$b = \frac{4}{\sqrt{100 \cdot 0.15(1 - 0.15)}} \approx 1.12.$$

Der Zentrale Grenzwertsatz liefert die Näherung

$$\begin{aligned} P\left\{a \leq \frac{\sum_{i=1}^{100} Y_i - 15}{\sqrt{100 \cdot 0.15(1 - 0.15)}} \leq b\right\} &\approx \Phi(b) - \Phi(a) \\ &\approx \Phi(1.12) - \Phi(-1.12) \\ &= 2\Phi(1.12) - 1 \approx 0.737 \end{aligned}$$

Wegen der Symmetrie der Standardnormalverteilung gilt  $\Phi(a) = 1 - \Phi(-a)$ .

- (ii) Für diskrete Zufallsvariable wird häufig eine **Stetigkeitskorrektur** durchgeführt. Da  $\sum_{i=1}^n Y_i$  nur ganzzahlige Werte annehmen kann, gilt auch

$$P\left\{\left|\sum_{i=1}^{100} Y_i - 15\right| \leq 4\right\} = P\left\{-4.5 \leq \sum_{i=1}^{100} Y_i - 15 \leq 4.5\right\}$$

und man ersetzt  $a$  und  $b$  durch

$$a^* = \frac{-4.5}{\sqrt{100 \cdot 0.15(1 - 0.15)}} \approx -1.26 \quad \text{und} \quad b^* = -a^*$$

und erhält

$$\Phi(b^*) - \Phi(a^*) = 2 * \Phi(b^*) - 1 \approx 0.79$$

Wegen

$$b^* - b = \frac{0.5}{\sqrt{np(1-p)}}$$

nähern sich die Ergebnisse mit und ohne Stetigkeitskorrektur mit wachsendem Stichprobenumfang  $n$  an. (Exakter Wert: 0.794008....)

**Bemerkung:** Die Verteilungsfunktion einer  $B_{(n,p)}$ -verteilten Zufallsvariablen  $Z_n$  kann also durch die Verteilungsfunktion einer geeigneten Normalverteilung oder, falls  $p$  klein ist, durch die Verteilungsfunktion einer Poisson-verteilten Zufallsvariablen approximiert werden. D.h:

$$P\{Z_n \leq t\} \approx \Phi\left(\frac{t - np}{\sqrt{np(1-p)}}\right)$$

bzw. mit Stetigkeitskorrektur, falls  $t$  ganzzahlig ist,

$$P\{Z_n \leq t\} \approx \Phi\left(\frac{(t + 0.5) - np}{\sqrt{np(1-p)}}\right),$$

falls  $n$  hinreichend groß ist.

Ist  $p$  klein, dann wählt man die Poisson-Approximation

$$P\{Z_n \leq t\} \approx Po(np)([0, t]) = \sum_{0 \leq k \leq t} \frac{(np)^k}{k!} e^{-np}.$$

## 9 Parameterschätzung

In Anwendungen ist die vorliegende Verteilung in der Regel nicht bekannt (z.B. Fluthöhen, Messfehler) oder nur von der Art her bekannt (z.B. Qualitätskontrolle). Gesucht sind nun Schätzverfahren für gewisse Parameter dieser Verteilung. Ein Schätzverfahren ist eine Vorschrift, nach der basierend auf einer Stichprobe ein Schätzwert für einen unbekanntem Parameter berechnet wird.

**Beispiel:** Aus der Ungleichung von Tschebyschev folgt, dass das arithmetische Mittel von  $n$  unabhängigen und identisch verteilten Zufallsvariablen  $X_1, \dots, X_n$  für hinreichend großen Stichprobenumfang nahe an dem Erwartungswert  $\mu = E(X_1)$  liegt, falls  $Var(X_1) < \infty$ .

Soll nun der unbekannte Parameter  $\mu = E(X_1)$  geschätzt werden, dann liefert das arithmetische Mittel  $g(x) := \bar{x} = (x_1 + \dots + x_n)/n$  basierend auf der beobachteten Stichprobe  $x_1, \dots, x_n$  einen Schätzwert für  $\mu$ .

Das Schätzverfahren im Beispiel kann durch die Funktion  $g$  beschrieben werden. Für eine konkrete Stichprobe  $x_1, \dots, x_n$  heißt eine Funktion  $g$  oder

$$g(X_1, \dots, X_n)$$

**Schätzfunktion** oder **Schätzer** und

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

**Schätzwert** oder **Schätzung**.

Ein Schätzer  $g(X_1, \dots, X_n)$  ist eine Zufallsvariable und besitzt eine Verteilung.

Es ergeben sich folgende Fragestellungen:

1. Wie konstruiert man Schätzverfahren für unbekanntem Parameter?
2. Wie genau sind die Schätzwerte?

Ein Konstruktionsprinzip wurde im Beispiel schon angedeutet. Ein weiteres Konstruktionsprinzip, die Maximum-Likelihood-Methode, folgt später.

Im Folgenden seien  $x_1, \dots, x_n$  die Werte einer Stichprobe, die als Realisationen von unabhängigen und identisch verteilten Zufallsvariablen  $X_1, \dots, X_n$  mit Verteilungsfunktion  $F$  betrachtet werden. Der Erwartungswert von  $X_1$  heißt auch **Mittelwert** von  $F$  oder Mittelwert der Verteilung von  $X$ .

**Schätzer für den Mittelwert**

Als Schätzer für den Mittelwert wählen wir das arithmetische Mittel der Zufallsvariablen  $X_1, \dots, X_n$

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

Die zugehörige Schätzung ist das Stichprobenmittel

$$\hat{\mu} := \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

### Schätzer für die Varianz $\sigma^2$

Die Stichprobenvarianz

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

ist eine Schätzung für die Varianz  $\sigma^2 = \text{Var}(X_1)$ . Die zugehörige Schätzfunktion lautet

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Es gilt

$$E(S^2) = \sigma^2,$$

der Schätzer ist **erwartungstreu**. Der Schätzer  $S$  für  $\sigma$  ist nicht erwartungstreu.

### Schätzer für die Wahrscheinlichkeit $p = P\{X_1 \in A\}$

Als Schätzung für  $p$  wählen wir die relative Häufigkeit

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n 1_A(x_i).$$

Der zugehörige Schätzer lautet  $\hat{K}/n$  mit  $\hat{K} := \sum_{i=1}^n 1_A(X_i)$ . Offenbar ist  $\hat{K}$  eine  $B_{(n,p)}$ -verteilte Zufallsvariable, wobei  $p$  unbekannt ist. Wegen

$$E(\hat{K}/n) = \frac{1}{n} \sum_{i=1}^n E(1_A(X_i)) = P\{X_1 \in A\}$$

ist der Schätzer erwartungstreu.

**Definition 9.1** Seine  $X_1, \dots, X_n$  unabhängige Zufallsvariable mit identischer Verteilungsfunktion  $F$  und  $\vartheta_F$  ein zu schätzender unbekannter Parameter.

Ein Schätzer  $\hat{g}_n = g_n(X_1, \dots, X_n)$  heißt

- (i) **erwartungstreu**, falls für alle zulässigen Verteilungsfunktionen  $F$  gilt

$$E_F(\hat{g}_n) = \vartheta_F,$$

- (ii) **konsistent**, falls für alle zulässigen Verteilungsfunktionen  $F$  und alle  $t > 0$  gilt

$$\lim_{n \rightarrow \infty} P_F\{|\hat{g}_n - \vartheta_F| > t\} = 0.$$

Der Index  $F$  deutet die Abhängigkeit von der jeweiligen Verteilungsfunktion an.

**Bemerkung:** Die Erwartungstreue eines Schätzers ist kein wesentliches Kriterium. Tatsächlich existiert häufig kein erwartungstreuer Schätzer.

## Maximum-Likelihood-Methode

Die Maximum-Likelihood-Methode liefert ein allgemeines Konstruktionsprinzip für Schätzfunktionen.

Zur Motivation betrachten wir eine diskrete Zufallsvariable  $X$  mit Zähldichte  $f_k(x) := P\{X = x\}$ . Die Zähldichte  $f_k(x)$  sei bis auf einen Parameter  $k$  bekannt. Es sei

$$f_1(0) = \frac{1}{2}, \quad f_1(1) = \frac{1}{4}, \quad f_1(2) = \frac{1}{4}$$

und

$$f_2(0) = \frac{1}{5}, \quad f_2(1) = \frac{4}{5}, \quad f_2(2) = 0.$$

Sei nun  $x$  eine Realisation von  $X$  (Stichprobe vom Umfang 1). Gesucht ist ein Schätzwert für  $k$  basierend auf  $x$ .

Nach der Maximum-Likelihood-Methode wird der Parameter  $\hat{k}(x)$  gewählt, dessen zugehörige Zähldichte die größte Wahrscheinlichkeit für  $x$  aufweist.

$x$	0	1	2
$f_1(x)$	0.5	0.25	0.25
$f_2(x)$	0.2	0.8	0
$\hat{k}(x)$	1	2	1

Die Erweiterung auf einen Stichprobenumfang  $n > 1$  ist unmittelbar möglich. Sind  $x_1, \dots, x_n$  Realisationen der unabhängigen Zufallsvariablen  $X_1, \dots, X_n$ , die alle die gleiche Zähldichte  $f_1(x_i)$  oder  $f_2(x_i)$  haben, dann gilt

$$\begin{aligned} L(k) &:= L_{x_1, \dots, x_n}(k) \\ &:= P\{X_1 = x_1, \dots, X_n = x_n\} \\ &= P\{X_1 = x_1\} \cdot \dots \cdot P\{X_n = x_n\} \\ &= f_k(x_1) \cdot \dots \cdot f_k(x_n), \end{aligned}$$

wobei  $k \in \{1, 2\}$  unbekannt ist. Nun wird der Parameter  $\hat{k}(x_1, \dots, x_n)$  gewählt, für den die Zähldichte  $f_k(x_1) \cdot \dots \cdot f_k(x_n)$  der gemeinsamen Verteilung von  $X_1, \dots, X_n$  bzw. die **Likelihood-Funktion**  $L(k)$  für die konkrete Stichprobe  $x_1, \dots, x_n$  maximal ist.

Ist z.B.  $n = 3$ ,  $x_1 = 1$ ,  $x_2 = 0$ ,  $x_3 = 1$ , dann gilt

$$L_{1,0,1}(k) = \begin{cases} 0.25 \cdot 0.5 \cdot 0.25 = 0.03125, & k = 1 \\ 0.8 \cdot 0.2 \cdot 0.8 = 0.128, & k = 2. \end{cases}$$

Also:  $\hat{k}(1, 0, 1) = 2$ .

**Zur Erinnerung:**  $X$  besitzt die Dichte  $f$ , falls  $P\{X \leq x\} = \int_{-\infty}^x f(t) dt$  für alle reellen Zahlen  $t$ .

**Definition 9.2** Seien  $X_1, \dots, X_n$  unabhängige Zufallsvariable, die alle dieselbe Zähldichte oder Dichte  $f_\vartheta$  besitzen, wobei der Parameter  $\vartheta$  unbekannt ist. Sei  $x_1, \dots, x_n$  eine Stichprobe vom Umfang  $n$ .

(i) Die Funktion

$$L(\vartheta) := L_{x_1, \dots, x_n}(\vartheta) := f_\vartheta(x_1) \cdot \dots \cdot f_\vartheta(x_n)$$

heißt **Likelihood-Funktion**.  $L$  ist bei fest vorgegebenen Werten  $x_1, \dots, x_n$  eine Funktion des unbekanntem Parameters  $\vartheta$ .

(ii) Der Schätzwert

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

für den die Likelihood-Funktion maximal wird, also

$$L(\hat{\vartheta}) = L(g(x_1, \dots, x_n)) = \max_{\vartheta} L(\vartheta),$$

heißt **Maximum-Likelihood-Schätzung** zur Stichprobe  $x_1, \dots, x_n$ . Die zugehörige Schätzfunktion  $g(X_1, \dots, X_n)$  heißt **Maximum-Likelihood-Schätzfunktion**.

### Berechnung einer Maximum-Likelihood-Schätzung

Die Maximum-Likelihood-Schätzung ist unter Differenzierbarkeitsbedingungen eine Lösung der Gleichungen

$$\frac{\partial L(\vartheta)}{\partial \vartheta_j} = 0, \quad j = 1, \dots, d,$$

falls der Parameter  $\vartheta = (\vartheta_1, \dots, \vartheta_d)$  die Dimension  $d$  besitzt.

Wegen der Produktstruktur von  $L$  ist es in der Regel einfacher, die **Log-Likelihood-Funktion**  $\log L(\vartheta)$  zu betrachten und die **Likelihood-Gleichungen**

$$\frac{\partial \log L(\vartheta)}{\partial \vartheta_j} = \sum_{i=1}^n \frac{\partial \log f_{\vartheta}(x_i)}{\partial \vartheta_j} = 0, \quad j = 1, \dots, d,$$

zu lösen. Häufig können die Lösungen dieser Gleichungen nur numerisch bestimmt werden.

Falls mehrere verschiedene Lösungen existieren, die alle die Likelihood-Funktion maximieren, dann ist die Maximum-Likelihood-Schätzung nicht eindeutig bestimmt.

### Beispiel 9.3 (Maximum-Likelihood-Schätzungen)

(i) (Bernoulli-Verteilung/Binomialverteilung)

Die Zähldichte der Binomialverteilung  $B_{(1,p)}$  lautet

$$f_p(x) = p^x(1-p)^{1-x}, \quad x = 0, 1.$$

Zu einer Stichprobe  $x_1, \dots, x_n \in \{0, 1\}$  erhält man

$$\begin{aligned} \log L(p) &= \sum_{i=1}^n \log f_p(x_i) = \sum_{i=1}^n (x_i \cdot \log(p) + (1-x_i) \cdot \log(1-p)) \\ &= k \cdot \log(p) + (n-k) \cdot \log(1-p), \end{aligned}$$

wobei  $k := x_1 + \dots + x_n$ . Nun gilt

$$0 = \frac{\partial \log L(p)}{\partial p} = \frac{k}{p} - \frac{n-k}{1-p}$$

Nach Auflösen dieser Gleichung nach  $p$  erhält man

$$\hat{p} = \frac{k}{n} = \bar{x}.$$

als Maximum-Likelihood-Schätzung zum Stichprobenumfang  $n$ .

Betrachtet man nicht die einzelnen Werte  $x_1, \dots, x_n$  sondern nur  $k := x_1 + \dots + x_n$ , dann wählt man die Dichte einer Binomialverteilung  $B_{(n,p)}$  und maximiert

$$\begin{aligned} \log \tilde{L}(p) &= \log \left( \binom{n}{k} p^k (1-p)^{n-k} \right) \\ &= \log \left( \binom{n}{k} \right) + k \cdot \log(p) + (n-k) \cdot \log(1-p) \end{aligned}$$

Als Lösung erhält man ebenfalls  $\hat{p} = k/n$ .

(ii) (Exponentialverteilung)

Die Exponentialverteilung  $Exp(\lambda)$  besitzt die Dichte  $f_\lambda(x) = \lambda e^{-\lambda x}$ ,  $x > 0$ . Die Log-Likelihood-Funktion zum Stichprobenumfang  $n$  lautet

$$\log L(\lambda) = \sum_{i=1}^n (\log(\lambda) - \lambda \cdot x_i) = n \cdot \log(\lambda) - \lambda \cdot n \cdot \bar{x}.$$

Damit gilt

$$0 = \frac{\partial \log L(\lambda)}{\partial \lambda} = \frac{n}{\lambda} - n \cdot \bar{x}.$$

Auflösen nach  $\lambda$  ergibt

$$\hat{\lambda} = \frac{1}{\bar{x}} = \frac{n}{\sum_{i=1}^n x_i}.$$

(iii) (Normalverteilung)

Die Likelihood-Funktion lautet

$$\begin{aligned} L(\vartheta) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_1 - \mu)^2}{2\sigma^2}\right) \cdot \dots \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_n - \mu)^2}{2\sigma^2}\right) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right), \end{aligned}$$

wobei  $\vartheta = (\mu, \sigma^2)$  unbekannt. Zu lösen sind die Likelihood-Gleichungen

$$\begin{aligned} 0 &= \frac{\partial \log L(\vartheta)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = \frac{n}{\sigma^2} (\bar{x} - \mu), \\ 0 &= \frac{\partial \log L(\vartheta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

Die Lösung ist

$$\hat{\vartheta} = \left( \bar{x}, \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right).$$



## 10 Konfidenzintervalle

Die Schätzwerte aus dem vorherigen Abschnitt liefern keine Aussage über die Abweichung vom unbekanntem Parameter. Gesucht ist nun ein Intervall, das den wahren aber unbekanntem Parameter mit einer vorgegebenen Wahrscheinlichkeit überdeckt.

Sei  $0 < \beta < 1$  gegeben. Wir suchen eine untere und eine obere Schätzfunktion  $g_u(X_1, \dots, X_n)$  bzw.  $g_o(X_1, \dots, X_n)$ , so dass für alle zulässigen Parameter  $\vartheta$  und alle unabhängigen Zufallsvariable  $X_1, \dots, X_n$  mit Verteilungsfunktion  $F_\vartheta$  gilt

$$P\{g_u(X_1, \dots, X_n) \leq \vartheta \leq g_o(X_1, \dots, X_n)\} \geq \beta.$$

Im Mittel liegt dann der wahre Parameter in  $\beta \cdot 100\%$  (oder mehr) der Fälle für eine konkrete Stichprobe  $x_1, \dots, x_n$  zwischen den Werten  $g_u(x_1, \dots, x_n)$  und  $g_o(x_1, \dots, x_n)$ . Das Intervall

$$C(x) := [g_u(x_1, \dots, x_n), g_o(x_1, \dots, x_n)]$$

heißt **Konfidenz-** oder **Vertrauensintervall für den Parameter  $\vartheta$  zum Niveau  $\beta$** . Der Wert  $\alpha := 1 - \beta$  heißt **Irrtumswahrscheinlichkeit**.

Je nach Problemstellung wird man ein möglichst kleines zweiseitiges Konfidenzintervall oder einseitiges Konfidenzintervall z.B. von der Form  $(-\infty, g_o(x_1, \dots, x_n)]$  suchen.

**Definition 10.1** Sei  $F$  eine Verteilungsfunktion und  $0 < q < 1$ . Der Wert

$$u_q := \inf\{t : F(t) \geq q\}$$

heißt  **$q$ -Quantil** von  $F$ . Die Funktion  $F^{-1} : (0, 1) \rightarrow \mathbb{R}$ ,  $F^{-1}(q) = u_q$ , heißt **Quantilfunktion** oder **verallgemeinerte Inverse**.

### Bemerkung

- (i) Falls  $F$  streng monoton wachsend und stetig ist, dann ist  $F^{-1}$  die Umkehrfunktion von  $F$ .
- (ii) Ist  $F$  stetig in  $u_q$ , dann gilt  $F(u_q) = q$ .
- (iii) Für alle  $0 < q < 1$  und alle reellen Zahlen  $x$  gilt

$$F^{-1}(q) \leq x \iff q \leq F(x).$$

### Vertrauensintervalle für den unbekanntem Mittelwert einer Normalverteilung bei bekannter Varianz

Seien  $X_1, \dots, X_n$  unabhängige normalverteilte Zufallsvariable mit unbekanntem Mittelwert  $\mu$  und bekannter Varianz  $\sigma^2$ . Dann gilt

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

Sei  $c > 0$ . Dann

$$\begin{aligned} \Phi(c) - \Phi(-c) &= P\left\{-c \leq \frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n (X_i - \mu) \leq c\right\} \\ &= P\left\{-c \frac{\sigma}{\sqrt{n}} \leq \frac{1}{n} \sum_{i=1}^n (X_i - \mu) \leq c \frac{\sigma}{\sqrt{n}}\right\} \\ &= P\left\{\bar{X} - c \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + c \frac{\sigma}{\sqrt{n}}\right\} \end{aligned}$$

Wählt man  $u_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ , dann gilt

$$\Phi(u_{1-\alpha/2}) - \Phi(-u_{1-\alpha/2}) = 1 - \alpha = \beta$$

und damit

$$P\left\{\bar{X} - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right\} = \beta.$$

**Bemerkung 10.2** Das Konfidenzintervall zum Niveau  $\beta = 1 - \alpha$  für den Mittelwert  $\mu$  bei bekannter Varianz  $\sigma^2$  lautet

$$C(x) = \left[ \bar{x} - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

### Vertrauensintervalle für den unbekanntem Mittelwert einer Normalverteilung bei unbekannter Varianz

Ist zusätzlich die Varianz unbekannt, so wird diese durch

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{bzw.} \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

geschätzt und

$$\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n (X_i - \mu) \quad \text{durch} \quad T = \frac{1}{\sqrt{n}S} \sum_{i=1}^n (X_i - \mu)$$

ersetzt.

**Satz 10.3** Die Zufallsvariable  $T$  besitzt eine  $t$ -Verteilung mit  $n - 1$  Freiheitsgraden, d.h.

$$P\{T \leq t\} = \int_{-\infty}^t \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} dx.$$

**Bemerkung 10.4** Als Konfidenzintervall zum Niveau  $\beta = 1 - \alpha$  für den Mittelwert  $\mu$  bei unbekannter Varianz erhält man das Intervall

$$\tilde{C}(x) = \left[ \bar{x} - t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}} \right]$$

wobei  $t_{n-1, q}$  das  $q$ -Quantil der  $t_{n-1}$ -Verteilung bezeichnet.

**Bemerkung:** Für  $n \rightarrow \infty$  gilt  $t_{n-1, q} \rightarrow u_q$ .

### Vertrauensintervalle für den unbekanntem Mittelwert bei bekannter oder unbekannter Varianz

Sind  $X_1, \dots, X_n$  unabhängige Zufallsvariable mit Erwartungswert  $\mu = E(X_1)$  und Varianz  $\sigma^2$ , dann gilt nach dem Zentralen Grenzwertsatz 8.4

$$\lim_{n \rightarrow \infty} P\left\{\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n (X_i - \mu) \leq t\right\} = \Phi(t)$$

Ebenso kann man die folgende Konvergenz zeigen:

$$\lim_{n \rightarrow \infty} P\left\{\frac{1}{\sqrt{n}S} \sum_{i=1}^n (X_i - \mu) \leq t\right\} = \Phi(t)$$

**Bemerkung 10.5** Die Konfidenzintervalle  $C$  und  $\tilde{C}$  aus Bemerkung 10.2 bzw. 10.4 sind **asymptotische Konfidenzintervalle** zum Niveau  $\beta = 1 - \alpha$ . Es gilt

$$\lim_{n \rightarrow \infty} P \left\{ \bar{X} - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right\} = \beta$$

und

$$\lim_{n \rightarrow \infty} P \left\{ \bar{X} - t_{n-1, 1-\alpha/2} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1, 1-\alpha/2} \frac{S}{\sqrt{n}} \right\} = \beta$$

Das Quantil  $t_{n-1, 1-\alpha/2}$  der  $t$ -Verteilung kann auch durch  $u_{1-\alpha/2}$  ersetzt werden.

### Vertrauensintervalle für die unbekannte Varianz einer Normalverteilung

Unter der Normalverteilungsannahme können Konfidenzintervalle für die unbekannte Varianz mit Hilfe der  $\chi^2$ -Verteilung berechnet werden. Die Zufallsvariable

$$Z := \frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2$$

ist  $\chi_{n-1}^2$ -verteilt. Da die Chi-Quadrat-Verteilung nicht symmetrisch ist, wählt man  $c_1 = F^{-1}(\alpha/2)$  und  $c_2 = F^{-1}(1 - \alpha/2)$ , wobei  $F$  die Verteilungsfunktion der  $\chi_{n-1}^2$ -Verteilung bezeichnet. Dann gilt

$$P \left\{ c_1 \leq Z \leq c_2 \right\} = F(c_2) - F(c_1) = 1 - \alpha$$

und

$$P \left\{ c_1 \leq Z \leq c_2 \right\} = P \left\{ c_1 \leq \frac{n-1}{\sigma^2} S^2 \leq c_2 \right\} = P \left\{ \frac{n-1}{c_2} S^2 \leq \sigma^2 \leq \frac{n-1}{c_1} S^2 \right\}.$$

**Bemerkung 10.6** Als Konfidenzintervall zum Niveau  $\beta = 1 - \alpha$  für die Varianz  $\sigma^2$  bei unbekanntem Mittelwert  $\mu$  erhält man das Intervall

$$C_v(x) = \left[ \frac{n-1}{\chi_{n-1, 1-\alpha/2}^2} s^2, \frac{n-1}{\chi_{n-1, \alpha/2}^2} s^2 \right]$$

wobei  $\chi_{n-1, q}^2$  das  $q$ -Quantil der  $\chi_{n-1}^2$ -Verteilung bezeichnet.

## 11 Korrelation und Regression

Wir betrachten in diesem Abschnitt Zufallsexperimente, in denen zwei Merkmale beobachtet werden. Die zufälligen Ergebnisse werden durch die Zufallsvariablen  $X$  und  $Y$  beschrieben. Das Tupel  $Z := (X, Y)$  wird als **zweidimensionale Zufallsvariable** oder auch als **Zufallsvektor** bezeichnet.

Die Verteilung einer zweidimensionalen Zufallsvariablen kann vollständig durch die Verteilungsfunktion

$$F(x, y) = P\{X \leq x, Y \leq y\}$$

beschrieben werden.  $F(x, y)$  gibt die Wahrscheinlichkeit an, dass sowohl  $X \leq x$  als auch  $Y \leq y$  eintreten.

### Diskrete zweidimensionale Zufallsvariable

Sind  $X$  und  $Y$  diskrete Zufallsvariable, die nur die Werte  $x_1, \dots, x_m$  und  $y_1, \dots, y_k$  annehmen, dann kann die Verteilung von  $Z = (X, Y)$  vollständig durch die Zähldichte

$$f(x, y) = P\{X = x, Y = y\} = \begin{cases} P\{X = x_i, Y = y_j\}, & x = x_i, y = y_j \\ 0, & \text{sonst} \end{cases}$$

beschrieben werden. Es gilt

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f(x_i, y_j).$$

Weiterhin ist  $f$  **normiert**, d.h.

$$\sum_{i=1}^m \sum_{j=1}^k f(x_i, y_j) = 1.$$

### Bemerkung:

- (i) Ist die Zähldichte  $f(x, y)$  einer zweidimensionalen Zufallsvariablen  $(X, Y)$  gegeben, dann gilt für die Zähldichten  $f_X(x)$  und  $f_Y(y)$  der eindimensionalen Zufallsvariablen  $X$  und  $Y$

$$f_X(x) = \sum_{j=1}^k f(x, y_j), \quad f_Y(y) = \sum_{i=1}^m f(x_i, y).$$

Die Verteilungen von  $X$  und  $Y$  heißen **Randverteilungen**.

- (ii) Sind  $X$  und  $Y$  **unabhängig**, dann gilt  $f(x, y) = f_X(x) \cdot f_Y(y)$ , wobei  $f_X(x) := P\{X = x\}$  und  $f_Y(y) := P\{Y = y\}$ . Im Allgemeinen gilt dies jedoch nicht!

**Beispiel 11.1** Ein Experiment mit Erfolgswahrscheinlichkeit  $2/3$  wird zweimal durchgeführt. Es sei  $X = 1$ , falls das 1. Experiment erfolgreich ist und  $X = 0$  sonst. Sei  $Y$  die Anzahl der Erfolge in beiden Experimenten. Die Zähldichte von  $(X, Y)$  ist

$x_i \backslash y_j$	0	1	2
0	1/9	2/9	0
1	0	2/9	4/9

$X$  und  $Y$  sind keine unabhängigen Zufallsvariablen. Es gilt z.B.

$$P\{X = 0, Y = 2\} = 0 \neq P\{X = 0\} \cdot P\{Y = 2\} = 1/3 \cdot 4/9.$$

## Stetige zweidimensionale Zufallsvariable

Stetige zweidimensionale Zufallsvariable  $(X, Y)$  werden durch eine Dichte  $f(x, y) \geq 0$  beschrieben. Es gilt

$$P\{X \leq x, Y \leq y\} = F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) \, dv \, du$$

Die Dichte  $f$  ist **normiert**, d.h.

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

### Bemerkung:

- (i) Ist die Dichte  $f(x, y)$  einer zweidimensionalen Zufallsvariablen  $(X, Y)$  gegeben, dann gilt für die Dichten  $f_X(x)$  und  $f_Y(y)$  der Zufallsvariablen  $X$  und  $Y$

$$f_X(x) = \int_{-\infty}^{\infty} f(x, v) \, dv, \quad f_Y(y) = \int_{-\infty}^{\infty} f(u, y) \, du.$$

Die Verteilungen von  $X$  und  $Y$  heißen **Randverteilungen**.

- (ii) Sind  $X$  und  $Y$  **unabhängig**, dann gilt  $f(x, y) = f_X(x) \cdot f_Y(y)$ , wobei  $f_X$  und  $f_Y$  die Dichten von  $X$  bzw.  $Y$  sind. Im Allgemeinen gilt dies jedoch nicht!

**Beispiel 11.2** (i) Die Dichte der zweidimensionalen Gleichverteilung auf  $[0, 3] \times [-1, 1]$  lautet

$$f(x, y) = 1/3 \cdot 1/2 = 1/6, \quad 0 \leq x \leq 2, -1 \leq y \leq 1$$

und  $f(x, y) = 0$  sonst. Ist  $(X, Y)$  eine zweidimensionale Zufallsvariable mit der Dichte  $f(x, y)$ , dann sind  $X$  und  $Y$  unabhängige und auf  $[0, 3]$  bzw.  $[-1, 1]$  gleichverteilte Zufallsvariable.

- (ii) Die Dichte der zweidimensionalen Normalverteilung  $N_{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)}$  mit den Parametern  $\mu_1, \mu_2 \in \mathbb{R}$ ,  $\sigma_1^2, \sigma_2^2 > 0$  und  $-1 < \rho < 1$  ist gegeben durch

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right)$$

Ist  $(X, Y) \sim N_{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)}$ , dann ist  $X \sim N_{(\mu_1, \sigma_1^2)}$  und  $Y \sim N_{(\mu_2, \sigma_2^2)}$ .  $X$  und  $Y$  sind genau dann unabhängig, wenn  $\rho = 0$  gilt.

## Kovarianz und Korrelationskoeffizient

Sei  $(X, Y)$  eine zweidimensionale Zufallsvariable. Die **Kovarianz von  $X$  und  $Y$**  ist definiert durch

$$\sigma_{XY} := E\left[\left(X - E(X)\right)\left(Y - E(Y)\right)\right] = E(XY) - E(X) \cdot E(Y).$$

Es gilt  $\sigma_X^2 := \text{Var}(X) = \sigma_{XX}$  und  $\sigma_Y^2 := \text{Var}(Y) = \sigma_{YY}$ .

Der **Korrelationskoeffizient von  $X$  und  $Y$**  ist gegeben durch

$$\rho = \rho_{xy} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y},$$

wobei  $\sigma_X, \sigma_Y > 0$  vorausgesetzt wird. Die Zufallsvariablen  $X$  und  $Y$  heißen **unkorreliert**, wenn  $\rho = 0$  gilt.

**Bemerkung 11.3** Seien  $\sigma_X, \sigma_Y > 0$ .

- (i) Unabhängige Zufallsvariable sind unkorreliert. Die Umkehrung gilt im Allgemeinen nicht.
- (ii) Es gilt stets  $-1 \leq \rho \leq 1$ .
- (iii) Der lineare Zusammenhang  $Y = aX + b$  gilt genau dann für geeignete Zahlen  $a > 0$  und  $b \in \mathbb{R}$ , wenn  $|\rho| = 1$ .
- (iv) Seien  $c > 0$ ,  $d \in \mathbb{R}$  und  $Z := cX + d$ . Dann gilt  $\rho_{XY} = \rho_{ZY}$ .
- (v) Für diskrete Zufallsvariable  $(X, Y)$  mit Zähldichte  $f$  berechnet sich  $E(XY)$  wie folgt:

$$E(XY) = \sum_{i=1}^n \sum_{j=1}^k x_i \cdot y_j \cdot f(x_i, y_j)$$

Für stetige Zufallsvariable  $(X, Y)$  mit Dichte  $f$  gilt

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y \cdot f(x, y) dx dy.$$

**Beispiel 11.4** Sei  $(X, Y) \sim N_{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)}$  zweidimensional normalverteilt. Dann ist  $\rho_{XY} = \rho$ . Weiterhin sind  $X$  und  $Y$  genau dann unabhängig, wenn  $\rho = 0$  erfüllt ist.

Es gilt der folgende Satz.

**Satz 11.5** (Transformationsformel)

Sei  $(X, Y)$  eine zweidimensionale Zufallsvariable und  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$  eine Abbildung. Dann gilt für die Zufallsvariable  $Z := g(X, Y)$ :

$$E(Z) := E(g(X, Y)) = \begin{cases} \sum_{i=1}^m \sum_{j=1}^k g(x_i, y_j) \cdot f(x_i, y_j), & \text{falls } (X, Y) \text{ diskret} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot f(x, y) dx dy, & \text{falls } (X, Y) \text{ stetig,} \end{cases}$$

falls die linke Seite für  $|g(x, y)|$  statt  $g(x, y)$  endlich ist.

**Bemerkung** Ist  $(x_1, y_1), \dots, (x_n, y_n)$  eine Stichprobe von unabhängigen Wiederholungen des durch die Zufallsvariable  $(X, Y)$  beschriebenen Zufallsexperimentes, dann kann der Korrelationskoeffizient  $\rho$  durch den **empirischen Korrelationskoeffizienten**

$$r := \frac{s_{xy}}{s_x \cdot s_y}$$

geschätzt werden, wobei

$$s_{xy} := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

die **empirische Kovarianz** ist und

$$s_x^2 := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad s_y^2 := \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

## Lineare Regression

Im Folgenden soll ein Verfahren zur Beschreibung eines funktionalen Zusammenhangs der Form

$$Y = f(X)$$

zwischen einer (Zufalls-)Variablen  $X$  und einer Zufallsvariablen  $Y$  vorgestellt werden. (Beispiele: Auslenkung  $y_i$  einer Federwaage bei einem Gewicht  $x_i$ , Fahrbahnverschleiß  $y_i$  in Abhängigkeit von dem Gewicht  $x_i$  eines Fahrzeugs). Dabei wird  $X$  als **unabhängige** und  $Y$  als **abhängige Variable** bezeichnet.  $X$  kann eine Zufallsvariable (z.B. das zufällig beobachtete Fahrzeuggewicht) oder eine Variable (genau definiertes Gewicht in einem Versuchsaufbau) sein. Basierend auf einer Stichprobe

$$(x_1, y_1), \dots, (x_n, y_n),$$

soll eine **Regressionsfunktion**  $f$  bestimmt werden. Dabei beschränkt man sich auf eine gewisse Klasse von Funktionen und versucht, eine Funktion aus dieser Klasse möglichst gut anzupassen.

Wir betrachten die Klasse der linearen Funktionen, d.h.

$$f(x) = ax + b$$

für alle reellen Zahlen  $x$ , wobei  $a$  und  $b$  unbekannt sind. Die Regressionsfunktion heißt dann auch **Regressionsgerade**. Als Verfahren verwenden wir die **Methode der kleinsten Quadrate**.

Gesucht ist eine Lösung des Minimierungsproblems

$$\min_{a,b} \sum_{i=1}^n (y_i - ax_i - b)^2$$

Nach der Methode der kleinsten Quadrate wird die Summe der quadratischen Abstände  $(y_i - f(x_i))^2$  minimiert. Als Lösungen erhält man

$$\hat{a} = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

und

$$\hat{b} = \bar{y} - \hat{a} \cdot \bar{x}.$$

Eine Erweiterung auf mehrere erklärende Variable  $X_1, \dots, X_d$  ist möglich. Bei der **multiplen linearen Regression** lautet die Regressionsfunktion

$$f(x_1, \dots, x_d) = \sum_{i=1}^d a_i x_i + b.$$

Ein Spezialfall ist die **polynomiale Regression** mit den erklärenden Variablen  $X_1 = X, X_2 = X^2, \dots, X_d = X^d$ ) und der Regressionsfunktion

$$f(x) = \sum_{i=1}^d a_i x^i + b.$$

## 12 Hypothesentests

Eine statistische Hypothese ist eine Annahme über die Verteilung einer Zufallsgröße. Mit Hilfe eines Hypothesentests soll eine Entscheidung getroffen werden, ob diese Annahme gilt oder nicht.

Die Hypothese, die auf ihre Gültigkeit überprüft werden soll, wird als **Alternative** oder **Alternativhypothese**  $H_1$  bezeichnet. Als gegenteilige Annahme wird die **Nullhypothese** formuliert. Nullhypothese und Alternative schließen einander aus. Wesentlich ist, dass die nachzuweisende Eigenschaft oder Verteilungsannahme immer als Alternative zu formulieren ist.

Ein Verfahren, das basierend auf einer Stichprobe eine Entscheidung liefert, ob die Nullhypothese zu verwerfen ist oder nicht, wird als **statistischer Test** bezeichnet. Da die Beobachtungen zufällig sind, wird die getroffene Entscheidung nicht immer richtig sein. Folgende Konstellationen können eintreten:

Entscheidung	In der Grundgesamtheit ist	
	$H_0$ richtig	$H_1$ richtig
$H_0$ annehmen	Entscheidung richtig	Fehler 2. Art
$H_0$ ablehnen	Fehler 1. Art	Entscheidung richtig

Eine Fehlentscheidung liegt vor, wenn für die Alternative  $H_1$  entschieden wird, obwohl  $H_0$  richtig ist (Fehler 1. Art) bzw. wenn für die Nullhypothese  $H_0$  entschieden wird, obwohl  $H_1$  richtig ist (Fehler 2. Art). Der Fehler 1. Art wird als schwerwiegender eingeschätzt als der Fehler 2. Art.

**Beispiel 12.1** Die Geschwindigkeit eines PKW sei  $N_{(\mu, \sigma^2)}$ -verteilt. Die Auswertung einer Stichprobe von 50 PKW ergab eine mittlere Geschwindigkeit von  $\bar{x} = 76$  km/h und eine Stichprobenvarianz von  $s^2 = 120$ . Aus Gründen der Verkehrssicherheit wird eine mittlere Geschwindigkeit von  $\mu < 80$  gefordert. Hat die Einhaltung der Verkehrssicherheit oberste Priorität, dann kann das statistische Testproblem wie folgt formuliert werden:

$$H_0 : \mu \geq 80 \quad \text{gegen} \quad H_1 : \mu < 80.$$

Eine Fehlentscheidung zugunsten von  $H_1$ , obwohl tatsächlich  $\mu \geq 80$  gilt, wird als schwerwiegender Fehler erachtet.

Der Fehler 1. Art soll nur mit Wahrscheinlichkeit  $\leq \alpha$  (z.B.  $\alpha = 0.05$ ) eintreten. Kann ein Testverfahren gefunden werden, das einen Fehler 1. Art nur mit einer Wahrscheinlichkeit  $\leq \alpha$  besitzt und der Test eine Entscheidung für  $H_1$  liefert, dann gilt die Alternative  $H_1$  als **statistisch gesichert zum Signifikanzniveau  $\alpha$** .

Ein statistischer Test hat häufig die folgende Form: Basierend auf einer Stichprobe  $X_1, \dots, X_n$  wird eine **Teststatistik** oder **Prüfgröße**  $T = T(X_1, \dots, X_n)$  berechnet. Der Wert wird verglichen mit einem **kritischen Wert**  $c$ .

**Beispiel 12.2** (Fortsetzung von Beispiel 12.1) Als Teststatistik wählen wir

$$T(X_1, \dots, X_n) := \frac{\sum_{i=1}^n X_i - n\mu_0}{\sqrt{n}S} = \sqrt{n} \frac{\bar{X} - \mu_0}{S},$$

wobei  $\mu_0 = 80$  und

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$



Da  $\bar{X}$  ein Schätzer für den wahren Parameter  $\mu$  ist, wird man  $H_0$  ablehnen, wenn  $T$  hinreichend klein ist, also  $T < c$ .

Also:

Lehne  $H_0$  ab, falls  $T < c$ .

Das Signifikanzniveau sei  $\alpha = 0.05$ . Die Bestimmung des kritischen Wertes  $c$  erfolgt gemäß der Vorschrift

$$P\{H_0 \text{ wird abgelehnt, obwohl } H_0 \text{ richtig ist}\} \leq \alpha,$$

d.h.

$$P_{\mu,\sigma}\{T < c\} \leq \alpha, \quad \text{für alle } \mu \geq 80, \sigma > 0.$$

Ist  $\mu = \mu_0 = 80$  der wahre Parameter, dann ist  $T$   $t_{n-1}$ -verteilt. Es gilt

$$P_{80,\sigma}\{T < t_{n-1,\alpha}\} = \alpha,$$

wobei  $t_{n-1,q}$  das  $q$ -Quantil der  $t_{n-1}$ -Verteilung bezeichnet. Für  $\mu > 80$  gilt sogar

$$P_{\mu,\sigma}\{T < t_{n-1,\alpha}\} < \alpha,$$

Als kritischer Wert kann also  $c = t_{n-1,\alpha}$  gewählt werden.

Im vorliegenden Beispiel gilt

$$T(x_1, \dots, x_n) = \sqrt{n} \frac{\bar{x} - \mu_0}{s} = \sqrt{50} \frac{76 - 80}{\sqrt{120}} \approx -2.58$$

Wegen  $c := t_{n-1,\alpha} = -t_{49,1-0.05} \approx -1.677$  gilt  $T < c$ . Die Nullhypothese wird abgelehnt. Die Stichprobe ist signifikant zum Niveau  $\alpha = 0.05$ .

Wählt man als Signifikanzniveau  $\tilde{\alpha} = 0.005$ , dann erhält man den kritischen Wert  $\tilde{c} = -t_{49,0.995} \approx -2.68$ . Zu diesem Fehlerniveau wird die Hypothese nicht abgelehnt. Die Alternative kann nicht als statistisch gesichert zum Niveau 0.005 angesehen werden.

Im Folgenden werden einige Testverfahren für den Mittelwert einer Normalverteilung angegeben. Für weitere relevante Testverfahren wird auf Band 3 von Papula, 1994, verwiesen. Es seien  $u_q$  das  $q$ -Quantil der Standardnormalverteilung und

$$S^2 = S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

### Einseitige Tests für den Mittelwert einer Normalverteilung

Verteilungsannahme:  $N_{(\mu,\sigma^2)}$ ,  $\mu$  unbekannt,  $\sigma^2$  bekannt oder unbekannt

Testproblem:  $H_0 : \mu \leq \mu_0$  gegen  $H_1 : \mu > \mu_0$

Gaußtest ( $\sigma^2$  bekannt)

Prüfgröße:  $T = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma}$

$H_0$  wird abgelehnt:  $T > u_{1-\alpha}$

t-Test ( $\sigma^2$  unbekannt)

Prüfgröße:  $T = \sqrt{n} \frac{\bar{X} - \mu_0}{S}$

$H_0$  wird abgelehnt:  $T > t_{n-1,1-\alpha}$

### Zweiseitige Tests für den Mittelwert einer Normalverteilung

Verteilungsannahme:  $N_{(\mu, \sigma^2)}$ ,  $\mu$  unbekannt,  $\sigma^2$  bekannt oder unbekannt

Testproblem:  $H_0 : \mu = \mu_0$  gegen  $H_1 : \mu \neq \mu_0$

Gaußtest ( $\sigma^2$  bekannt)

Prüfgröße:  $T = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma}$

$H_0$  wird abgelehnt:  $|T| > u_{1-\alpha/2}$

t-Test ( $\sigma^2$  unbekannt)

Prüfgröße:  $T = \sqrt{n} \frac{\bar{X} - \mu_0}{S}$

$H_0$  wird abgelehnt:  $|T| > t_{n-1, 1-\alpha/2}$