

Materialien zu Stochastik 1

# Einführung in die Wahrscheinlichkeitsrechnung und Statistik

Dr. Christian Kredler

WS 2003/04

# Inhaltsverzeichnis

<b>Teil 1: Wahrscheinlichkeitsrechnung</b>	<b>1</b>
<b>1 Grundlagen der Wahrscheinlichkeitsrechnung</b>	<b>1</b>
1.1 Ereignisse, Wahrscheinlichkeit . . . . .	1
1.2 Wahrscheinlichkeitsaxiome von Kolmogorov . . . . .	4
1.3 Kombinatorik, Abzählen . . . . .	7
1.4 Unabhängigkeit, bedingte Wahrscheinlichkeit . . . . .	10
1.5 Zufallsvariable . . . . .	17
1.5.1 Diskrete Zufallsvariable . . . . .	18
1.5.2 Verteilungsfunktion . . . . .	21
1.5.3 Stetige Zufallsvariable (ZV) . . . . .	23
1.5.4 Prozentpunkte und Quantilfunktion . . . . .	29
1.5.5 Eindimensionale Transformationen von ZV . . . . .	32
1.6 Erwartungswerte . . . . .	36
1.6.1 Erwartungswert und Varianz . . . . .	36
1.6.2 Momente und Kumulanten . . . . .	42
<b>2 Mehrdimensionale Verteilungen</b>	<b>46</b>
2.1 Gemeinsame Verteilung von $n$ Zufallsvariablen . . . . .	46
2.1.1 Diskrete $n$ -dim. Zufallsvariable . . . . .	47
2.1.2 Stetige $n$ -dim. Zufallsvariable . . . . .	49
2.1.3 Randverteilungen . . . . .	51
2.1.4 Erwartungswerte; $n$ -dim. . . . .	54
2.2 Unabhängige Zufallsvariable . . . . .	59
2.3 Transformationen von $n$ Zufallsvariablen . . . . .	63
2.3.1 Transformationssatz für Dichten . . . . .	63
2.3.2 Lineare Transformationen . . . . .	64
2.3.3 Summen, Quotienten . . . . .	68
2.3.4 Minimum, Maximum . . . . .	70
2.4 Bedingte Verteilungen und Erwartungswerte . . . . .	72
2.4.1 Bedingung bzgl. Ereignis $B$ . . . . .	72
2.4.2 Bedingte Verteilungen . . . . .	74
2.4.3 Bedingte Erwartung . . . . .	76
2.5 Ordnungsstatistiken . . . . .	79

<b>3 Erzeugende Funktionen</b>	<b>82</b>
3.1 Nichtnegative, ganzzahlige Zufallsvariable . . . . .	82
3.2 Momenterzeugende Funktion . . . . .	87
<b>4 Gesetze der grossen Zahlen</b>	<b>92</b>
4.1 Ungleichungen . . . . .	92
4.2 Gesetze der grossen Zahlen . . . . .	95
4.3 Grenzwertsätze . . . . .	97
4.3.1 Zentraler Grenzwertsatz . . . . .	97
4.3.2 Approximation der Binomial- durch die Poisson-Verteilung . . .	99
<b>5 Einfache Irrfahrt (Random Walk)</b>	<b>101</b>
5.1 Definition, Rekurrenz . . . . .	101
5.2 Stoppzeiten . . . . .	106
<b>6 Eine Auswahl wichtiger Verteilungen</b>	<b>108</b>
6.1 Einige diskrete Verteilungen . . . . .	109
6.2 Beispiele stetiger Verteilungen . . . . .	112
<b>Teil 2: Statistik</b>	<b>119</b>
<b>7 Deskriptive Statistik</b>	<b>120</b>
7.1 Grundgesamtheit, Merkmale . . . . .	121
7.1.1 Klassifizierung von Merkmalen, Skalentypen . . . . .	122
7.1.2 Messreihen, Stichproben . . . . .	124
7.2 Empirische Verteilung eindim. diskreter Merkmale . . . . .	125
7.2.1 Tabellierung und grafische Darstellung . . . . .	125
7.2.2 Relative Summenhäufigkeit . . . . .	126
7.3 Empirische Verteilung eindim. stetiger Merkmale . . . . .	129
7.4 Lageparameter, Lokationsmaße . . . . .	132
7.4.1 Modus (Modalwert) . . . . .	132
7.4.2 Empirische Quantile, Quartile, Median . . . . .	133
7.4.3 Arithmetisches Mittel . . . . .	135
7.5 Streuungsparameter . . . . .	135
7.5.1 Spannweite, Variationsbreite . . . . .	135
7.5.2 Mittlere quadratische Abweichung, Standardabweichung . . . .	136
7.5.3 Mittlere lineare Streuung . . . . .	137
7.5.4 Interquartilsabstand, inter quartile range ( <i>IQR</i> ) . . . . .	137
7.5.5 Variationskoeffizient . . . . .	138
7.6 Schiefe und Exzess . . . . .	139
7.7 Mittelwerte, Streuungen für gruppierte Daten . . . . .	140
7.8 Affine Transformationen . . . . .	141
7.9 Empirische Regression . . . . .	142
7.9.1 Statistische Maßzahlen . . . . .	142
7.9.2 KQ-Gerade, Bestimmtheitsmaß . . . . .	144
7.10 Beispiele . . . . .	154
7.10.1 Grafiken und Ausdrücke von Statistikprogrammen . . . . .	154

7.10.2	Einige Datenbeispiele . . . . .	157
<b>8</b>	<b>Schätzfunktionen, ML-Prinzip</b>	<b>158</b>
8.1	Erwartungstreue und Konsistenz . . . . .	159
8.2	Schätzfunktionen minimaler Varianz . . . . .	160
8.3	Konstruktion von Schätzern . . . . .	161
8.3.1	Methode der Kleinsten Quadrate (KQ) . . . . .	162
8.3.2	Momentenmethode . . . . .	162
8.3.3	Maximum-Likelihood (ML)-Methode . . . . .	162
<b>9</b>	<b>Statistik normalverteilter Daten</b>	<b>165</b>
9.1	Stichprobenverteilungen . . . . .	165
9.1.1	Normalverteilung . . . . .	165
9.1.2	$\chi^2$ -Verteilung . . . . .	168
9.1.3	Student- oder t-Verteilung . . . . .	169
9.1.4	Stichprobenmittel und -varianz . . . . .	170
9.1.5	F-Verteilung . . . . .	170
9.2	Konfidenzintervalle, Tests . . . . .	172
9.2.1	Konfidenzintervalle . . . . .	172
9.2.2	Tests, Konstruktion des kritischen Bereichs . . . . .	173
9.2.3	Fehler 2. Art . . . . .	176
9.2.4	Zweiseitiger Test . . . . .	178
9.3	Konfidenzintervalle und t-Test; $\sigma$ unbekannt . . . . .	178
9.3.1	Quantile der t-Verteilung . . . . .	179
9.3.2	Konfidenzintervalle . . . . .	179
9.3.3	Konstruktion von $T$ -Statistiken, t-Tests . . . . .	180
9.3.4	t-Test verbundener Stichproben; matched pairs . . . . .	181
9.3.5	t-Test zum Vergleich von Mittelwerten . . . . .	183
9.4	Vergleich von Varianzen, F-Test . . . . .	185
9.5	Der $p$ -Wert . . . . .	187
9.6	Einfache lineare Regression . . . . .	189
9.7	Multiple Regression . . . . .	193
<b>10</b>	<b>Explorative Datenanalyse</b>	<b>196</b>
10.1	Kerndichte-Schätzer . . . . .	197
10.2	QQ-Plots . . . . .	201
10.3	Box-Plots . . . . .	207
<b>11</b>	<b>Anhang, Tabellen, Approximationen</b>	<b>212</b>
11.1	Asymptotische Formeln . . . . .	212
11.2	Rationale Bestapproximationen für $\Phi$ und $\Phi^{-1}$ . . . . .	214
11.3	Tabellen . . . . .	214
11.4	Normal- bzw. Poissonapprox. der Binomialverteilung . . . . .	219
	<b>Literatur</b>	<b>220</b>
	<b>Index</b>	<b>222</b>

# Einführung

Die Vorlesung **Einführung in die Wahrscheinlichkeitsrechnung und Statistik** (Stochastik 1) stellt den ersten Teil des viersemestrigen Zyklus mit Wahrscheinlichkeitstheorie (Stochastik 2), Stochastischen Prozessen (Stochastik 3) und der Mathematischen Statistik (Stochastik 4) an der TU München dar. In Stochastik 1 werden elementare Grundlagen der Wahrscheinlichkeitsrechnung und der Statistik vermittelt. Bei der einfachen Irrfahrt (random walk) kommen bereits erste Eigenschaften stochastischer Prozesse zur Sprache. Die Anwendung auf praxisrelevante Szenarien, in diesem Fall auf den *Ruin des Spielers*, wird stets im Auge behalten.

Die Vorlesung ist für Diplom-Mathematiker, Techno-, Finanz- und Wirtschaftsmathematiker sowie Studierende des Lehramts Mathematik an Gymnasien konzipiert. Sie kann ab dem dritten Semester gehört werden, verwendet noch keine abstrakte Maßtheorie und bietet sich ebenso für Physiker und Ingenieure an.

Der wahrscheinlichkeitstheoretische Teil (Kapitel 1 bis 6) ist also elementar gehalten, er orientiert sich an klassischen Lehrbüchern wie [Chung (1979)] und [Stirzaker (1994)]. Schon aus Zeitgründen wird in der Vorlesung auf manche statistischen Aspekte nicht mit voller Ausführlichkeit eingegangen werden können. Hier sind diese Materialien als Ergänzung und Anregung zum Selbststudium gedacht. Die einführenden statistischen Kapitel 7 bis 10 dienen aber in erster Linie als Beiheft des vorlesungsbegleitenden Statistik-Praktikums. Dieses wird vorwiegend mit den Programmpaketen R und S-Plus durchgeführt.

Ein Durcharbeiten dieser Materialien ersetzt nicht den Besuch der Vorlesung. Auf die Erstellung eines kompletten *Skriptums* wurde in Anbetracht vorhandener Lehrbücher (siehe oben) verzichtet. Im wahrscheinlichkeitstheoretischen Teil sind Beweise fast durchwegs weggelassen. Statt dessen findet man genaue Referenzen auf die Bücher von [Chung (1979)] und [Stirzaker (1994)].

Ein weiterer Zweck dieser Ausarbeitung ist die komplette Darstellung umfangreicher Rechnungen, die sich schlecht für eine Darbietung an der Tafel eignen, aber dem Leser beim Durcharbeiten einen gründlichen Einblick in wichtige Zusammenhänge vermitteln; zum Beispiel Integrationen bei der zweidimensionalen Normalverteilung.

Nun zur Gliederung dieser Materialien (in der Vorlesung wird das Kapitel über deskriptive Statistik aus didaktischen Gründen an den Anfang gestellt):

- Die sechs ersten Kapitel stellen wichtige Grundlagen der Wahrscheinlichkeitsrechnung zusammen. Ein Großteil der in der Vorlesung verwendeten grafischen

Darstellungen sind hier bereits mit abgebildet, so dass Hörerinnen und Hörern ein mühsames Reproduzieren dieses Materials erspart bleibt.

- Die Beschreibung und Klassifikation von Merkmalen sowie die Erläuterung einfacher Methoden der deskriptiven Statistik sind Gegenstand des siebten Kapitels. Weitere Stichworte lauten hier Histogramme, empirische Verteilungsfunktion, empirische Momente und Quantile, sowie empirische Korrelation und Kleinste-Quadrate-Gerade durch eine Punktwolke.
- Der propädeutischen Ausrichtung folgend werden im Anschluss Eigenschaften statistischer Schätzfunktionen diskutiert und u.a. das Maximum-Likelihood-Prinzip als Verfahren zur Gewinnung von Punktschätzern dargestellt.
- Wichtige Elemente der mathematischen Statistik normalverteilter Zufallsvariablen bis hin zur einfachen linearen Regression mit Gaußschen Fehlern sind Gegenstand von Kapitel 9.
- Das letzte Kapitel setzt beim Leser Fertigkeiten in Wahrscheinlichkeitsrechnung voraus und skizziert einige Standardmethoden der explorativen Datenanalyse, die in Statistik-Programmpaketen wie SAS, S-Plus oder SPSS angeboten werden. Dazu gehören u.a. nichtparametrische Kerndichteschätzer sowie Box- und QQ-Plots zur Auswahl geeigneter Verteilungsmodelle. Die beiden letzten Kapitel liefern die theoretischen Grundlagen für das Statistik-Praktikum am Rechner.
- Schließlich folgen die wichtigsten Verteilungstafeln und einige Approximationsformeln, insbesondere für Näherungen des Normalverteilungsintegrals.
- Im Literaturverzeichnis ist eine knappe Auswahl einführender Lehrbücher angegeben. Neben den oben bereits erwähnten Standardwerken stützt sich diese Ausarbeitung im statistischen Teil vorwiegend auf [Falk et al. (1995)].
- Diese Materialien sind in manchen Details umfangreicher als die Einführungsvorlesung. So dient ein ausführliches Stichwortverzeichnis dem schnellen Auffinden wichtiger Begriffe und weiterführender Referenzen. Ein Anspruch auf Vollständigkeit wird nicht erhoben.

Die Gliederung der Vorlesung Stochastik 1 weicht, wie schon erwähnt, zu Beginn vom Aufbau dieser Materialien ab. Dies hat folgenden Grund:

1. Die Studierenden sollen von Anfang an in die Lage versetzt werden, Statistik-Programmpakete zur Analyse einfacher Datensätze zu benützen und die ausgedruckten Maßzahlen und Grafiken zu interpretieren. Deswegen befassen wir uns in den ersten Wochen – vor allem im vorlesungsbegleitenden Praktikum – kurz mit den deskriptiven Methoden, die im siebten Kapitel dieser Materialien zum weitgehenden Selbststudium ausführlicher und mit zahlreichen Beispielrechnungen dargestellt sind.
2. Erst danach folgt das Kernstück der Vorlesung. Ausbildungsziel ist es, mit den Studierenden den Kalkül der elementaren Wahrscheinlichkeitsrechnung gründlich einzuüben. Diskrete und stetige Wahrscheinlichkeitsräume sowie die Begriffe

Zufallsexperiment, Ereignis, Wahrscheinlichkeit, Zufallsvariable, Dichte und Verteilungsfunktion werden eingeführt und illustriert. In den Übungen und Hausaufgaben sind der Stoff und die nötigen Rechenfertigkeiten von den Studierenden an weiteren Beispielen zu vertiefen. Vorkenntnisse aus der Kollegstufe sind dabei hilfreich, aber keine Voraussetzung für das Verständnis. Ein weiteres Ziel ist der sichere Umgang mit Erwartungswerten, Transformationen und bedingten Verteilungen diskreter und stetiger Zufallsvariablen einschließlich der Betrachtung von Ordnungsstatistiken. Es folgen Gesetze der großen Zahlen und eine einfache Version des Zentralen Grenzwertsatzes. Mit dem sog. *Random Walk* unternehmen wir erste Gehversuche im wichtigen Gebiet der Stochastischen Prozesse.

3. Die wahrscheinlichkeitstheoretischen Grundlagen aus dem Hauptteil erlauben einen Ausbau der eingangs behandelten empirischen Methoden in den Kapiteln über mathematische Statistik. Dort folgt die Vorlesung im wesentlichen dem Aufbau dieser Materialien.

Herrn Dipl. Math. Martin Severin möchte ich für Verbesserungsvorschläge beim Erstellen dieser Materialien danken.

Mein besonderer Dank gilt Frau Prof. Dr. Klüppelberg, die beim Konzeptentwurf mit wertvollen Hinweisen zur Seite stand.

# Kapitel 1

## Grundlagen der Wahrscheinlichkeitsrechnung

Die deskriptive Statistik (vgl. Kapitel 7) arbeitet mit anschaulichen Begriffen wie **Population**, **Merkmal**, (relative) **Häufigkeit** etc.

In der axiomatisch fundierten Wahrscheinlichkeitstheorie werden entsprechende abstrakte Formulierungen verwendet. Die Kenntnis einiger Begriffe der beschreibenden Statistik ist zwar keine Voraussetzung zum Verständnis der folgenden Abschnitte, stellt sich aber erfahrungsgemäß in einem ersten Stochastik-Kurs als hilfreich heraus.

### 1.1 Ereignisse, Wahrscheinlichkeit

**Def. 1.1 (Zufallsexperiment, Ergebnismenge, sample space)**

$\Omega \neq \emptyset$  sei die Menge aller möglichen Ergebnisse eines mathematischen Zufallsexperiments, die sog. **Ergebnismenge**. Man spricht auch vom **Stichprobenraum** (sample space).

#### Beispiel 1.2 (Zufallsexperimente)

1. Bernoulli-Experiment: Werfen einer Münze  
 $\Omega = \{Kopf, Wappen\}$  oder  $\Omega = \{0, 1\}$
2. Würfeln:  $\Omega = \{1, 2, 3, 4, 5, 6\}$
3. Lotto 6 aus 49:  $\Omega = \{\omega \mid \omega = \{j_1, \dots, j_6\}, j_1, \dots, j_6 \in \{1, 2, 3, \dots, 48, 49\}\}$   
Da Mengen nur verschiedene Elemente enthalten, gilt  $|\{j_1, \dots, j_6\}| = 6$ .
4. Anzahl der Anrufe in einer Telefonvermittlung pro Tag :  $\Omega = \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ .
5.  $\Omega = \{\omega \mid \omega = \text{Matrikelnummer eines TU - Studenten im WS 1997/98}\}$ .
6. Schuss auf eine unendlich große Scheibe:  $\Omega = \mathbb{R}^2$ .



7. Verlauf der Körpertemperatur eines Lebewesens:

$$\{ \omega = (id, f) \mid id \in \mathbb{N}, f \in C(\mathbb{R}_+) \}.$$

Ergebnis des Experiments ist also die Identifikationsnummer  $id$  des Lebewesens und eine (beschränkte) stetige Funktion auf der nichtnegativen reellen Achse.  $f(0)$  ist die Körpertemperatur bei der Geburt. Nach dem Tod  $T > 0$  des Lebewesens könnte man etwa die Umgebungstemperatur zur Fortsetzung der Funktion  $f$  heranziehen.

Das letzte Beispiel zeigt, dass auch Funktionen als Ergebnisse eines Zufallsexperiments auftreten können.  $\Omega$  kann also endlich, abzählbar oder sogar überabzählbar unendlich sein.

Man interessiert sich nun dafür, ob bei Durchführung des Zufallsexperiments bestimmte **Ereignisse** eintreten. Zum Beispiel, ob

1. beim Wurf einer Münze  $A = \{Kopf\}$  gefallen ist
2. beim Würfeln eine 5 oder 6, d. h.  $B = \{5, 6\}$  herauskam
3. im Lotto 6 aus 49 "sechs Richtige" angekreuzt wurden
4. mehr als 1000 Anrufe pro Tag in der Telefonvermittlung,  $D = \{n \mid n > 1000\}$ , auftraten
5.  $K = \{\omega \mid \text{Matrikelnummer } \omega \text{ beginnt mit einer } 7\}$
6. ein Treffer im Ziel  $Z$  landete; z. B. mit  $Z = \{(x = r \cos \varphi, y = r \sin \varphi) \mid 0 \leq r < 5; 0 \leq \varphi < 2\pi\}$
7. die Körpertemperatur nie den Wert  $40^\circ C$  überschritt.

In jedem Fall handelt es sich bei Ereignissen um Teilmengen von  $\Omega$ .

**Def. 1.3 (Ereignis, event)**

Gewisse  $A \subset \Omega$  heißen **Ereignis** (event).  $A$  tritt ein, falls sich bei Versuchsdurchführung ein  $\omega \in A$  ergibt.

Mit  $A$  und  $B$  sind auch  $\bar{A} = A^c := \Omega \setminus A$ ,  $A \cap B$ ,  $A \cup B$  Ereignisse.

$\Omega$  heißt **sicheres Ereignis** (tritt also immer ein)

$\emptyset$  heißt **unmögliches Ereignis** (kann nie eintreten)

$A^c$  heißt **Komplementärereignis**

$A, B$  heißen **unvereinbar** oder **disjunkt** (disjoint), falls  $A \cap B = \emptyset$

$\{\omega\}$  heißt **Elementarereignis** (singleton).

Vereinbarung: Falls nichts anderes gesagt wird, verwenden wir " $\subset$ " im Sinne von " $\subseteq$ ".

Bei abzählbarem  $\Omega$  ist jede Teilmenge als Ereignis zugelassen. Um bei überabzählbarem  $\Omega$  Widersprüche auszuschließen, beschränkt man sich auf sog. *Ereignis- $\sigma$ -Algebren*. Dies sind ausreichend umfangreiche Mengensysteme von Ereignissen, die  $\Omega$  enthalten und bzgl. Komplementbildung und abzählbarer Vereinigung (und damit auch abzählbarem Durchschnitt) abgeschlossen sind.

**Def. 1.4** ( $\sigma$ -Algebra  $\mathfrak{F}$ ,  $\sigma$ -field, event space)

Eine  $\sigma$ -Algebra  $\mathfrak{F}$  ist ein Mengensystem  $\mathfrak{F} \subset \mathcal{P}(\Omega)$  mit folgenden Eigenschaften

$$(A1) \quad \Omega \in \mathfrak{F}$$

$$(A2) \quad \text{mit } A \in \mathfrak{F} \text{ folgt auch } A^c \in \mathfrak{F}$$

$$(A3) \quad \text{falls } A_i \in \mathfrak{F}, i = 1, 2, \dots, \text{ dann ist auch } \bigcup_{i=1}^{\infty} A_i \in \mathfrak{F}.$$

Eine genaue Diskussion der Eigenschaften von  $\sigma$ -Algebren erfolgt in weiterführenden Vorlesungen. Hier werden nur die wichtigsten Resultate zusammengestellt. Details entnehme man z. B. [Williams (1991)], ch. 1.

### Beispiel 1.5

1. Die Potenzmenge  $\mathcal{P}(\Omega)$  ist die größte  $\sigma$ -Algebra auf  $\Omega$ .
2.  $\{\emptyset, \Omega\}$  ist die kleinste  $\sigma$ -Algebra auf  $\Omega$ .
3.  $\{\emptyset, A, A^c, \Omega\}$  ist die kleinste  $\sigma$ -Algebra, die eine nichtleere Menge  $A \neq \Omega$  enthält.

### Bemerkung 1.6

1. Das Präfix  $\sigma$ - weist darauf hin, dass die Eigenschaft (A3) für abzählbar unendlich viele Mengen gilt.
2. Natürlich gilt die Aussage (A3) auch für endliche Vereinigungen. Dazu beachtet man  $\emptyset \in \mathfrak{F}$  (wegen (A1) und (A2)) und setzt  $A_k := \emptyset$ , z. B. für  $k > n$ .
3. Mit den de Morgan'schen Gesetzen

$$\left( \bigcup_{i \in I} A_i \right)^c = \bigcap_{i \in I} A_i^c \quad \text{und} \quad \left( \bigcap_{i \in I} A_i \right)^c = \bigcup_{i \in I} A_i^c,$$

die für beliebige Indexmengen  $I$  gelten, folgt aus (A2) und (A3):

falls  $A_i \in \mathfrak{F}$ ,  $i = 1, 2, \dots$ , dann ist auch  $\bigcap_{i=1}^{\infty} A_i \in \mathfrak{F}$ .

**Satz 1.7 (Erzeugung von  $\sigma$ -Algebren, Durchschnittseigenschaft)**

1. Der Durchschnitt beliebig vieler  $\sigma$ -Algebren ist wieder eine  $\sigma$ -Algebra.
2. Zu jedem Mengensystem  $\mathfrak{A} \subset \mathcal{P}(\Omega)$  gibt es eine eindeutig bestimmte kleinste  $\sigma$ -Algebra  $\mathfrak{F}$  mit  $\mathfrak{A} \subset \mathfrak{F}$ , die von  $\mathfrak{A}$  **erzeugte**  $\sigma$ -Algebra  $\mathfrak{F}$ .

**Def. 1.8 (Borel -  $\sigma$ -Algebra)**

Die von den offenen Mengen in  $\mathbb{R}$  erzeugte  $\sigma$ -Algebra heißt **Borel** -  $\sigma$ -Algebra.

**Satz 1.9 (Borel -  $\sigma$ -Algebra)**

1.  $\mathfrak{B}$  enthält u.a. alle reellen Intervalle.
2.  $\mathfrak{B}$  wird auch von den Intervallen  $(-\infty, x]$ ,  $x \in \mathbb{R}$ , erzeugt.
3. Zwar ist  $\mathfrak{B} \neq \mathcal{P}(\mathbb{R})$ , dennoch erfordert die Konstruktion einer Menge  $A \subset \mathbb{R}$  und  $A \notin \mathfrak{B}$  erhebliche Mühe.

Es genügt für praktische Anwendungen, Wahrscheinlichkeiten auf  $\sigma$ -Algebren  $\mathfrak{F}$ , d. h. geeigneten Untersystemen der Potenzmenge von  $\Omega$  zu definieren. Ab jetzt heißen nur noch solche Teilmengen  $C \subset \Omega$  *Ereignisse*, für die  $C \in \mathfrak{F}$  gilt. Weiterhin seien von nun an generell  $A, A_1, A_2, \dots \in \mathfrak{F}$  und  $B, B_1, B_2, \dots \in \mathfrak{F}$ .

## 1.2 Wahrscheinlichkeitsaxiome von Kolmogorov

**Def. 1.10 (Wahrscheinlichkeit)**

Die **Wahrscheinlichkeit** ist eine Funktion  $P : \mathfrak{F} \rightarrow \mathbb{R}$  für Ereignisse  $A, A_1, A_2, \dots$  aus einer  $\sigma$ -Algebra  $\mathfrak{F}$  auf  $\Omega$  mit folgenden Eigenschaften

$$(P1) \quad P(A) \geq 0 \quad \text{für alle } A \in \mathfrak{F}$$

$$(P2) \quad P(\Omega) = 1$$

$$(P3) \quad P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i), \text{ falls } A_i \cap A_j = \emptyset \text{ für } i \neq j.$$

Man nennt  $P$  auch **Wahrscheinlichkeitsmaß** oder kurz **W-Maß**.

In der Wahrscheinlichkeitsrechnung wird stets davon ausgegangen, dass die Funktion  $P$  bekannt ist. Bei praxisrelevanten Problemen ist diese Annahme meist nicht erfüllt. Mit den Methoden der Mathematischen Statistik kann man aber Schätzungen für die Wahrscheinlichkeiten  $P(A)$ , jedenfalls für gewisse  $A \in \mathfrak{F}$  gewinnen.

**Def. 1.11 (Wahrscheinlichkeitsraum)**

Das Tripel  $(\Omega, \mathfrak{F}, P)$  nennt man **Wahrscheinlichkeitsraum** oder kurz **W-Raum**.  $\Omega$  heißt **diskret**, falls es höchstens abzählbar unendlich viele Elemente hat.

**Beispiel 1.12 (Wahrscheinlichkeitsräume)**

1. Würfeln:  $\Omega = \{1, 2, 3, 4, 5, 6\}$ ,  $P\{i\} = \frac{1}{6}$ ,  $i = 1, \dots, 6$

2. Münzwurf (Standardmodellierung):  
 $\Omega = \{0, 1\}$ ,  $P\{1\} = P(\text{Kopf}) = \frac{1}{2}$ ,  $P\{0\} = P(\text{Wappen}) = \frac{1}{2}$

3. Münzwurf (Präzisierung):

Beim Münzwurf ist auch die Situation denkbar, dass die Münze auf der Kante stehenbleibt. Die Menge **aller** möglichen Ergebnisse lautet dann  $\tilde{\Omega} = \{0, 1, 2\}$ , wobei  $P\{1\} = \frac{1}{2}$ ,  $P\{0\} = \frac{1}{2}$ ,  $P\{2\} = P(\text{Kante}) = 0$ .

Hierbei bedeutet das Ergebnis "2", dass die Münze weder auf Kopf noch auf Wappen fällt. Indem man die Wahrscheinlichkeit (Wkt) für das Ereignis  $\{2\}$  auf 0 festlegt, werden im W-Raum  $\tilde{\Omega}$  alle wichtigen Aussagen genauso gelten wie in dem erweiterten  $\Omega$  mit dem zusätzlichen Ergebnis 2. Deswegen arbeitet man in der Regel gleich mit  $\Omega$  statt  $\tilde{\Omega}$ . Ähnliches gilt für viele andere W-Räume.

Etwas unpräzise formuliert kann man vielleicht sagen, dass Ereignisse mit Wkt 0 in diskreten W-Räumen praktisch nicht auftreten.

4. In überabzählbaren W-Räumen, auf denen z. B. normalverteilte Merkmale modelliert werden, kann die Wkt für jedes Elementarereignis  $\omega \in \Omega$  gleich Null sein, was nur scheinbar ein Widerspruch zu dem vorher Gesagten ist. Auf dieses Paradox wird später nochmals eingegangen.

5. Würfeln bzw. Münzwurf sind Beispiele sog. Laplace-W-Räume, vgl. Def. 1.16.

**Satz 1.13 (Folgerungen)**

1. Setzt man  $A_i := \emptyset$ ,  $i = n + 1, n + 2, \dots$ , so gilt (P3) auch für endliche disjunkte Vereinigungen, d. h. für  $A_1, \dots, A_n \in \mathfrak{F}$

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i), \text{ falls } A_i \cap A_j = \emptyset, i \neq j.$$

2. Weiterhin folgt aus den drei Axiomen für  $A, B \in \mathfrak{F}$

$$0 \leq P(A) \leq 1, \quad P(\emptyset) = 0$$

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

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

3. Besonders wichtig ist der **Additionssatz**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B), \quad (1.1)$$

den man sich am besten in einem Venn-Diagramm für beliebige Mengen  $A, B \in \mathfrak{F}$  verdeutlicht.

**Satz 1.14 (Ungleichungen, Boole's inequalities)**

Über den Additionssatz folgt für  $A \subset B$  zunächst  $A \cap B = A$  und somit

$$P(B) = P(A) + P(B \cap A^c) \geq P(A); \quad A, B \in \mathfrak{F}.$$

Für beliebige  $A, B \in \mathfrak{F}$  gelten die **Ungleichungen von Boole**

$$\begin{aligned} P(A) + P(B) &\geq P(A \cup B) \\ &\geq \max\{P(A), P(B)\} \\ &\geq P(A \cap B) \\ &\geq P(A) + P(B) - 1. \end{aligned}$$

Die Aussage des folgenden Satzes spielt bei vielen Beweisen eine Rolle.

**Satz 1.15 (Folgen von Ereignissen)**

Seien  $A_{n+1} \subset A_n \in \mathfrak{F}$ ,  $n = 1, 2, \dots$  und  $\bigcap_{n=1}^{\infty} A_n = \emptyset$ , dann gilt folgender

**Stetigkeitssatz**

$$\lim_{n \rightarrow \infty} P(A_n) = 0.$$

**Def. 1.16 (Laplace-Raum; Abzählregel)**

Sind bei endlichem  $\Omega$  alle Elementarereignisse gleichwahrscheinlich, d. h.

$$P(\{\omega\}) = \frac{1}{|\Omega|}, \quad \text{für alle } \omega \in \Omega,$$

so spricht man von einem **Laplace-Wahrscheinlichkeitsraum** oder einfach von der **Laplace-Annahme**.

Als  $\sigma$ -Algebra wählt man in diesem Fall  $\mathfrak{F} = \mathcal{P}(\Omega)$ . Damit folgt aus dem Axiom (P3) für beliebige Ereignisse in Laplace-Räumen die sog. **Abzählregel**:

$$P(A) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Ergebnisse}} = \frac{|A|}{|\Omega|}. \quad (1.2)$$

## 1.3 Kombinatorik, Abzählen

Erste systematische Untersuchungen zu Fragen der Wahrscheinlichkeitstheorie wurden im 17. Jahrhundert vor allem im Zusammenhang mit Glücksspielen durchgeführt (Bernoulli, Fermat, Laplace, Pascal, ...). Unter anderem spielten damals Abzählaufgaben eine wichtige Rolle. Nachdem dieser Aufgabentyp in der Kollegstufe ausführlich behandelt wird, sollen hier nur die allerwichtigsten Aspekte zur Sprache kommen. Ausführliche Darstellungen findet man u.a. in [Chung (1979)], [Stirzaker (1994)], jeweils ch. 3 (Counting), [Isaac (1995)], ch. 2 (How to Count ...) oder [Henze (1997)].

Viele Abzählaufgaben basieren auf

**Satz 1.17 (Multiplikationsregel der Kombinatorik)**

Es sei eine mehrfache Auswahl zu treffen, wobei es  $m_1$  Möglichkeiten für die erste Wahl,  $m_2$  Möglichkeiten für die zweite Wahl,  $m_3$  für die dritte usw. gibt. Können alle Möglichkeiten nach Belieben kombiniert werden, so lautet die Gesamtzahl aller möglichen Fälle

$$m_1 \cdot m_2 \cdot m_3 \cdot \dots$$

Wichtigste Bausteine von Kombinatorikformeln sind die Fakultät und Binomialkoeffizienten.

**Def. 1.18 (Permutationen)**

Für  $n \in \mathbb{N}_0$  gibt

$$n! = n \cdot (n - 1) \cdot \dots \cdot 1, \quad 0! = 1,$$

die Anzahl der möglichen Permutationen (= Vertauschungen) von  $n$  verschiedenen Objekten an.

Zum Beispiel gibt es für  $n$  Personen  $(n - 1)!$  Möglichkeiten im Kreis zu sitzen. Dabei sind zwei Kreisarrangements gleich, wenn jede Person in jeder Anordnung denselben Nachbarn hat.

**Satz 1.19 (Variationen mit Wiederholung)**

Nach der Multiplikationsregel 1.17 gibt es  $n^k$  Möglichkeiten, aus einer Menge von  $n$  Elementen  $k$  Elemente unter Beachtung der Reihenfolge mit Zurücklegen zu ziehen.

**Satz 1.20 (Kombinationen ohne Wiederholung)**

Es gibt

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1) \cdot \dots \cdot (n-k+1)}{1 \cdot 2 \cdot \dots \cdot k}, \quad n, k \in \mathbb{N}_0, \quad k \leq n$$

$k$ -elementige Teilmengen einer Menge von  $n$  Elementen.

Es folgen zwei Anwendungsbeispiele.

**Satz 1.21 (Ziehen mit Zurücklegen)**

In einer Urne seien  $N$  verschiedene Kugeln. Davon seien  $M$  rot gefärbt und der Rest weiß ( $0 < M < N$ ).  $A_n^m$  sei das Ereignis, dass beim  $n$ -maligen (unabhängigen) Ziehen einer Kugel mit Zurücklegen genau  $m$  rote Kugeln auftreten,  $0 \leq m \leq n$ . Nach der Abzählregel 1.16, der Multiplikationsregel 1.17 und Satz 1.19 erhält man

$$P(A_n^m) = \binom{n}{m} \frac{M^m (N - M)^{n-m}}{N^n}.$$

Die Gesamtzahl aller Möglichkeiten bei  $n$  Zügen mit Zurücklegen unter Beachtung der Reihenfolge (wir denken uns die Kugeln von 1 bis  $N$  nummeriert) ist  $N^n$ . Seien die

Nummern der roten Kugeln  $1, \dots, M$  und die der weißen  $M + 1, \dots, N$ . Man erhält zum Beispiel genau  $m$  rote Kugeln, wenn zuerst  $m$  rote und danach  $n - m$  weiße Kugeln gezogen werden. Nach der Multiplikationsregel hat man dafür  $M^m(N - M)^{n-m}$  Möglichkeiten. Es sind aber auch andere Farbreihenfolgen möglich. Man könnte etwa zunächst eine rote, dann  $n - m$  weiße und zum Schluß die restlichen  $m - 1$  roten Kugeln ziehen. Markieren wir diejenigen  $m$  Stellen, an denen eine rote Kugel gezogen wird mit 1, und die anderen mit 0, so erhalten wir binäre  $n$ -Tupel, die mit den  $m$ -elementigen Teilmengen einer  $n$ -elementigen Menge identifiziert werden können. Davon gibt es eine Anzahl von genau  $\binom{n}{m}$ . Mit der Abzählregel erhält man die Formel von Satz 1.21. Ähnliche Überlegungen führen zur Herleitung der Wahrscheinlichkeitsfunktion der Binomialverteilung; siehe u.a. Formel (1.16).

In der Fragestellung kommt es gar nicht darauf an, ob zuerst die roten oder andere Kugeln gezogen werden. Trotzdem dürfen wir entsprechende Formeln - mit Beachtung der Reihenfolge - benutzen, wenn dies sowohl beim Abzählen der möglichen als auch der für  $A_n^m$  günstigen Fälle geschieht.

**Satz 1.22 (Ziehen ohne Zurücklegen)**

*Wie vorher seien in einer Urne  $N$  Kugeln, davon  $0 < M < N$  rot gefärbt und der Rest weiß.  $B_n^m$  sei das Ereignis, dass beim  $n$ -maligen (unabhängigen) Ziehen einer Kugel ohne Zurücklegen genau  $m$  rote Kugeln auftreten,  $0 \leq m \leq \min\{M, n\}$ . Wieder mit der Abzähl- und Multiplikationsregel sowie diesmal mit Satz 1.20 erhält man*

$$P(B_n^m) = \frac{\binom{M}{m} \binom{N - M}{n - m}}{\binom{N}{n}}.$$

Hier kann man die Formeln über Kombinationen ohne Wiederholung heranziehen. Dort kommt es ohnehin bei den Ergebnissen nicht auf die Reihenfolge der Elemente an.

Als mögliche Ergebnisse beim Ziehen ohne Zurücklegen kann man etwa alle  $n$ -elementigen Teilmengen einer  $N$ -elementigen Grundmenge (Kugeln mit Nummern  $1, \dots, N$ ) ansehen. Bei der Definition von Mengen spielt bekanntlich die Reihenfolge der Elemente in der Auflistung keine Rolle.

Zum Abzählen der für  $B_n^m$  günstigen Ergebnisse denken wir uns wieder die roten Kugeln als von  $1, \dots, M$  durchnummeriert, die  $N - M$  weißen von  $M + 1, \dots, N$ .  $m$  rote Kugeln treten beim  $n$ -maligen Ziehen ohne Zurücklegen genau dann auf, wenn am Ende Kugeln mit genau  $m$  verschiedenen Nummern aus  $\{1, \dots, M\}$  und  $n - m$  verschiedenen Nummern aus  $\{M + 1, \dots, N\}$  vorliegen. Gemäß Satz 1.20 ist die Anzahl dieser Ergebnisse  $\binom{M}{m}$  bzw.  $\binom{N - M}{n - m}$ .



Ein typisches Beispiel für 'Ziehen ohne Zurücklegen' ist etwa eine Ausspielung beim Lotto "6 aus 49":

- $N = 49$  Gesamtzahl der Kugeln in der Trommel
- $M = 6$  von einer Person  $P$  (am Freitag) angekreuzte Zahlen  
(in Gedanken rot gefärbt)
- $n = 6$  am Samstagabend gezogene Zahlen
- $m$  'Richtige' der Person  $P$   
(Anzahl der gezogenen 'roten Kugeln').

## 1.4 Unabhängigkeit, bedingte Wahrscheinlichkeit

### Def. 1.23 (Bedingte Wahrscheinlichkeit, conditional probability)

Seien  $A, B$  Ereignisse in  $\mathfrak{F}$ . Für  $P(B) \neq 0$  definiert

$$P(A|B) = P_B(A) := \frac{P(A \cap B)}{P(B)} \quad (1.3)$$

die Wahrscheinlichkeit von  $A$  unter der **Bedingung**  $B$ . Die folgende Identität, in der  $P(A|B)$  auf der rechten Seite vorkommt, ist für  $P(B) > 0$  wohldefiniert. Um in Zukunft die Diskussion von Sonderfällen zu vermeiden, legen wir zusätzlich für  $B \in \mathfrak{F}$  und  $P(B) = 0$  fest

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

Für ein fest gewähltes  $B$  mit  $P(B) > 0$  und variables  $A \in \mathfrak{F}$  definiert (1.3) sogar einen neuen Wahrscheinlichkeitsraum.

### Satz 1.24 (Wahrscheinlichkeit unter der Bedingung $B$ )

Für  $B \in \mathfrak{F}$  mit  $P(B) > 0$ ,  $\mathfrak{F}_B := \{ A_B \mid A_B := A \cap B, A \in \mathfrak{F} \}$  und  $P_B(A_B) := P(A|B)$  ist

$$(B, \mathfrak{F}_B, P_B)$$

ein Wahrscheinlichkeitsraum.

Der nächste Satz ist eine einfache Folgerung der vorhergehenden Definition und doch von weitreichender Bedeutung bei späteren Überlegungen.

**Satz 1.25 (Bedingte Zerlegung, conditioning)**

Für Ereignisse  $A, B \in \mathfrak{F}$  mit  $0 < P(B) < 1$  gilt

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) = \\ &= P(A|B)P(B) + P(A|B^c)P(B^c). \end{aligned} \quad (1.4)$$

Dies ist offensichtlich ein Spezialfall vom

**Satz 1.26 (von der totalen Wahrscheinlichkeit, total probability)**

Seien  $B_1, B_2, \dots$  höchstens abzählbar viele, paarweise disjunkte Ereignisse in  $\mathfrak{F}$  und  $A \subset \bigcup_{i=1}^{\infty} B_i$ ,  $A \in \mathfrak{F}$ , dann gilt

$$P(A) = \sum_{i=1}^{\infty} P(B_i) P(A|B_i). \quad (1.5)$$

bzw. für endlich viele  $B_i$

$$P(A) = \sum_{i=1}^n P(B_i) P(A|B_i). \quad (1.6)$$

Der folgende Satz ist eine direkte Konsequenz. Er wird nur für endlich viele  $B_i$  formuliert:

**Satz 1.27 (von Bayes)**

Seien  $A, B_1, \dots, B_n \in \mathfrak{F}$ ,  $B_1, \dots, B_n$  paarweise disjunkt,  $A \subset \bigcup_{i=1}^n B_i$  und  $P(A) \neq 0$ . Weiterhin seien die bedingten Wahrscheinlichkeiten  $P(A|B_i)$  gegeben. Dann gilt

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^n P(B_i)P(A|B_i)} \quad \forall j = 1, \dots, n.$$

Für  $P(B) > 0$  kann folgende Situation eintreten:

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

Dies motiviert die Definition

**Def. 1.28 (Unabhängigkeit)**

1. Zwei Ereignisse  $A$  und  $B \in \mathfrak{F}$  heißen **unabhängig** (independent), falls

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

2. Eine (nicht unbedingt abzählbare) Familie von Ereignissen  $A_i \in \mathfrak{F}$ ,  $i \in I$ , heißt (stochastisch) **unabhängig**, falls

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j), \quad (1.7)$$

für alle endlichen Indermengen  $J \subset I$ .

3. Eine Familie von Ereignissen  $A_i \in \mathfrak{F}$ ,  $i \in I$ , heißt **paarweise unabhängig**, falls

$$P(A_i \cap A_j) = P(A_i)P(A_j), \quad \text{für alle } i \neq j.$$

Paarweise unabhängige Ereignisse sind nicht notwendig (stochastisch) unabhängig.

**Folgerung 1.29**

Seien  $A, B \in \mathfrak{F}$ .

1. Für alle  $A \in \mathfrak{F}$  sind  $\emptyset$  und  $A$  unabhängig.
2. Für alle  $A \in \mathfrak{F}$  sind  $\Omega$  und  $A$  unabhängig.
3. Für alle  $A \in \mathfrak{F}$  mit  $0 < P(A) < 1$  sind  $A$  und  $A^c$  **nicht** unabhängig.
4. Falls  $A$  und  $B$  unabhängig, so auch  $A$  und  $B^c$  sowie  $A^c$  und  $B$  sowie  $A^c$  und  $B^c$ .
5. Für  $P(B) > 0$  sind  $A$  und  $B$  genau dann unabhängig, falls  $P(A|B) = P(A)$ .

**Beispiel 1.30 (Ruin des Spielers, gambler's ruin)**

Am Ende dieses Abschnitts soll zum ersten Mal ein Beispiel diskutiert werden, das uns in dieser Einführungsvorlesung noch öfter begegnen wird. Es handelt sich um den

**”Ruin des Spielers”**,

vgl. etwa [Stirzaker (1994)], pp. 39 und 145.

Sie betreten ein Spielkasino mit einem Kapital von  $k$  DM und setzen bei jedem Rouletteinsatz 1 DM auf das Ereignis  $R$ , dass die Kugel auf ”Rot” fällt. Im Falle  $R$  erhalten Sie mit dem Einsatz 2 DM zurück, andernfalls gehört Ihr Einsatz von 1 DM der Spielbank. Wir wissen  $p = P(R) = 18/37 < 1/2$ . Sie legen von vorneherein ein Zielkapital von  $K \geq k$  fest und beenden das Spiel, wenn Sie  $K$  DM besitzen oder wenn Sie alles verloren haben.

Gesucht ist zunächst  $p_k$ , die Wahrscheinlichkeit dafür, alles zu verlieren.

Lösung analog [Chung (1979)], pp. 242-245.

Nun sei allgemein  $0 < p < 1$ . Bezeichne  $A_k$  das Ereignis Ruin bei Anfangskapital  $k$ . Falls das erste Spiel mit ”Rot” endet, so ist dies dieselbe Situation als hätte man mit  $k+1$  DM begonnen. Falls die Kugel ”nicht auf Rot” fällt (d. h.  $R^c$ ), so hätte man gleich mit  $k-1$  DM beginnen können. Damit ist der Satz von der totalen Wahrscheinlichkeit bzgl. des Ereignisses  $R$  anwendbar und es gilt für  $0 < k < K$

$$\begin{aligned} P(A_k) &= P(R) P(A_k|R) + P(R^c) P(A_k|R^c) \\ &= P(R) P(A_{k+1}) + P(R^c) P(A_{k-1}) \\ p_k &= p p_{k+1} + (1-p) p_{k-1}. \end{aligned} \tag{1.8}$$

Wir lösen nun die Differenzgleichung (1.8) unter den trivialen Anfangsbedingungen

$$p_0 = 1 \quad \text{und} \quad p_K = 0.$$

In keinem dieser beiden Fälle findet ein Spiel statt: im ersten, weil kein Kapital vorhanden ist; im zweiten, weil schon alles gewonnen ist. Seien jetzt allgemein  $0 < p < 1$ ,  $r := (1-p)/p$  und  $d_k := p_k - p_{k+1}$ . Mit  $p_k = p p_{k+1} + (1-p) p_{k-1}$  folgt aus (1.8) durch Umordnen

$$\begin{aligned} p_k - p_{k+1} &= \frac{1-p}{p} (p_{k-1} - p_k) = \left( \frac{1-p}{p} \right)^k (1 - p_1) \\ \implies d_k &= r d_{k-1} = r^k d_0. \end{aligned}$$

Schrittweises Aufsummieren liefert mit der endlichen geometrischen Reihe

$$\begin{aligned} 1 = p_0 - p_K &= \sum_{k=0}^{K-1} (p_k - p_{k+1}) = \sum_{k=0}^{K-1} d_k = \sum_{k=0}^{K-1} r^k d_0 \\ &= \begin{cases} K d_0 & , \text{ falls } p = 1-p = \frac{1}{2} \\ \frac{1-r^K}{1-r} d_0 & , \text{ falls } p \neq 1-p \neq \frac{1}{2}, \end{cases} \end{aligned}$$

und daher

$$d_0 = \begin{cases} 1/K & , \text{ falls } p = \frac{1}{2} \\ \frac{1-r}{1-r^K} & , \text{ falls } p \neq \frac{1}{2}. \end{cases} \quad (1.9)$$

Wieder durch Summieren erhält man mit  $p_K = 0$

$$\begin{aligned} p_k = p_k - p_K &= \sum_{i=k}^{K-1} (p_i - p_{i+1}) = \sum_{i=k}^{K-1} d_i = \sum_{i=k}^{K-1} r^i d_0 = \\ &= \begin{cases} (K-k) d_0 & , \text{ falls } p = \frac{1}{2} \\ \frac{r^k - r^K}{1-r} d_0 & , \text{ falls } p \neq \frac{1}{2}. \end{cases} \end{aligned} \quad (1.10)$$

Zusammen ergeben (1.9) und (1.10) für  $0 \leq k \leq K$

$$p_k = \begin{cases} \frac{K-k}{K} & , \text{ falls } p = \frac{1}{2} \\ \frac{r^k - r^K}{1-r^K} & , \text{ falls } p \neq \frac{1}{2}. \end{cases} \quad (1.11)$$

Durch Vertauschen der Rollen von  $p$  und  $1-p$ , d. h.  $r \rightarrow 1/r$ , sowie dem Übergang von  $k \rightarrow K-k$  erhält man die Wahrscheinlichkeiten  $q_k$ , nämlich dafür, dass die Spielstrategie mit dem angestrebten Gewinn  $K$  endet:

$$q_k = \begin{cases} \frac{k}{K} & , \text{ falls } p = \frac{1}{2} \\ \frac{1-r^k}{1-r^K} & , \text{ falls } p \neq \frac{1}{2}. \end{cases} \quad (1.12)$$

Durch Einsetzen verifiziert man

$$p_k + q_k = 1$$

und erhält sofort den wichtigen

**Satz 1.31 (Endliches Ende des Spiels)**

*Das beschriebene Spiel endet mit Wahrscheinlichkeit 1 entweder im Ruin des Spielers oder mit dem Erreichen des Zielkapitals  $K$ .*

Diese Aussage ist alles andere als trivial und folgt nur wegen der separaten Herleitung der Wahrscheinlichkeiten  $q_k$ . (Warum darf man nicht einfach  $q_k := 1 - p_k$  setzen?) Zwar ist es denkbar, dass der Spieler jeweils abwechselnd ein Spiel gewinnt und verliert. In diesem Fall würde das Spiel für  $2 \leq k$  und  $K \geq k+2$  nie enden. Man kann aber zeigen,

dass derart spezielle Folgen von Spielausgängen nur mit Wahrscheinlichkeit 0 auftreten.

Ohne es zu bemerken, haben wir mit dem letzten Satz und der expliziten Angabe der  $p_k$  und  $q_k$  bereits die Analyse eines wichtigen Stochastischen Prozesses eingeleitet. Es handelt sich um die **einfache Irrfahrt** oder den (simple) **Random Walk**.

Dabei bewegt man sich in diskreter Zeit  $t = 0, 1, 2, \dots$  auf dem Gitter der ganzen Zahlen nach folgenden Regeln:

1. Zum Zeitpunkt  $t$  befindet sich der Prozeß im Zustand  $S_t \in \mathbb{Z}$ .
2. Der Ausgangszustand  $S_0 = s_0 \in \mathbb{Z}$  ist vorgegeben.
3. Auf den Zustand  $S_t$  kann entweder der Zustand  $S_{t+1} = S_t + 1$  oder der Zustand  $S_{t+1} = S_t - 1$  folgen.
4. Für  $0 < p < 1$  gilt

$$P(S_{t+1} = j + 1 \mid S_t = j) = p \quad \text{und} \quad P(S_{t+1} = j - 1 \mid S_t = j) = 1 - p.$$

5. Weiterhin gelten noch Unabhängigkeitsannahmen, die an dieser Stelle noch nicht präzisiert werden können.

Trägt man die Zustände  $S_t$  über der Zeit ab, so könnte der **Pfad** eines typischen Random Walks etwa folgendermaßen aussehen

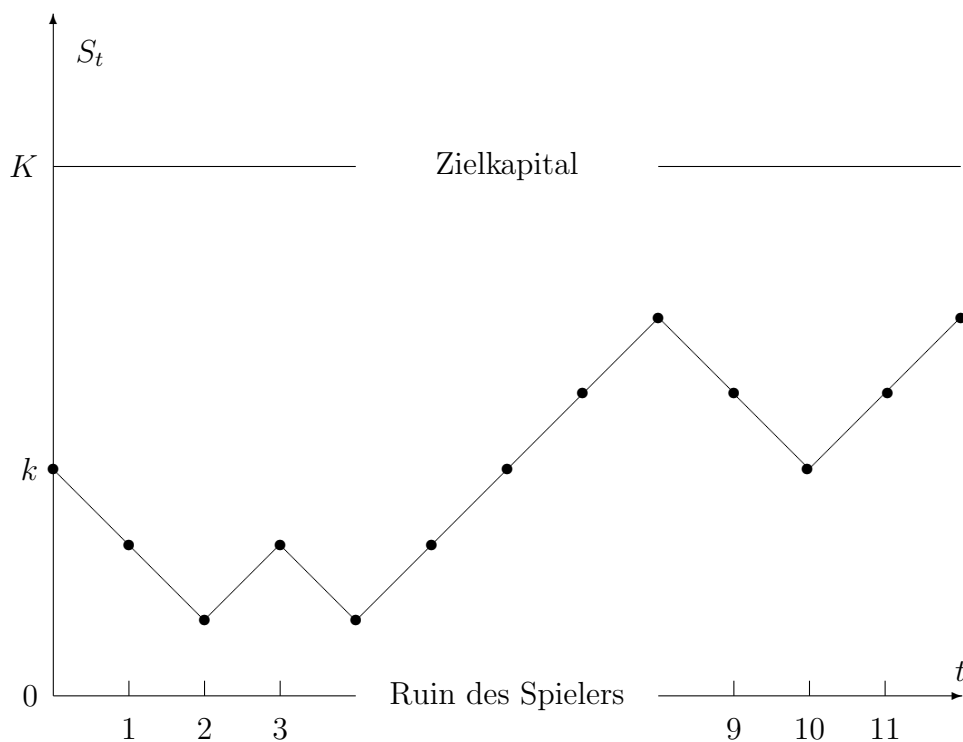


Bild 1: Random Walk

Der Zusammenhang zum Ruin des Spielers ergibt sich folgendermaßen

1. Startkapital :  $k = s_0 \geq 0$
2.  $p = P(\text{Gewinn}) = P(S_{t+1} = j + 1 | S_t = j)$
3.  $S_t = 0 \iff$  Ruin des Spielers
4.  $S_t = K \iff$  Spielende mit gewonnenem Zielkapital  $K$ .

Für den Random Walk kann man ähnliche und eine Reihe weiterer Fragen stellen wie beim Ruin des Spielers. Dies sind u.a.

1. Verläßt ein beliebiger Pfad des Random Walk ein vorgegebenes endliches Intervall  $[a, b]$ ,  $a < b$ ; womöglich sogar mit Wahrscheinlichkeit 1?
2. Trifft ein solcher Pfad einen beliebig vorgegebenen Wert  $z \in \mathbb{Z}$ ; und wie oft?
3.  $T_k$  sei die Zeit, bei der das Kapital des Spielers (ausgehend von  $S_0 = k$ ) zum ersten Mal einen der Werte  $S_{T_k} = 0$  oder  $S_{T_k} = K$  erreicht. Es ist u.a. zu klären, ob  $T_k$  endlich ist.
4. Was geschieht im Fall  $K \rightarrow \infty$ , wenn die Bank unbeschränktes Kapital besitzt?

Diesen und ähnlichen Fragen wenden wir uns im Kapitel 5 zu. Zuvor benötigen wir aber die Einführung weiterer wichtiger Begriffe der Wahrscheinlichkeitsrechnung, mit denen wir uns zunächst befassen.

## 1.5 Zufallsvariable

Die Festlegung von Wahrscheinlichkeiten für Ereignisse in den einfachen Laplace-Räumen wurde bereits diskutiert. In komplexeren Modellen benötigt man neue Hilfsmittel, nämlich sog. Zufallsvariable (vgl. auch die Ausführungen über Merkmale im Kapitel 'Empirische Statistik').

Oft ist man gar nicht an den Ergebnissen  $\omega \in \Omega$  selbst interessiert, sondern an deren Merkmalen; z. B. an der "Verteilung" von Größe oder Gewicht von Individuen  $\omega$  in einer Population  $\Omega$ . Es zeigt sich, dass die Festlegung interessierender Wahrscheinlichkeiten besonders einfach für bestimmte Abbildungen erfolgen kann.

### Def. 1.32 (Zufallsvariable, ZV)

Gegeben sei ein Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{F}, P)$ . Eine Abbildung

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

$$\text{mit } X^{-1}(B) \in \mathfrak{F} \text{ für alle Borelmengen } B \in \mathfrak{B} \text{ (vgl. Def. 1.8)} \quad (1.13)$$

heißt (eindimensionale) **Zufallsvariable** oder auch **Zufallsgröße** (random variable).

Die reelle Zahl  $x = X(\omega)$ , die sich bei Durchführung des Zufallsexperiments durch Auswerten von  $X$  ergibt, heißt **Realisierung** von  $X$ .

Das Wahrscheinlichkeitsmaß  $P_X$  auf  $(\mathbb{R}, \mathfrak{B})$  definiert durch

$$P_X(B) := P(X^{-1}(B)) \quad \text{für alle Borelmengen } B \in \mathfrak{B} \quad (1.14)$$

heißt **Verteilung** (distribution) von  $X$ .

### Bemerkung 1.33

1. Für eine ZV genügt es schon, wenn  $X^{-1}((-\infty, x]) \in \mathfrak{F}$ , für alle  $x \in \mathbb{R}$ .
2. Die Definition von ZV wird später verallgemeinert, zunächst auf Zufallsvektoren  $\mathbf{X} : (\Omega, \mathfrak{F}) \rightarrow (\mathbb{R}^n, \mathfrak{B}^n)$ .
3. Bei sog. **Wartezeitproblemen** kann die ZV  $X$  den Wert  $\infty$  annehmen. In diesem Einführungskurs bleibt dieser Fall ausgeschlossen, solange nicht explizit darauf Bezug genommen wird.
4. Falls  $\Omega$  höchstens abzählbar ist, und falls  $\mathfrak{F} = \mathcal{P}(\Omega)$ , so ist jede Abbildung  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable.
5. Für  $\mathfrak{F} = \{\emptyset, A, A^c, \Omega\}$  sind genau die Abbildungen

$$X(\omega) = \begin{cases} c & , \text{ falls } \omega \in A \\ d & , \text{ sonst,} \end{cases}$$

mit  $c, d \in \mathbb{R}$  Zufallsvariable.



6. Das letzte Beispiel demonstriert, wie einschränkend die Forderung (1.13) im Spezialfall sein kann. In einem Stochastik-Einführungskurs wollen wir aber diese abstrakte Messbarkeitseigenschaft nicht weiter verfolgen und fortan stets voraussetzen, dass die verwendeten ZV im Sinne von (1.13) wohldefiniert sind. Bei höchstens abzählbarem  $\Omega$  treten ohnehin keine Probleme auf, da für  $\mathfrak{F} = \mathcal{P}(\Omega)$  jede Abbildung messbar ist, und die Wahrscheinlichkeiten für die Elementarereignisse festgelegt werden können.

### 1.5.1 Diskrete Zufallsvariable

Für **diskrete** ZV nimmt  $X$  nur die Werte  $x_i$ ,  $i \in I$  an, wobei  $I$  höchstens abzählbar ist; z. B.  $I = \{1, \dots, n\}$ ,  $I = \mathbb{Z}$ ,  $I = \mathbb{N}$ ,  $I = \mathbb{N}_0$  usw. Häufig gilt:  $x_i \in \{0, 1, 2, \dots\}$ .

**Def. 1.34 (Diskrete Verteilung)**

$X$  sei eine diskrete ZV, und die Indexmenge  $I$  höchstens abzählbar unendlich.

$$p_i = P(X = x_i), \quad i \in I \quad (1.15)$$

mit  $p_i \geq 0$ ,  $i \in I$  und  $\sum_{i \in I} p_i = 1$

legen die in (1.14) definierte **Verteilung** von  $X$  eindeutig fest.

$$f(x) = f_X(x) = P(X = x) := \begin{cases} p_i & , \text{ falls } x = x_i, i \in I \\ 0 & , \text{ sonst,} \end{cases}$$

heißt **Wahrscheinlichkeitsfunktion** (*probability mass function, p.m.f.*) von  $X$ .

Auch umgekehrt bestimmt im diskreten Fall die Verteilung von  $X$  eindeutig die  $p_i > 0$ . Deswegen spricht man gelegentlich bei den  $p_i$  oder der Wahrscheinlichkeitsfunktion  $f_X$  von (diskreter) *Verteilung*, obwohl dieser Begriff genau genommen für das Wahrscheinlichkeitsmaß von (1.14) reserviert ist.

Vereinbarung:

1. An dieser Stelle scheint das Subskript bei  $f_X(x)$  noch überflüssig. Falls keine Verwechslungen zu befürchten sind, wird es auch weggelassen. In späteren Abschnitten, z. B. bei der Transformation von ZV, ermöglicht jedoch der Gebrauch des Subskripts eine verständlichere Darstellung.
2. Je nach Sachlage werden wir einmal die Schreibweise  $p_i$ ,  $i \in I$  gebrauchen, um an die diskreten Werte  $x_i$  zu erinnern, und ein anderes Mal die Darstellung über die Wahrscheinlichkeitsfunktion  $f_X$ , um eher den Bezug zur Zufallsvariablen  $X$  herauszuheben.
3. Man gibt die Wahrscheinlichkeitsfunktion in der Regel nur für Argumente  $x$  mit  $f_X(x) > 0$  an.
4. Im Zusammenhang mit diskreten ZV bedeutet die verkürzte Schreibweise  $\sum_i$  stets  $\sum_{i \in I}$ , also z. B.  $\sum_{i=0}^n$  oder  $\sum_{i=0}^{\infty}$  etc.

**Beispiel 1.35 (diskrete Verteilungen)****1. Würfeln:**

$$|\Omega| = 6, \quad X : \Omega \rightarrow \{1, 2, \dots, 6\}, \quad P(X = i) = \frac{1}{6}, \quad i = 1, \dots, 6.$$

**2. Bernoullisches Experiment:**  $\Omega = A \cup A^c$ ,  $X : \Omega \rightarrow \{0, 1\}$ .

$$P(A) = P(X = 1) = p, \quad P(A^c) = P(X = 0) = 1 - p, \quad 0 < p < 1.$$

Das Ereignis  $A$  steht etwa für "Gewinn" bei einem Glücksspiel oder für "Ausschuß" bei der Kontrolle einer Lieferung.

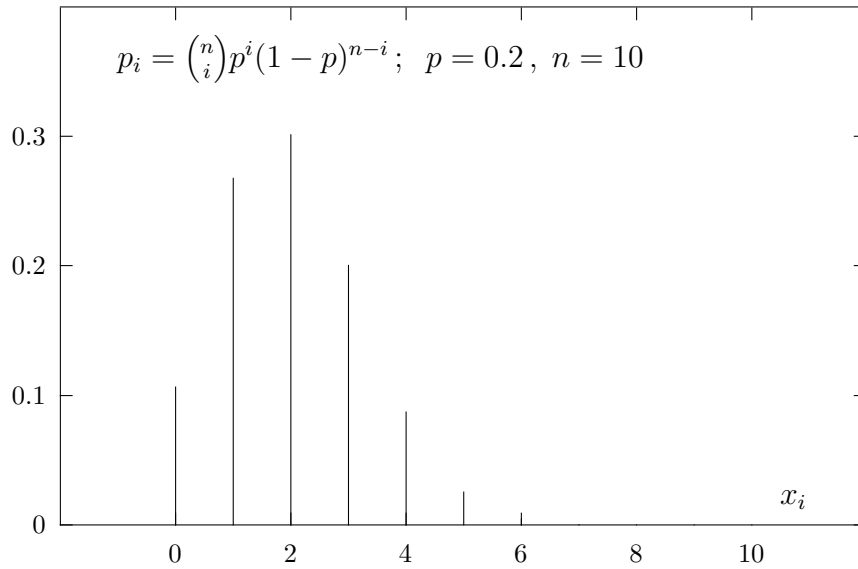
**3. Binomialverteilung:**  $B(n, p)$ 

$n$ -maliges Durchführen unabhängiger Bernoulliexperimente.  $P(A) = p$ ,  
 $X : \Omega \rightarrow \{0, 1, 2, \dots, n\}$ ;  $X$  bezeichnet die Anzahl des Auftretens von  $A$ .

$$p_i = P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}, \quad i = 0, 1, \dots, n, \quad 0 < p < 1. \quad (1.16)$$

Beweis von  $\sum_i p_i = 1$  über Binomische Formel.

Bedeutet  $A$  Ausschuß, so gibt  $X$  die zufällige Anzahl von nicht brauchbaren Teilen an, die in einer Stichprobe vom Umfang  $n$  enthalten sind.



**Bild 2:** Wahrscheinlichkeitsfunktion der Binomialverteilung  $B(10, 0.2)$

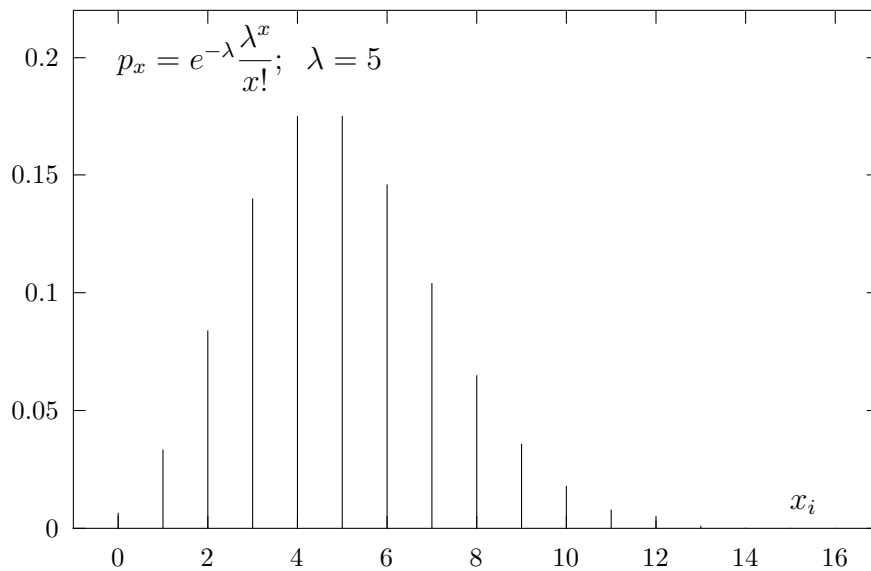
#### 4. Poisson-Verteilung: $Poi(\lambda)$

z. B. Anzahl der Telefonanrufe in einer Vermittlung pro Stunde.

$$X : \Omega \rightarrow \{0, 1, 2, \dots\}, \quad p_x = f(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \dots, \quad \lambda > 0, \quad (1.17)$$

(z. B.  $\lambda$  durchschnittliche Zahl von Anrufen).

Beweis von  $\sum_x p_x = 1$  über Taylorreihe der Exponentialfunktion.



**Bild 3:** Wahrscheinlichkeitsfunktion der Poisson-Verteilung  $Poi(5)$

**5. Geometrische Verteilung:**

$X$  sei die Anzahl der Fehlversuche bei (unabhängigen) Bernoulli-Experimenten, bis zum ersten Mal das Ereignis  $A = \text{Erfolg}$ , mit  $P(A) = p$ ,  $0 < p < 1$  eintritt, z. B.  $A = \text{”Rot”}$  beim Roulette. Somit ist  $X : \Omega \rightarrow \mathbb{N}_0$  mit der Verteilung

$$p_i = P(X = i) = p(1-p)^i, \quad i = 0, 1, 2, \dots \quad (1.18)$$

Beweis von  $\sum_i p_i = 1$  über die geometrische Reihe.  $X = 0$  bedeutet, dass bereits im 1. Versuch ”Erfolg” eintritt.

**6. Wartezeitprobleme**

Hier betrachtet man u.a. ZV  $X : \Omega \rightarrow \{0, 1, 2, \dots, \infty\}$  mit

$$\sum_{i=0}^{\infty} p_i = \bar{p} < 1 \quad \text{und} \quad P(X = \infty) = 1 - \bar{p} > 0. \quad (1.19)$$

Auf Beispiele dieser Art wird der Vollständigkeit halber gelegentlich hingewiesen. Falls jedoch nicht ausdrücklich ausgeschlossen, gelte hier stets  $\sum_i p_i = 1$ .

**1.5.2 Verteilungsfunktion**

Eine Standardmethode zur Festlegung von Wahrscheinlichkeiten für beliebige Zufallsvariable benützt die

**Def. 1.36 (Verteilungsfunktion, distribution function)**

Die Funktion  $F : \mathbb{R} \rightarrow \mathbb{R}$  mit

$$F(x) = F_X(x) := P(X \leq x) = P(\{\omega \mid X(\omega) \leq x\})$$

heißt **Verteilungsfunktion** der ZV  $X$ .

**Folgerung 1.37**

1.  $F(x)$  ist also die Wahrscheinlichkeit, dass der Wert der ZV bei Versuchsdurchführung kleiner gleich  $x$  ausfällt.
2.  $F$  ist wohldefiniert, da  $\{\omega \mid X(\omega) \leq x\} \in \mathfrak{F}$  für alle  $x \in \mathbb{R}$ .
3. Da die Borelsche  $\sigma$ -Algebra  $\mathfrak{B}$  auch von den Intervallen  $(-\infty, x]$ ,  $x \in \mathbb{R}$ , erzeugt wird, legt die Funktion  $F$  die Verteilung von  $X$  gemäß (1.14) eindeutig fest.
4.  $\bar{F}(x) := 1 - F(x) = P(X > x)$  wird auch ”**tail probability**” genannt.

5. Nach dem Additionssatz gilt für  $a < b$

$$P(a < X \leq b) = F(b) - F(a). \quad (1.20)$$

Man beachte den Unterschied zwischen  $a < X$  und  $a \leq X$  bei diskreten ZV.

**Satz 1.38 (Eigenschaften)**

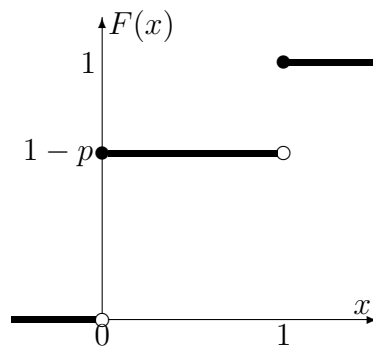
Für die Verteilungsfunktion  $F$  gilt

1.  $F(x) \leq F(y)$ , falls  $x \leq y$  (monoton nicht fallend)
2.  $F(-\infty) := \lim_{x \rightarrow -\infty} F(x) = 0$ ,  $F(\infty) := \lim_{x \rightarrow \infty} F(x) = 1$
3.  $F$  ist rechtsseitig stetig.

Manche Autoren definieren  $F(x) := P(X < x)$ . In diesem Fall ist  $F$  linksseitig stetig. Für diskrete ZV erhält man mit unserer Definition

$$F(x) = \sum_{i: x_i \leq x} p_i = \sum_{i: x_i \leq x} f(x_i), \quad (1.21)$$

also eine Treppenfunktion mit Sprüngen  $p_i = P(X = x_i)$  in  $x_i$ .



**Bild 4:** Verteilungsfunktion für diskretes  $X$ :  $P(X = 0) = 1 - p$ ,  $P(X = 1) = p$

Bei diskreten ZV gibt es einen eindeutigen Zusammenhang zwischen den Wahrscheinlichkeiten  $p_i = f_X(x_i) > 0$ ,  $i \in I$  und der Verteilungsfunktion  $F$ , wobei die  $p_i$  anschaulicher zu interpretieren sind; vgl. etwa die Verwandtschaft zu relativen Häufigkeiten.

Entsprechende Analogien gelten für eine andere wichtige Klasse von Zufallsvariablen, die Werte in einem ganzen Intervall annehmen können.

### 1.5.3 Stetige Zufallsvariable (ZV)

Beispiele stetiger ZV sind u.a.

- Körpergröße einer Grundgesamtheit  $\Omega$ ,  $X : \Omega \rightarrow (0, \infty)$ .
- Kontostand einer Gruppe von Kunden  $\Omega$ ,  $X : \Omega \rightarrow \mathbb{R}$ .

**Def. 1.39 (Dichtefunktion, density)**

Eine Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$  heißt **Dichte**, falls

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

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

**Def. 1.40 (Stetige ZV)**

$X : \Omega \rightarrow \mathbb{R}$  sei eine ZV. Man nennt  $X$  **stetig**, falls es eine (integrierbare) Dichte  $f_X$  gibt, so dass für alle  $x \in \mathbb{R}$

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt.$$

Vereinbarung:

1. Wir schreiben wieder  $F$  statt  $F_X$  bzw.  $f$  statt  $f_X$ , falls klar ist, zu welcher Zufallsvariablen  $F$  bzw.  $f$  gehören.
2. Die Verteilungsfunktion  $F_X$ , und damit die Verteilung von  $X$  gemäß (1.14), werden eindeutig durch eine Dichte  $f_X$  festgelegt. Wenn im folgenden von einer (stetigen) *Verteilung* die Rede ist, so genügt es also zur Charakterisierung, eine Dichte  $f_X$  anzugeben.
3. In dieser Vorlesung werden nur Dichten betrachtet, die stetig oder stückweise stetig sind (also insbesondere Regelfunktionen).

Für diesen praktisch wichtigen Sonderfall folgt aus dem Hauptsatz der Differential- und Integralrechnung

**Satz 1.41 (Beziehung Verteilungsfunktion - Dichte)**

Falls die stetige Verteilungsfunktion  $F$  einer ZV  $X$  bis auf endlich viele Stellen stetig differenzierbar ist, so ist  $X$  stetig mit der Dichte

$$f(x) = f_X(x) = F'_X(x) = F'(x). \quad (1.22)$$

Die Festlegung von  $f$  an den Sprungstellen kann für  $f(x) \geq 0$  beliebig erfolgen.

**Folgerung 1.42**

1. Die Verteilungsfunktion einer stetigen ZV ist stetig.
2. Nach dem Additionssatz gilt für  $a < b$

$$P(a < X \leq b) = F(b) - F(a) = \int_a^b f(t) dt = P(a \leq X \leq b). \quad (1.23)$$

3. Für stetige ZV spielt also wegen

$$P(X = a) = P(a \leq X \leq a) = \int_a^a f(t) dt = 0$$

die Festlegung der Dichte an einzelnen Punkten keine Rolle.

4. Bei stetigen ZV kann man die Verteilungsfunktion eindeutig aus einer gegebenen Dichte gewinnen. Umgekehrt ist die Dichte einer stetigen ZV bei gegebener Verteilungsfunktion  $F_X$  punktweise nicht eindeutig festgelegt. Für Dichten  $f_X$  und  $\tilde{f}_X$  von  $X$  gilt allerdings für alle  $-\infty \leq a \leq b \leq \infty$

$$\int_a^b f_X(x) dx = \int_a^b \tilde{f}_X(x) dx.$$

5. Die Analogie zwischen der hier definierten und der empirischen Dichte (Histogramm) in der beschreibenden Statistik ist offensichtlich.
6. Wegen Axiom (P3) von Def. 1.10 gilt für diskrete und stetige ZV sowie für beliebige Borelmengen  $B \subset \mathbb{R}$

$$P(X \in B) = \begin{cases} \sum_{x \in B} f_X(x) & , \text{ falls } X \text{ diskret} \\ \int_{x \in B} f_X(x) dx & , \text{ falls } X \text{ stetig,} \end{cases} \quad (1.24)$$

wobei etwa für  $B = [a, b] \cup [c, d]$  mit  $a < b < c < d$

$$\int_{x \in B} f_X(x) dx = \int_a^b f_X(x) dx + \int_c^d f_X(x) dx$$

zu lesen ist.

7. Trotz vieler formaler Ähnlichkeiten bestehen wesentliche Unterschiede zwischen der Wahrscheinlichkeitsfunktion diskreter und der Dichtefunktion stetiger ZV. Wir haben gesehen, dass die Aussage

$$P(X = x) = f_X(x)$$

für diskrete ZV richtig und für stetige i. a. falsch ist.

Häufige Wertebereiche für stetige ZV sind u.a. die Intervalle  $[a, b]$ ,  $a < b$ ,  $[0, \infty)$ ,  $\mathbb{R}$ . Jeder Wert im entsprechenden Intervall ist möglich. Die Kuriosität, dass ein konkreter Wert  $x^*$  nur mit Wahrscheinlichkeit 0 auftritt, soll an einem Beispiel erläutert werden. Für  $X =$  Körpergewicht würde dies etwa bedeuten

$$P(X = 70) = 0, \quad \text{aber} \quad P(69.999 \leq X \leq 70.001) > 0.$$

Der Fall, dass jemand genau 70.0000... kg wiegt, tritt also nur mit Wahrscheinlichkeit 0 auf. Für sinnvoll gewählte, nichtleere Intervalle erhält man dagegen in der Regel positive Wahrscheinlichkeiten.

**Def. 1.43 (Identisch verteilt, identically distributed)**

Zwei ZV  $X$  und  $Y$  heißen **identisch verteilt** ( $X \sim Y$ ), falls die entsprechenden Verteilungen übereinstimmen, d. h.

$$P(X^{-1}(B)) = P(Y^{-1}(B)) \quad \text{für alle Borelmengen } B \in \mathfrak{B}$$

oder äquivalent dazu, falls

$$F_X(z) = F_Y(z) \quad \text{für alle } z \in \mathbb{R}.$$

Vor der Diskussion einiger Beispiele noch eine weitere Definition:

**Def. 1.44 (Indikatorfunktion)**

$A$  sei eine nichtleere Teilmenge von  $\Omega$ .  $1_A : \Omega \rightarrow \mathbb{R}$  mit

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

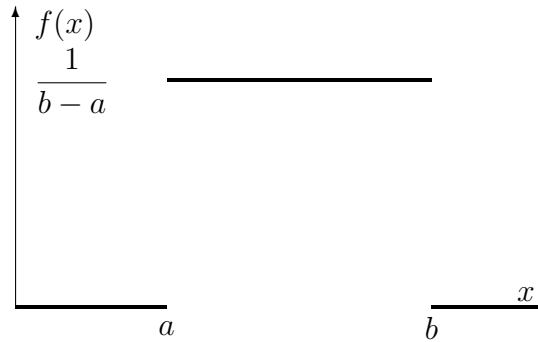
heißt **Indikatorfunktion** von  $A$ .

**Beispiel 1.45 (stetige Verteilungen)**



1. **Gleichverteilung** (uniform distribution)  $U(a, b)$  in  $(a, b)$ ,  $a < b$ .  
Man spricht auch von der Rechteckverteilung. Diese hat die Dichte

$$f(x) = \frac{1}{b-a} 1_{(a,b)}(x) = \begin{cases} \frac{1}{b-a} & , \text{ falls } a < x < b \\ 0 & , \text{ sonst,} \end{cases} \quad (1.25)$$



**Bild 5:** Dichte der Rechteckverteilung

Die Verteilungsfunktion der Gleichverteilung in  $(a, b)$  lautet

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

Da es sich um eine stetige Verteilung handelt, spielt es keine Rolle, ob das offene oder das abgeschlossene Intervall zwischen  $a$  und  $b$  betrachtet wird. Gemäß obiger Definition sind dann die ZV  $X$  mit  $U(a, b)$  und  $Y$  mit  $U[a, b]$  ( $a$  und  $b$  eingeschlossen) identisch verteilt.

2. **Normal (Gauß)-Verteilung**  $N(\mu, \sigma^2)$

Die Normalverteilung spielt eine zentrale Rolle. Für  $\mu \in \mathbb{R}$ ,  $\sigma^2 > 0$  lautet die Dichte

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1.26)$$

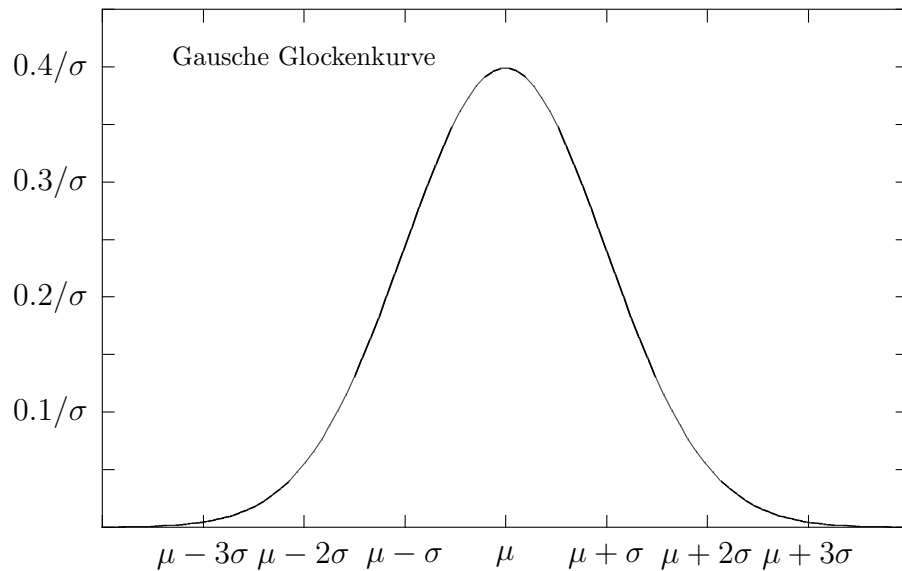
$N(0, 1)$  heißt **Standardnormalverteilung** und hat die spezielle Dichte

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

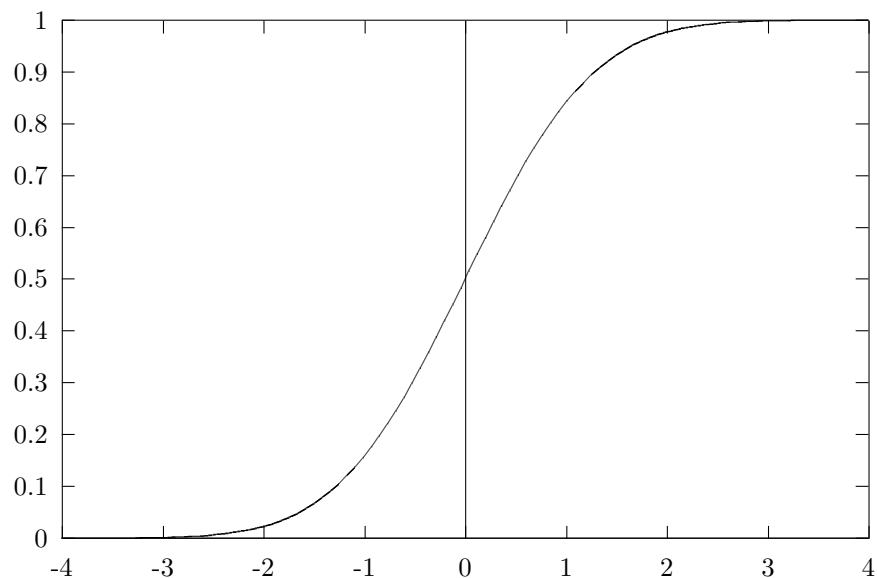
Die Verteilungsfunktion

$$\Phi(u) = \int_{-\infty}^u \varphi(t) dt$$

der Standardnormalverteilung  $N(0, 1)$  ist analytisch. Eine geschlossene Darstellung mit bekannten Standardfunktionen ist jedoch für  $\Phi$  nicht möglich. Numerisch kann  $\Phi$  genauso berechnet werden wie etwa die Exponentialfunktion oder andere wichtige Funktionen. Zum Ablesen von Quantilen (vgl. Abschnitt 1.5.4) verwendet man Tabellen von  $\Phi$ , die in fast allen Statistikbüchern zu finden sind, so auch im Anhang dieser Materialien.



**Bild 6:** Dichte der  $N(\mu, \sigma^2)$ -Verteilung



**Bild 7:** Verteilungsfunktion  $\Phi$  der Standardnormalverteilung  $N(0, 1)$

### 3. Exponentialverteilung $ED(\lambda)$

Die Bezeichnung steht für "exponential distribution". Bei Lebensdauervertei-

gen nehmen die entsprechenden ZV nur nichtnegative Werte an. Ein einfaches, aber wichtiges Beispiel ist die Exponentialverteilung mit Parameter  $\lambda > 0$ . Die Dichte ergibt sich zu

$$f_T(t) = \lambda e^{-\lambda t} 1_{(0,\infty)}(t) = \begin{cases} \lambda e^{-\lambda t} & , \text{ falls } t > 0 \\ 0 & , \text{ sonst.} \end{cases} \quad (1.27)$$

Ist etwa die Ausfallzeit  $T$  eines Bauteils exponentialverteilt mit Parameter  $\lambda = 2$ , so bedeutet

$$P(1 < T \leq 2) = \int_1^2 \lambda e^{-\lambda t} dt = -e^{-2t} \Big|_1^2 = -e^{-4} + e^{-2} \simeq 0.135 - 0.018 = 0.117,$$

dass das Bauteil mit einer Wahrscheinlichkeit von 11.7% im zweiten Betriebsjahr ausfällt. Die Verteilungsfunktion der Exponentialverteilung lautet

$$F_T(t) = (1 - e^{-\lambda t}) 1_{(0,\infty)}(t) = \begin{cases} 0 & , \text{ falls } t \leq 0 \\ 1 - e^{-\lambda t} & , \text{ falls } t > 0. \end{cases} \quad (1.28)$$

Die *tail probability*  $\bar{F}(t) = e^{-\lambda t}$  gibt die Wkt an, dass das Bauteil nicht vor dem Zeitpunkt  $t$  ausfällt (Überlebenswahrscheinlichkeit).

Eine besondere Eigenschaft der Exponentialverteilung ist die sog. **Gedächtnislosigkeit**. Hat nämlich ein Bauteil das Alter  $t_0$  erreicht, so gilt mit (1.3)

$$\begin{aligned} P(T > t + t_0 | T > t_0) &= \frac{P(T > t + t_0, T > t_0)}{P(T > t_0)} = \frac{P(T > t + t_0)}{P(T > t_0)} \\ &= \frac{e^{-\lambda(t+t_0)}}{e^{-\lambda t_0}} = e^{-\lambda t} = P(T > t). \end{aligned}$$

Damit ist die Wahrscheinlichkeit für das Bauteil, im Intervall  $(0, t)$  auszufallen genauso groß, wie diejenige, im Intervall  $(t_0, t_0 + t)$  auszufallen, nachdem das Bauteil bereits das Alter  $t_0$  erreicht hat.

#### 4. Gammaverteilung $\Gamma(q, \lambda)$

Für  $z > 0$  bezeichne

$$\Gamma(z) := \int_0^{\infty} t^{z-1} e^{-t} dt \quad (1.29)$$

die **Gammfunktion**. Es gelten die bekannten Identitäten

$$\Gamma(z + 1) = z \Gamma(z); \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}; \quad \Gamma(n) = (n - 1)!, \quad n \in \mathbb{N}.$$

Für die Parameter  $q > 0$  und  $\lambda > 0$  ist

$$f_Z(z) = \frac{\lambda^q}{\Gamma(q)} z^{q-1} e^{-\lambda z} 1_{(0,\infty)}(z) \quad (1.30)$$

die Dichte einer Gamma-verteilten ZV. Spezialfälle dieser Verteilungsfamilie sind u.a. die Exponentialverteilung  $ED(\lambda) = \Gamma(1, \lambda)$ . Später in Bsp. 3.14 werden

wir sehen, dass Summen von  $n$  sog. "unabhängigen"  $ED(\lambda)$ -verteilten ZV eine Erlangverteilung besitzen. Wegen  $Erlang(n, \lambda) = \Gamma(n, \lambda)$  handelt es sich auch hier um ein Element der Familie von Gammaverteilungen. Eine weitere spezielle Gammaverteilung ist die  $\chi^2$ -Verteilung, die bei Summen von Normalverteilungsquadraten eine wichtige Rolle spielt; vgl. Kap. 6.

**Def. 1.46 (Bezeichnung)**

Anstelle von "  $X$  ist normalverteilt mit den Parametern  $\mu$  und  $\sigma^2$  " schreiben wir kurz

$$X \sim N(\mu, \sigma^2)$$

oder entsprechend bei einer  $(n, p)$ -binomialverteilten Größe  $Y$

$$Y \sim B(n, p)$$

usw.

### 1.5.4 Prozentpunkte und Quantilfunktion

Zur Simulation einer ZV mit beliebiger Verteilungsfunktion  $F$  und zur Definition von sog. **Quantilen** benötigt man eine verallgemeinerte Inversion  $F^{\leftarrow}$  von  $F$ .

**Def. 1.47 (Verallgemeinerte Inverse, Quantilfunktion)**

$X$  sei eine ZV mit Verteilungsfunktion  $F$ .

$$F^{\leftarrow}(p) := \inf \{ x \in \mathbb{R} \mid F(x) \geq p \}, \quad p \in (0, 1).$$

heißt **verallgemeinerte Inverse** oder **Quantilfunktion** zu  $F$ .

**Bemerkung 1.48**

Für streng monotonen  $F : \mathbb{R} \rightarrow (0, 1)$  gilt  $F^{\leftarrow} = F^{-1}$ .

**Lemma 1.49 (Eigenschaften der Quantilfunktion)**

Für beliebige Verteilungsfunktionen gilt

$$F^{\leftarrow}(p) \leq t \iff p \leq F(t) \quad \text{für alle } t \in \mathbb{R} \text{ und } p \in (0, 1).$$

**Def. 1.50 (p-Quantil, Median, Quartil, IQR)**

$X$  sei eine ZV mit Verteilungsfunktion  $F$ , dann heißt für  $0 < p < 1$  die Zahl

$$x_p := F^{\leftarrow}(p) = \inf \{ x \in \mathbb{R} \mid F(x) \geq p \}$$

**p-Quantil** oder **p-Prozentpunkt** von  $X$ .

Das wichtigste Quantil  $x_{0.5} = Q_2$  heißt **Median** oder 2. Quartil. Erstes bzw. drittes **Quartil** lauten  $Q_1 = x_{0.25}$  bzw.  $Q_3 = x_{0.75}$ .

Den Abstand  $IQR := x_{0.75} - x_{0.25}$  nennt man **Interquartilsabstand** (inter quartile range).

Gelegentlich findet man leicht abweichende Definitionen von  $x_p$ , die aber für streng monotonen, stetigen  $F$  stets übereinstimmen. In diesem Fall gilt dann

$$P(X \leq x_p) = p, \quad \text{für } 0 < p < 1.$$

**Bemerkung 1.51 (Lageparameter)**

1. p-Quantile und insbesondere der Median sind **Lageparameter** der Verteilung von  $X$ .
2. Weitere Lageparameter sind der Mittelwert (Erwartungswert)  $\mu$ , den wir in den folgenden Abschnitten näher behandeln und der **Modus** oder **Modalwert**  $x_M$ . Bei sog. **unimodalen** (eingipfeligen) Verteilungen bezeichnet  $x_M$  das eindeutig bestimmte Maximum von  $f_X$ .

**Beispiel 1.52 (Wichtige Quantile der Normalverteilung)**

Am häufigsten benötigt man die Quantile der Standardnormalverteilung  $\Phi$ . Da deren Dichte eine gerade Funktion ist, folgt

$$\Phi(-x) = 1 - \Phi(x) \quad \text{für alle } x \in \mathbb{R}. \quad (1.31)$$

Damit kann die in den meisten Statistikbüchern zu findende Tabellierung auf  $x \geq 0$  beschränkt werden. Es folgen oft benützte Quantile der Normalverteilung.

$p$	Quantil $x_p$
0.5	0
0.9	1.282
0.95	1.645
0.975	1.960
0.99	2.326
0.995	2.576

**Beispiel 1.53 (Median, Modus)**1. Normalverteilung  $N(\mu, \sigma^2)$ 

$$\text{Median } x_{0.5} = \mu, \quad \text{Modus } x_M = \mu.$$

2. Exponentialverteilung  $ED(\lambda)$ 

$$\text{Median } x_{0.5} = \frac{\ln 2}{\lambda}, \quad \text{Modus } x_M = 0.$$

### 1.5.5 Eindimensionale Transformationen von ZV

$X$  sei eine ZV mit Verteilungsfunktion  $F_X$  und Wahrscheinlichkeitsfunktion bzw. Dichte  $f_X$ .

**Def. 1.54 ( $\mathfrak{B} - \mathfrak{B}$ -messbar)**

Eine Funktion  $g : \mathbb{R} \rightarrow \mathbb{R}$  heißt  $\mathfrak{B} - \mathfrak{B}$ -messbar oder einfach **messbar**, falls

$$g^{-1}(B) \in \mathfrak{B} \text{ für alle } B \in \mathfrak{B}.$$

**Satz 1.55 (Messbare Funktionen)**

Messbar sind u. a.

1. stetige Funktionen; z. B.  $\max\{0, f\}$ ,  $\min\{0, f\}$  bei stetigem  $f$
2. stückweise stetige Funktionen; z. B. die Indikatorfunktion  $1_A$
3. Verteilungsfunktionen und die Quantilfunktion aus Def. 1.47.

Zufallsgeneratoren am Rechner erzeugen in fast allen Fällen (Pseudo-) Zufallszahlen, die in  $(0, 1)$  gleichverteilt sind. Zur Simulation einer ZV mit beliebiger Verteilungsfunktion  $F$  benötigt man die Quantilfunktion  $F^{\leftarrow}$  von  $F$  aus Def. 1.47.

**Korollar 1.56 (Quantiltransformation)**

$X$  sei eine ZV mit Verteilungsfunktion  $F$  und Quantilfunktion  $F^{\leftarrow}$ . Weiterhin sei  $U$  im Intervall  $(0, 1)$  gleichverteilt.

1. Für beliebiges  $X$  hat die Zufallsvariable  $Y := F^{\leftarrow}(U)$  die Verteilungsfunktion  $F$ , d. h.

$$P(Y \leq y) = P(F^{\leftarrow}(U) \leq y) = F(y), \quad y \in \mathbb{R}.$$

2. Für stetiges  $F$  ist  $Z := F(X)$  gleichverteilt in  $(0, 1)$ , d. h.

$$P(Z \leq z) = P(F(X) \leq z) = z, \quad z \in (0, 1).$$

Es folgt die Diskussion wichtiger Spezialfälle.

**Satz 1.57 (Transformationsatz)**

$X$  sei eine ZV und  $g : \mathbb{R} \rightarrow \mathbb{R}$  messbar, dann gilt:

1.  $Y := g(X)$  ist eine ZV.
2. Falls  $g$  stetig und streng monoton wachsend, so lautet die Verteilungsfunktion  $F_Y$  von  $Y := g(X)$

$$F_Y(y) = F_X[g^{-1}(y)] \quad \text{für } y \in g(\mathbb{R}).$$

**Beispiel 1.58**

1.  $N(0, 1)$

Gemäß Bemerkung 1.48 und Korollar 1.56 Nr. 1 ist bei  $(0, 1)$ -gleichverteiltem  $U$  die Zufallsvariable  $X := \Phi^{-1}(U)$  standardnormalverteilt. Zur numerischen Berechnung der Standardnormalverteilungsfunktion  $\Phi$  bzw. von  $\Phi^{-1}$  gibt es ausgezeichnete rationale Approximationen (siehe etwa [Kredler & Ritter (1995)]; Anhang).

2. Exponentialverteilung  $ED(\lambda)$

$$F(x) = (1 - e^{-\lambda x}) 1_{(0, \infty)}(x) \quad \implies \quad F^{\leftarrow}(y) = -\frac{\ln(1-y)}{\lambda} 1_{(0,1)}(y).$$

3. Bernoulli-Verteilung  $B(1, p)$

$$F(x) = \begin{cases} 0 & , \text{ falls } x < 0 \\ 1-p & , \text{ falls } 0 \leq x < 1 \\ 1 & , \text{ falls } x \geq 1, \end{cases}$$

und somit

$$F^{\leftarrow}(y) = \begin{cases} 0 & , \text{ falls } 0 < y \leq 1-p \\ 1 & , \text{ falls } 1-p < y < 1. \end{cases}$$

4. Die ZV  $X$  habe die Verteilungsfunktion  $F_X(x)$ . Seien  $a \in \mathbb{R}$  und  $b \neq 0$ . Dann lautet die Verteilungsfunktion  $F_Y$  von  $Y := a + bX$

$$F_Y(y) = \begin{cases} F_X\left(\frac{y-a}{b}\right) & , \text{ falls } b > 0 \\ 1 - F_X\left(\frac{y-a}{b}\right) & , \text{ falls } b < 0, \end{cases}$$

und falls  $X$  stetig ist mit Dichte  $f_X$ , so ist  $Y$  auch stetig mit Dichte

$$f_Y(y) = \frac{1}{|b|} f_X\left(\frac{y-a}{b}\right). \quad (1.32)$$



5. Eine parametrisierte Familie von Verteilungen heißt **stabil** bzgl. affinen Transformationen, wenn mit  $F(x)$  auch  $F([x - a]/b)$  für alle  $a$  und  $b \neq 0$  zu dieser Familie gehören. Wichtige Beispiele sind die Familien  $N(\mu, \sigma^2)$  der Normalverteilungen und der Gleichverteilungen  $U(c, d)$ ,  $c < d$ .

$X$  sei  $N(0, 1)$ -verteilt mit Dichte  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$  und  $Y := \mu + \sigma X$  mit  $\sigma > 0$ . Dann ist  $Y \sim N(\mu, \sigma^2)$  mit Dichte

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

6. Für nichtnegative  $X$  und  $\beta > 0$  betrachten wir  $Y := X^\beta$ . Man erhält die Verteilungsfunktion

$$F_Y(y) = F_X(y^{1/\beta})$$

und gegebenenfalls die Dichte

$$f_Y(y) = \frac{1}{\beta} y^{\frac{1-\beta}{\beta}} f_X(y^{1/\beta}). \quad (1.33)$$

7. Besonders wichtig ist der Fall  $Y := X^2$ , wobei  $X$  jetzt wieder beliebig sein kann. Es gilt für  $y \geq 0$

$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y} - 0).$$

Ist zusätzlich  $X$  stetig und symmetrisch zum Nullpunkt, d. h.  $F_X(-x) = 1 - F_X(x)$ , dann gilt

$$f_Y(y) = \frac{1}{\sqrt{y}} f_X(\sqrt{y}), \quad y > 0.$$

Eine Anwendung dazu ist  $X \sim N(0, \sigma^2)$  mit  $f_X(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-x^2/(2\sigma^2)}$  und

$$f_{X^2}(y) = \frac{1}{\sqrt{2\pi} \sigma} y^{-\frac{1}{2}} e^{-y/(2\sigma^2)}.$$

Damit ist  $X^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2\sigma^2})$ -verteilt. Man spricht auch von der  $\chi^2$ -Verteilung.

8. **Lognormalverteilung**  $\text{Log}N(\mu, \sigma^2)$

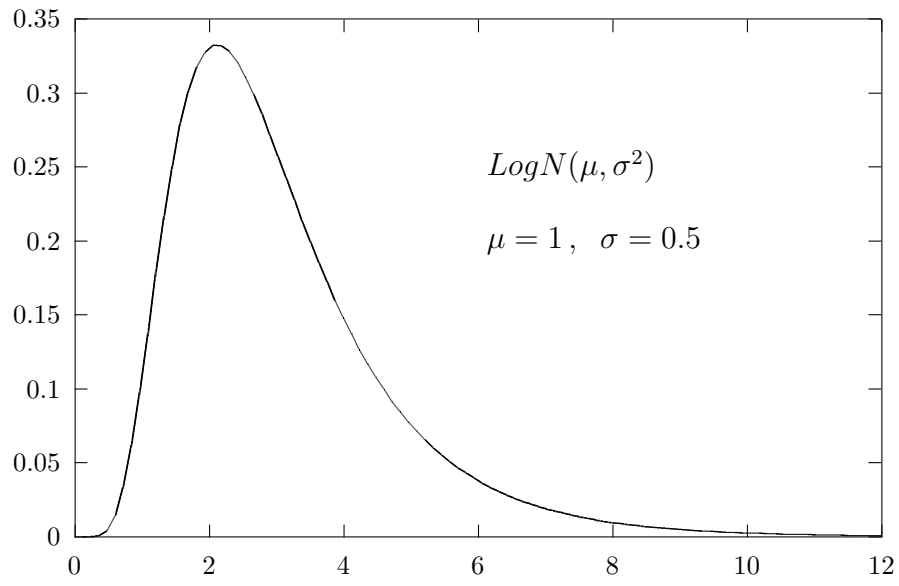
Eine positive ZV  $X$  heißt  $\text{Log}N(\mu, \sigma^2)$ -verteilt oder **logarithmisch normalverteilt**, wenn  $Y := \ln(X) \sim N(\mu, \sigma^2)$ -verteilt ist. Damit lautet die Dichte von  $X$

$$f_X(x) = \frac{1}{\sqrt{2\pi} \sigma x} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) 1_{(0, \infty)}(x). \quad (1.34)$$

Die Lognormalverteilung ist eine einseitige unsymmetrische Verteilung mit

$$\text{Median } x_{0.5} = e^\mu, \quad \text{Modus } x_M = e^{\mu - \sigma^2}$$

und wird u.a. zur Modellierung von Lebensdauern herangezogen.



**Bild 8:** Dichte der Lognormalverteilung  $LogN(\mu, \sigma^2)$

## 1.6 Erwartungswerte

Der abstrakte Mittelwert  $E(X)$  und die Varianz  $Var(X)$  sind so eingerichtet, dass sie für endliches  $\Omega$  mit den empirischen Mittelwert- bzw. Streuungsmaßen  $\bar{x}$  bzw.  $\bar{s}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$  zusammenfallen. Im allgemeinen kann  $\Omega$  überabzählbar viele Elemente enthalten. Für stetige Zufallsvariable (ZV) muss dies sogar so sein. Wir beschränken uns hier auf diskrete und stetige ZV.

### 1.6.1 Erwartungswert und Varianz

**Def. 1.59 (Erwartungswert (EW):**  $E(X) = E X = \mu$ )

Den **Erwartungswert** (auch Mittelwert) einer **diskreten** ZV  $X$  definiert man als

$$\mu = E(X) = \sum_i x_i f_X(x_i) = \sum_i x_i p_i,$$

falls  $\sum_i |x_i| f_X(x_i) < \infty$  existiert.

Analog legt man den **Erwartungswert** einer **stetigen** ZV  $X$  fest als

$$\mu = E(X) = \int_{-\infty}^{\infty} x f_X(x) dx,$$

falls wiederum die Existenz von  $\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$  gesichert ist.

#### Bemerkung 1.60

1. Man nennt  $\mu = E(X)$  wie den Median oder Modus einen **Lageparameter**.
2. Gelegentlich schreibt man kurz  $E X$  statt  $E(X)$ .
3. Für beliebige nichtnegative ZV, d. h.  $P(X < 0) = 0$ , gilt

$$E(X) = \int_0^{\infty} \bar{F}(x) dx = \int_0^{\infty} [1 - F(x)] dx, \quad (1.35)$$

falls der Erwartungswert existiert (Beweis siehe weiterführende Vorlesung). Zum Beispiel erhält man für die Exponentialverteilung

$$\mu = E(X) = \int_0^{\infty} \bar{F}(x) dx = \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda} \left( -e^{-\lambda x} \right)_0^{\infty} = \frac{1}{\lambda}.$$

4. Läßt man auch den Wert  $\infty$  für den Erwartungswert zu, so hat jede nichtnegative ZV einen Erwartungswert. Dies ist insbesondere dann sinnvoll, wenn z. B. Wartezeiten mit  $P(X = \infty) > 0$  betrachtet werden.
5. Für beliebige ZV geht man dann über zur Zerlegung

$$X = X^+ - X^- := \max\{0, X\} - \max\{0, -X\}.$$

Falls  $E(X^+) < \infty$  oder  $E(X^-) < \infty$ , dann definiert man

$$E(X) := E(X^+) - E(X^-). \quad (1.36)$$

Man hat auf jeden Fall in (1.36) den Term  $\infty - \infty$  zu vermeiden; vgl. (6.42).

**Def. 1.61 (Erwartungswert bzgl. einer Funktion  $g(X)$ )**

Für (messbare) Funktionen  $g$  definiert man analog

$$E[g(X)] = \sum_i g(x_i) f_X(x_i) = \sum_i g(x_i) p_i \quad \text{für diskretes } X \quad \text{und}$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \quad \text{für stetiges } X,$$

$$\text{falls } \sum_i |g(x_i)| f_X(x_i) \quad \text{bzw.} \quad \int_{-\infty}^{\infty} |g(x)| f_X(x) dx \quad \text{existieren.}$$

Aus diesen Definitionen ergeben sich direkt eine Reihe von Eigenschaften:

**Satz 1.62 (Rechenregeln für Erwartungswerte)**

1. Falls  $X \geq 0$ , so gilt  $E(X) \geq 0$ .
2. Für Indikatorfunktionen  $X = 1_A$  bzgl.  $A \in \mathfrak{F}$  gilt  $E(1_A) = P(A)$ .
3. Linearität:  $X, X_1, \dots, X_n$  seien ZV bzgl. desselben Wahrscheinlichkeitsraums  $(\Omega, \mathfrak{F}, P)$  mit existierenden Erwartungswerten und  $a, b, a_i \in \mathbb{R}$ ,  $i = 1, \dots, n$ , dann gilt

$$\begin{aligned} E(a) &= a \\ E(a + bX) &= a + bE(X) \\ E(a_1X_1 + \dots + a_nX_n) &= a_1E(X_1) + \dots + a_nE(X_n). \end{aligned}$$

**Def. 1.63 (Varianz  $\sigma^2$ , Standardabweichung  $\sigma$ , Streuung)**

Die **Varianz** einer ZV  $X$  wird festgelegt als mittlere quadratische Abweichung

$$\sigma^2 = \text{Var}(X) = D^2(X) := E[(X - \mu)^2], \text{ wobei } \mu = E(X).$$

$$\sigma = D(X) = \sqrt{\text{Var}(X)}$$

heißt **Standardabweichung** und hat stets dieselbe Benennung wie  $X$  selbst.

Konkret erhält man für **diskretes**  $X$

$$\text{Var}(X) = \sum_i (x_i - \mu)^2 f_X(x_i) = \sum_i (x_i - \mu)^2 p_i$$

und im **stetigen** Fall

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

Varianz und Standardabweichung sind **Streuungsparameter**.

Zur Berechnung der Varianz verwendet man häufig die **Verschiebungsregel**

$$D^2(X) = E(X^2) - \mu^2, \quad (1.37)$$

wobei  $E(X^2) = \sum_i x_i^2 p_i$  bzw.  $E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx$ .

**Satz 1.64 (Rechenregeln für Varianzen)**

$X, X_1, \dots, X_n$  seien ZV bzgl. desselben Wahrscheinlichkeitsraums  $(\Omega, \mathfrak{F}, P)$  mit existierenden Varianzen und  $a, b \in \mathbb{R}$ .

$$\begin{aligned} D^2(X) &\geq 0 \\ D^2(a) &= 0 \\ D^2(X + a) &= D^2(X) \\ D^2(bX) &= b^2 D^2(X) \\ D^2(a + bX) &= b^2 D^2(X). \end{aligned} \quad (1.38)$$

In gewissem Sinn gilt auch die "Umkehrung" von (1.38):

$$\text{Var}(x) = D^2(X) = 0 \iff P(X = a) = 1, \text{ mit } a = E(X). \quad (1.39)$$

**Folgerung 1.65 (Standardisierung)**

Gegeben sei eine ZV  $X$ , für die  $E(X) = \mu$  und  $D^2(X) = \sigma^2 > 0$  existieren. Dann gilt für die **standardisierte ZV**

$$U := \frac{X - \mu}{\sigma} \quad \Rightarrow \quad E(U) = 0, \quad D^2(U) = 1. \quad (1.40)$$

**Beispiel 1.66 (Diskrete ZV)**

1. **Bernoulli-Variable**  $B(1, p)$ ,  $0 < p < 1$

$$P(X = 1) = p, \quad P(X = 0) = 1 - p,$$

$$E(X) = p, \quad D^2(X) = p(1 - p).$$

2. **Binomialverteilung**  $B(n, p)$ ,  $0 < p < 1$

$$p_i = \binom{n}{i} p^i (1 - p)^{n-i}, \quad i = 0, 1, \dots, n$$

$$E(X) = np, \quad D^2(X) = np(1 - p).$$

3. **Poisson-Verteilung**  $Poi(\lambda)$ ,  $\lambda > 0$

$$p_i = e^{-\lambda} \lambda^i / i!, \quad i = 0, 1, 2, \dots$$

$$E(X) = \lambda, \quad D^2(X) = \lambda.$$

4. **Geometrische Verteilung**

$$p_i = p(1 - p)^i, \quad 0 < p < 1, \quad i = 0, 1, \dots, n$$

$$E(X) = \frac{1 - p}{p}, \quad D^2(X) = \frac{1 - p}{p^2}.$$

**Beispiel 1.67 (Stetige ZV)**

1. **Normalverteilung**  $N(\mu, \sigma^2)$

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

$$E(X) = \mu, \quad D^2(X) = \sigma^2$$

2. **Gleich (Rechteck)-verteilung**  $U(a, b)$ 

$$f(x) = \frac{1}{(b-a)} 1_{(a,b)}(x); \quad a < b,$$

$$E(X) = (a+b)/2, \quad D^2(X) = (b-a)^2/12$$

3. **Exponentialverteilung**  $ED(\lambda)$ 

$$f(x) = \lambda e^{-\lambda x} 1_{(0,\infty)}(x); \quad \lambda > 0,$$

$$E(X) = \frac{1}{\lambda}, \quad D^2(X) = \frac{1}{\lambda^2}.$$

Die mittlere Lebensdauer ist also  $1/\lambda$ .

4. **Gammaverteilung**  $\Gamma(q, \lambda)$ 

$$f_Z(z) = \frac{\lambda^q}{\Gamma(q)} z^{q-1} e^{-\lambda z} 1_{(0,\infty)}(x); \quad q, \lambda > 0,$$

$$E(Z) = \frac{q}{\lambda}, \quad D^2(Z) = \frac{q}{\lambda^2} \quad (1.41)$$

5. **Cauchy-Verteilung**

Diese hat für  $a > 0$  die Dichte

$$f_X(x) = \frac{a}{\pi(a^2 + x^2)}. \quad (1.42)$$

$E(X)$  und  $D^2(X)$  existieren nicht.

6. **Pareto-Verteilung**

Diese spielt eine Rolle in der Extremwerttheorie und wurde in der Vergangenheit auch zur Beschreibung ökonomisch-soziologischer Größen wie Verteilung von Einkommen und Vermögen etc. herangezogen. Sei  $c > 0$  ein gewisser Schwellenwert, den die entsprechende ZV  $X$  nicht erreichen oder unterschreiten kann. Für  $\alpha > 0$  lauten dann Dichte bzw. Verteilungsfunktion einer Pareto-verteilten ZV  $X$

$$f(x) = \frac{\alpha}{c} \left(\frac{c}{x}\right)^{\alpha+1} 1_{(c,\infty)}(x), \quad F(x) = 1 - \left(\frac{c}{x}\right)^\alpha 1_{(c,\infty)}(x). \quad (1.43)$$

Falls  $\alpha > 1$ , so existiert  $E(X) = \frac{\alpha c}{\alpha - 1}$ .

Für  $\alpha > 2$  existiert auch die Varianz  $Var(X) = \frac{\alpha c^2}{(\alpha - 1)^2(\alpha - 2)}$ .

7. **Lognormalverteilung**  $LogN(\mu, \sigma^2)$ 

Aus der zugehörigen Dichte (1.34) leitet man ab, dass lognormalverteilte ZV  $X$  positiv sind mit

$$E(X) = e^{\mu+\sigma^2/2} \quad \text{und} \quad Var(X) = e^{2\mu+\sigma^2} (e^{\sigma^2} - 1). \quad (1.44)$$

Die speziellen Eigenschaften von Verteilungsfunktionen (Monotonie,  $0 \leq F(x) \leq 1$ ) erlauben eine einheitliche Schreibweise von Erwartungswerten für diskrete ZV bzw. solche mit (stückweise) stetigen Dichten.

**Def. 1.68 (Riemann-Stieltjes-Integral)**

*X sei eine diskrete ZV oder besitze eine Dichte, die bis auf endlich viele Stellen stetig ist; für die Funktion g existiere  $E[g(X)]$ . Dann definiert man*

$$\int_{-\infty}^{\infty} g(x) dF_X(x) := \begin{cases} \sum g(x_i) f_X(x_i) & , \text{ falls } X \text{ diskret} \\ \int_{-\infty}^{\infty} g(x) f_X(x) dx & , \text{ falls } X \text{ stetig.} \end{cases} \quad (1.45)$$

**Bemerkung 1.69**

1. Kurz schreibt man wieder  $f = f_X$ ,  $F = F_X$  und  $\int_{-\infty}^{\infty} g(x) dF(x)$ .
2. Unter den Voraussetzungen von Def. 1.68 gilt also:

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} dF(x) \\ E(X) &= \int_{-\infty}^{\infty} x dF(x) := \mu \\ \text{Var}(X) &= \int_{-\infty}^{\infty} (x - \mu)^2 dF(x). \end{aligned}$$

3. Für (stückweise) differenzierbares  $F$  mit  $F' = \frac{dF}{dx} = f$  leuchtet die Schreibweise  $dF(x) = f(x) dx$  formal sofort ein, kann aber analog zum nachfolgenden diskreten Fall über den Mittelwertsatz der Differentialrechnung hergeleitet werden.
4. Bei diskretem  $X$  nehmen wir an, dass die  $x_i$ ,  $i \in \mathbb{N}$ , in einem kompakten Intervall  $[a, b]$  liegen mit  $a < x_1 < x_2 < \dots \leq b$ . Für stets feiner werdende Unterteilungen

$$a = a_0^{(n)} < a_1^{(n)} < a_2^{(n)} < \dots < a_{n-1}^{(n)} < a_n^{(n)} = b$$

des Integrationsintervalls  $[a, b]$  mit

$$\lim_{n \rightarrow \infty} \max_{0 \leq j \leq n-1} \{ |a_{j+1}^{(n)} - a_j^{(n)}| \} = 0$$

existiert dann der Grenzwert

$$\int_{-\infty}^{\infty} g(x) dF(x) := \lim_{n \rightarrow \infty} \sum_{j=1}^n g(a_j^{(n)}) [F(a_j^{(n)}) - F(a_{j-1}^{(n)})] = \sum_{i \in \mathbb{N}} g(x_i) p_i,$$



da  $F$  stückweise konstant ist und nur an den  $x_i$  Sprünge der Höhe

$$p_i = F(x_i) - F(x_i - 0)$$

aufweist. Wegen der Eigenschaften der Verteilungsfunktion  $F$  ist eine Ausdehnung des Intervalls  $[a, b]$  auf  $(-\infty, \infty)$  nicht problematisch.

5. Die Schreibweise ...  $dF(x)$  wird für beliebige ZV verwendet. Allerdings ist dann das allgemeinere Lebesgue-Integral zugrunde zu legen. Auf Einzelheiten können wir hier nicht eingehen. Details findet man in den einschlägigen Lehrbüchern.

## 1.6.2 Momente und Kumulanten

Bei spezieller Wahl der Funktion  $g$  erhält man weitere Momente der Verteilung der ZV  $X$ , sofern diese existieren.

### Def. 1.70 (Momente)

Vorausgesetzt die entsprechenden Erwartungswerte existieren, definiert man mit  $\mu_1 = \mu := E(X)$  für (die Ordnung)  $k = 1, 2, \dots$

$$\mu_k = E[X^k] = \int_{-\infty}^{\infty} x^k dF(x) \quad k\text{-tes (gewöhnliches) Moment}$$

$$E[|X|^k] = \int_{-\infty}^{\infty} |x|^k dF(x) \quad k\text{-tes absolutes Moment}$$

$$E[(X - \mu)^k] = \int_{-\infty}^{\infty} (x - \mu)^k dF(x) \quad k\text{-tes zentrales Moment} \quad (1.46)$$

$$E[X(X-1) \cdot \dots \cdot (X-k+1)] \quad k\text{-tes faktorielles Moment} \quad (1.47)$$

**Satz 1.71 (Existenz von Momenten)**

Man zeigt leicht, vgl. etwa [Heinhold & Gaede (1979)], S. 105:

$$\left[ E(|X|^k) \right]^{k+1} \leq \left[ E(|X|^{k+1}) \right]^k, \quad (1.48)$$

also z. B. für  $k = 1$

$$\mu^2 \leq \left[ E(|X|^1) \right]^2 \leq \left[ E(X^2) \right]^1 = E(X^2).$$

Damit sichert die Existenz eines (absoluten) Moments der Ordnung  $k$  die der niedrigeren Momente zur Ordnung  $1, 2, \dots, k - 1$ .

Insbesondere sichert die Existenz eines zweiten Moments von  $X$ , dass  $E(X)$  und  $\text{Var}(X)$  wohldefiniert sind.

Aus den oben definierten Momenten leitet man weitere Kenngrößen her, deren empirische Varianten neben Mittelwert und Varianz eine wichtige Rolle in der explorativen Datenanalyse spielen.

**Def. 1.72 (Variationskoeffizient, Schiefe, Exzess)**

Vorausgesetzt die entsprechenden Momente existieren, so definiert man mit  $\mu := E(X)$ ,  $\sigma^2 := \text{Var}(X) > 0$

$$CV := \frac{\sigma}{\mu}, \quad \mu \neq 0 \quad \text{Variationskoeffizient}$$

*(coefficient of variation)*

$$\gamma_3 := \frac{E[(X - \mu)^3]}{\sigma^3} \quad \text{Schiefe (skewness)}$$

$$\delta_4 := \frac{E[(X - \mu)^4]}{\sigma^4} - 3 \quad \text{Exzess (kurtosis), Wölbung}$$

**Bemerkung 1.73**

1. Schiefe und Exzess verschwinden bei der Normalverteilung.
2. Bei symmetrischen Verteilungen verschwindet die Schiefe, falls diese existiert.
3. Varianz, Schiefe und Exzess sind invariant gegenüber Transformationen der Art

$$X \rightarrow (X + a), \quad a \in \mathbb{R}.$$

Die letzte Eigenschaft wird auch von den sog. **Semiinvarianten** oder **Kumulanten** ab der Ordnung 2 geteilt.

**Def. 1.74 (Kumulanten)**

*Vorausgesetzt die entsprechenden Momente existieren, so lauten die vier ersten Kumulanten*

$$\begin{aligned}\kappa_1 &:= \mu = E[X^1] \\ \kappa_2 &:= \sigma^2 = E[(X - \mu)^2] \\ \kappa_3 &:= E[(X - \mu)^3] \\ \kappa_4 &:= E[(X - \mu)^4] - 3\sigma^4\end{aligned}$$

**Bemerkung 1.75**

1. Der Vorsatz "Semi-" ist gerechtfertigt, da sich die erste Kumulante bei der Transformation  $X \rightarrow (X + a)$  für  $a \neq 0$  ändert.
2. Für die Normalverteilung verschwinden alle Kumulanten der Ordnung 3 und höher.  $N(\mu, \sigma^2)$  ist die einzige Verteilung mit dieser Eigenschaft.
3. Eine einfache Darstellung der Kumulanten (falls diese existieren) liefert die Entwicklung von  $\ln(M(s))$ , wobei  $M$  die momenterzeugende Funktion bezeichnet; vgl. Abschnitt 3.2.
4. Die Zusammenhänge zwischen Varianz, Schiefe, Exzess und den Kumulanten lauten:

$$\begin{aligned}\sigma^2 &= \kappa_2 && \text{(Varianz)} \\ \gamma_3 &= \frac{\kappa_3}{\sqrt{\kappa_2^3}} && \text{(Schiefe)} \\ \delta_4 &= \frac{\kappa_4}{\kappa_2^2} && \text{(Exzess)}\end{aligned}$$

**Beispiel 1.76 (Momente, Kumulanten)**1.  $N(\mu, \sigma^2)$ 

$$\text{Kumulanten: } E(X) = \kappa_1 = \mu; \text{Var}(X) = \kappa_2 = \sigma^2; \kappa_k = 0, \text{ sonst.} \quad (1.49)$$

$$E[(X - \mu)^k] = \begin{cases} 0 & , \text{ falls } k \text{ ungerade} \\ 1 \cdot 3 \cdot 5 \cdot \dots \cdot (k-1) \sigma^k & , \text{ sonst} \end{cases} \quad (1.50)$$

2.  $Poi(\lambda), \lambda > 0$ 

$$\kappa_k = \lambda, \text{ f\"ur alle } k = 1, 2, \dots \quad (1.51)$$

3.  $\Gamma(q, \lambda), q > 0, \lambda > 0$ 

$$E[X^k] = \frac{q(q+1) \cdot \dots \cdot (q+k-1)}{\lambda^k}, \quad k = 1, 2, \dots; \quad (1.52)$$

insbesondere gilt:

$$E(X) = \frac{q}{\lambda} \quad \text{und} \quad \text{Var}(X) = \frac{q}{\lambda^2},$$

und f\"ur den Spezialfall  $q = 1$ , die Exponentialverteilung:

$$E(X) = \frac{1}{\lambda} \quad \text{und} \quad \text{Var}(X) = \frac{1}{\lambda^2},$$

4. Gleichverteilung  $U(a, b), a < b$ Hier gilt  $\mu = E(X) = (a + b)/2$  und

$$E[(X - \mu)^k] = \begin{cases} 0 & , \text{ falls } k \text{ ungerade} \\ \frac{1}{k+1} \left(\frac{b-a}{2}\right)^k & , \text{ sonst.} \end{cases} \quad (1.53)$$

# Kapitel 2

## Mehrdimensionale Verteilungen

Eine natürliche Erweiterung des bisherigen Begriffs der eindimensionalen Zufallsvariablen (ZV) ergibt sich, wenn man mehrere Merkmale, wie z. B. Größe und Gewicht eines Individuums  $\omega \in \Omega$ , gleichzeitig betrachtet.

### 2.1 Gemeinsame Verteilung von $n$ Zufallsvariablen

Die Definition 1.32 läßt sich in kanonischer Weise auf den  $n$ -dimensionalen Fall übertragen. Zur Vermeidung von Namenskollisionen seien in diesem Abschnitt die Vektoren  $\mathbf{X} = (X_1, \dots, X_n)^T \in \mathbb{R}^n$  und  $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{R}^n$  fett bezeichnet;  $T$  bezeichnet Transposition.

#### Def. 2.1 (n-dim. Zufallsvariable (ZV), Zufallsvektor)

Gegeben sei ein Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{F}, P)$ . Eine Abbildung

$$\mathbf{X} : \Omega \rightarrow \mathbb{R}^n, \quad \text{mit}$$

$$\mathbf{X}^{-1}(\mathbf{B}) \in \mathfrak{F} \quad \text{für alle } n\text{-dim. Borelmengen } \mathbf{B} \in \mathfrak{B}^n$$

heißt ( $n$ -dimensionale) **Zufallsvariable** oder auch **Zufallsvektor**. Der Vektor  $\mathbf{x} = \mathbf{X}(\omega) \in \mathbb{R}^n$ , der sich bei Durchführung des Zufallsexperiments durch Auswerten von  $\mathbf{X}$  ergibt, heißt **Realisierung** von  $\mathbf{X}$ .

Das Wahrscheinlichkeitsmaß  $P_{\mathbf{X}}$  auf  $(\mathbb{R}^n, \mathfrak{B}^n)$  definiert durch

$$P_{\mathbf{X}}(\mathbf{B}) := P(\mathbf{X}^{-1}(\mathbf{B})) \quad \text{für alle Borelmengen } \mathbf{B} \in \mathfrak{B}^n$$

heißt **Verteilung** (*distribution*) von  $\mathbf{X} = (X_1, \dots, X_n)^T$ .

Analog zum eindimensionalen Fall wird die  $\sigma$ -Algebra  $\mathfrak{B}^n$  wieder von den offenen Mengen des  $\mathbb{R}^n$  erzeugt. Ein anderes Erzeugendensystem für  $\mathfrak{B}^n$  sind die links unbeschränkten  $n$ -dim. Intervalle

$$(-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_n], \quad x_i \in \mathbb{R}, \quad i = 1, \dots, n.$$

Deshalb kann auch hier die Verteilung durch eine geeignete Funktion festgelegt werden.

**Def. 2.2 (n-dim. Verteilungsfunktion)**

Für Argumente  $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{R}^n$  heißt die Funktion  $F_{X_1, \dots, X_n} : \mathbb{R}^n \rightarrow \mathbb{R}$  mit

$$\begin{aligned} F_{X_1, \dots, X_n}(\mathbf{x}) = F_{X_1, \dots, X_n}(x_1, \dots, x_n) &:= P(\mathbf{X} \leq \mathbf{x}) \\ &:= P(\{\omega \mid X_1(\omega) \leq x_1, \dots, X_n(\omega) \leq x_n\}) \end{aligned}$$

**Verteilungsfunktion** des Zufallsvektors  $\mathbf{X}$ . Falls Verwechslungen ausgeschlossen sind, schreiben wir wieder kurz  $F$  bzw.  $F_X$  statt  $F_{X_1, \dots, X_n}$ .

**Folgerung 2.3**

1.  $F$  ist bzgl. jeder Variablen monoton wachsend und rechtsseitig stetig.

$$2. F(x_1, \dots, -\infty, \dots, x_n) := \lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_i, \dots, x_n) = 0, \quad i = 1, \dots, n$$

$$3. F(\infty, \dots, \infty) := \lim_{x_1 \rightarrow \infty} \dots \lim_{x_n \rightarrow \infty} F(x_1, \dots, x_n) = 1.$$

4. Nach dem Additionssatz gilt für  $a_1 < b_1$ ,  $a_2 < b_2$  und  $n = 2$

$$P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2).$$

Weitere Details zu  $\mathfrak{B}^n$  entnehme man Analysis-Lehrbüchern oder z. B. [Williams (1991)], ch. 1.

Zweidimensionale Vektoren heißen bei uns in der Regel  $(Z, Y)$  und, falls keine Verwechslung mit dem Zufallsvektor  $\mathbf{X}$  zu befürchten ist, auch  $(X, Y)$ . Der Einfachheit halber wird hier meist  $(Z, Y)$  statt  $(Z, Y)^T$  geschrieben.

### 2.1.1 Diskrete n-dim. Zufallsvariable

Wir beschränken uns hier zur Darstellung auf den Fall  $n = 2$ . Sei also  $(Z, Y)$ , wobei  $Z$  nur die Werte  $z_i$ ,  $i \in I$  und  $Y$  nur die Werte  $y_j$ ,  $j \in J$  annehme;  $I$  und  $J$  höchstens abzählbar.

**Def. 2.4 (Diskrete 2-dim. Verteilung)**

$(Z, Y)$  sei eine 2-dim. diskrete ZV, und die Indexmengen  $I, J$  höchstens abzählbar unendlich. Durch die

$$p_{ij} := P(Z = z_i, Y = y_j), \quad i \in I, \quad j \in J$$

mit  $p_{ij} \geq 0, i \in I, j \in J$

und  $\sum_{i,j} p_{ij} = 1$

wird die **Verteilung** von  $(Z, Y)$  gemäß Definition 2.1 eindeutig festgelegt.

$$f_{Z,Y}(z, y) = P(Z = z, Y = y) := \begin{cases} p_{ij} & , \text{ falls } z = z_i; i \in I \\ & \text{ und } y = y_j; j \in J \\ 0 & , \text{ sonst} \end{cases}$$

heißt **Wahrscheinlichkeitsfunktion** von  $(Z, Y)$ .

Vereinbarung:

1. Falls keine Verwechslungen zu befürchten sind, wird wieder  $f$  statt  $f_{Z,Y}$  geschrieben.
2. Die verkürzte Schreibweise  $\sum_{i,j} = \sum_{z,y}$  steht für  $\sum_{i \in I} \sum_{j \in J}$ .
3. Wie schon im eindimensionalen Fall gibt es eine eindeutige Beziehung zwischen der Verteilung von  $(Z, Y)$  und den  $p_{ij} > 0$ . Deswegen genügt es stets, zur Charakterisierung der Verteilung die  $p_{ij}$  bzw. die Wahrscheinlichkeitsfunktion  $f_{Z,Y}$  anzugeben.

**Beispiel 2.5 (diskrete 2-dim. Verteilungen)**

1. **Doppelter Münzwurf**; jeweils mit Kopf (K) oder Wappen (W):  
 $\Omega = \{K, W\} \times \{K, W\}$ .  $Z$  sei die Anzahl von "Kopf" und  $Y$  die Anzahl von "Wappen" beim zweimaligen Wurf der Münze.

$$(Z, Y) : \Omega \rightarrow \{0, 1, 2\} \times \{0, 1, 2\}.$$

Wir erhalten mit  $P(K) = p, 0 < p < 1$

$$p_{02} = (1 - p)^2, \quad p_{11} = 2p(1 - p), \quad p_{20} = p^2$$

und  $p_{ij} = 0$  sonst. Damit gilt

$$\sum_{i,j} p_{ij} = (1 - p)^2 + 2p(1 - p) + p^2 = 1.$$

## 2. Zwei Würfel:

$$\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}, \quad Z : \Omega \rightarrow \{1, 2, \dots, 6\}, \quad Y : \Omega \rightarrow \{1, 2, \dots, 6\},$$

$$P(Z = i, Y = j) = \frac{1}{36}, \quad i, j = 1, \dots, 6.$$

## 2.1.2 Stetige n-dim. Zufallsvariable

**Def. 2.6 (Stetige n-dim. Verteilung)**

Der Zufallsvektor  $\mathbf{X} = (X_1, \dots, X_n)$  heißt **stetig n-dim. verteilt**, falls eine nichtnegative, integrierbare Funktion  $f_{X_1, \dots, X_n} : \mathbb{R}^n \rightarrow \mathbb{R}$  existiert mit

$$F_{X_1, \dots, X_n}(\mathbf{x}) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{X_1, \dots, X_n}(\mathbf{x}) \, d\mathbf{x} = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f_{X_1, \dots, X_n}(x_1, \dots, x_n) \, dx_1 \dots dx_n.$$

An den Stetigkeitsstellen  $\mathbf{x}$  von  $f_{X_1, \dots, X_n}$  gilt:

$$f_{X_1, \dots, X_n}(\mathbf{x}) = \frac{\partial^n F_{X_1, \dots, X_n}(\mathbf{x})}{\partial x_1 \dots \partial x_n}. \quad (2.1)$$

Vereinbarung:

1. Je nach Sachlage verwenden wir die Bezeichnungen  $f_{X_1, \dots, X_n}$  oder  $f$ .
2. In den Beispielen dieses Kurses sind die Dichten zumindest stückweise stetig.
3. Generell gelte für  $\mathbf{x} = (x_1, \dots, x_n)^T$  und integrierbares  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  die Abkürzung

$$\int_{\mathbb{R}^n} h(\mathbf{x}) \, d\mathbf{x} := \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h(x_1, \dots, x_n) \, dx_1 \dots dx_n. \quad (2.2)$$

**Folgerung 2.7 (n-dim. Dichte)**

$f_{X_1, \dots, X_n} : \mathbb{R}^n \rightarrow \mathbb{R}$  sei Dichte eines  $n$ -dimensionalen Zufallsvektors  $(X_1, \dots, X_n)$ , dann gilt mit  $\mathbf{x} \in \mathbb{R}^n$ :

1.  $f_{X_1, \dots, X_n} \geq 0$
2.  $\int_{\mathbb{R}^n} f_{X_1, \dots, X_n}(\mathbf{x}) \, d\mathbf{x} = 1.$

Auch hier gilt das gleiche wie für eindimensionale ZV: Zur eindeutigen Festlegung der Verteilung von  $(Z, Y)$  genügt die Angabe einer Dichte  $f_{Z, Y}$ .



**Beispiel 2.8** Wie muss man in der folgenden Formel die Konstante  $c$  wählen, damit  $f$  eine zweidimensionale Dichte ist?

$$f(z, y) = c e^{-z-2y} 1_{(0, \infty) \times (0, \infty)}(z, y)$$

**Def. 2.9 (n-dimensionale Normalverteilung  $N(\boldsymbol{\mu}, \mathbf{C})$ )**

Seien  $\boldsymbol{\mu} \in \mathbb{R}^n$  und  $\mathbf{C} = (c_{ij}) \in \mathbb{R}^{n,n}$  eine positiv definite Matrix (damit ist  $\mathbf{C}$  auch symmetrisch).  $\mathbf{X} = (X_1, \dots, X_n)^T$  heißt **n-dimensional normalverteilt**, wenn es eine Dichte der Form

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C})}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right] \quad (2.3)$$

besitzt. Dabei bezeichnet  $\det(\mathbf{C})$  die Determinante von  $\mathbf{C}$ .

Der Nachweis der Dichteeseigenschaft von  $f$  erfordert die n-dim. Transformationsregel für Dichten und wird später in Bsp. 2.33 geführt. In dieser Einführungsvorlesung werden wir uns fast ausschließlich auf den wichtigen Spezialfall  $n = 2$  beschränken. Dieser begegnet uns im folgenden noch mehrmals. Aus rechentechnischen Gründen gleichen wir uns der Notation von [Stirzaker (1994)] an. An die Stelle des n-dim. Zufallsvektors  $\mathbf{X}$  tritt nun das Paar  $(Z, Y)^T$  und der Vektor  $\boldsymbol{\mu}$  wird zu  $(\mu_Z, \mu_Y)^T$ .

**Def. 2.10 (2-dimensionale Normalverteilung)**

Mit  $(\mu_Z, \mu_Y) \in \mathbb{R}^2$ ,  $\sigma > 0, \tau > 0$  und  $-1 < \rho < 1$  ist die sog. Kovarianzmatrix

$$\mathbf{C} := \begin{pmatrix} \sigma^2 & \rho \sigma \tau \\ \rho \sigma \tau & \tau^2 \end{pmatrix}$$

positiv definit und  $\det(\mathbf{C}) = \sigma^2 \tau^2 (1 - \rho^2)$ . Die Inverse  $\mathbf{C}^{-1}$  ist explizit berechenbar und es gilt gemäß (2.3)

$$f_{Z,Y}(z, y) = \frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)} \left( \frac{(z-\mu_Z)^2}{\sigma^2} - \frac{2\rho(z-\mu_Z)(y-\mu_Y)}{\sigma\tau} + \frac{(y-\mu_Y)^2}{\tau^2} \right)\right]. \quad (2.4)$$

Zur Bedeutung der Parameter siehe Bsp. 2.21. Um in Bsp. 2.13 (2) zu sehen, dass es sich bei  $f_{Z,Y}$  wirklich um eine Dichte handelt, benötigen wir noch den Begriff der Randverteilungen.

Die Wahl  $\mu_Z = \mu_Y = 0$  ist o.B.d.A. mit den Substitutionen  $Z - \mu_Z \mapsto \tilde{Z}$  bzw.  $Y - \mu_Y \mapsto \tilde{Y}$  möglich und beeinflusst wegen  $dz = d\tilde{z}$  bzw.  $dy = d\tilde{y}$  evtl. Integrationen nicht wesentlich. In diesem Fall vereinfacht sich die Dichte von Def. 2.10 zu

$$f_{Z,Y}(z, y) = \frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)} \left( \frac{z^2}{\sigma^2} - \frac{2\rho z y}{\sigma\tau} + \frac{y^2}{\tau^2} \right)\right]. \quad (2.5)$$

### 2.1.3 Randverteilungen

**Def. 2.11 (Randverteilung, marginal distribution)**

Die Verteilungen der Komponenten  $X_k$  eines Zufallsvektors  $\mathbf{X} = (X_1, \dots, X_n)^T$  heißen **Randverteilungen**. Die entsprechenden Verteilungsfunktionen lauten

$$F_{X_k}(z) = P(X_k \leq z) = F_{X_1, \dots, X_n}(\infty, \dots, \infty, z, \infty, \dots, \infty).$$

↑  $k$ -te Komponente.

Zur Notation  $F_{X_1, \dots, X_n}(\infty, \dots)$  vergleiche man die Vereinbarung von Folgerung 2.3.

**Satz 2.12 (Diskrete und stetige Randverteilungen)**

Mit der Definition 2.11 erhält man beispielsweise für die zweidimensionale diskrete ZV  $(Z, Y)$  die Randverteilungen (RV)

$$f_Z(z_i) = P(Z = z_i) = p_{i\cdot} = \sum_{j \in J} p_{ij} = \sum_y f_{Z,Y}(z_i, y) \quad (\text{RV bzgl. } Z)$$

$$f_Y(y_j) = P(Y = y_j) = p_{\cdot j} = \sum_{i \in I} p_{ij} = \sum_z f_{Z,Y}(z, y_j) \quad (\text{RV bzgl. } Y).$$

Im zweidimensionalen stetigen Fall gilt für die Randdichten von  $(Z, Y)$

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z,Y}(z, y) dy \quad (\text{Dichte für } Z)$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Z,Y}(z, y) dz. \quad (\text{Dichte für } Y).$$

Analog geht man für mehr als zwei Variable vor.

#### Beispiel 2.13

1. **Doppelter Münzwurf** (siehe Beispiel 2.5)

$$p_i = P(Z = i) = \begin{cases} (1-p)^2 & , \text{ falls } i = 0 \\ 2p(1-p) & , \text{ falls } i = 1 \\ p^2 & , \text{ falls } i = 2 \end{cases}$$

und

$$p_{\cdot j} = P(Y = j) = \begin{cases} p^2 & , \text{ falls } j = 0 \\ 2p(1-p) & , \text{ falls } j = 1 \\ (1-p)^2 & , \text{ falls } j = 2 \end{cases}$$

Die Wahrscheinlichkeiten für die *Randverteilung* erhält man durch zeilen- bzw. spaltenweises Aufsummieren der Zellenwahrscheinlichkeiten. Das Resultat trägt man dann zum Beispiel am "Rand" der folgenden diskreten Verteilungstabelle ein.

<b>Tabelle 2:</b> Doppelter Münzwurf $p_{ij} = P(Z = i, Y = j)$				
$i$	0	1	2	$p_{\cdot j}$
$j$				
0	0	0	$p^2$	$p^2$
1	0	$2p(1-p)$	0	$2p(1-p)$
2	$(1-p)^2$	0	0	$(1-p)^2$
$p_{i \cdot}$	$(1-p)^2$	$2p(1-p)$	$p^2$	1

## 2. Zweidimensionale Normalverteilung

Nach Satz 2.12 und mit (2.5) (dort war o.B.d.A.  $E(Z) = 0$  und  $E(Y) = 0$ ) gilt für die Randdichte von  $Z$  mit  $\sigma > 0$ ,  $\tau > 0$  und  $-1 < \rho < 1$  (siehe auch [Stirzaker (1994)], p. 269)

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f(z, y) dy = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{z^2}{\sigma^2} - \frac{2\rho zy}{\sigma\tau} + \frac{y^2}{\tau^2}\right)\right] dy \\ &= \frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}} \int_{y=-\infty}^{\infty} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\left\{\frac{y}{\tau} - \frac{\rho z}{\sigma}\right\}^2 + \frac{z^2}{\sigma^2} - \frac{\rho^2 z^2}{\sigma^2}\right)\right] dy \\ &\quad \text{und weiter mit der Substitution } u := \frac{y}{\tau} - \frac{\rho z}{\sigma}, \quad du = \frac{dy}{\tau} \\ &= \frac{1}{2\pi\sigma\sqrt{1-\rho^2}} \int_{u=-\infty}^{\infty} \exp\left(-\frac{u^2}{2(1-\rho^2)}\right) \cdot \exp\left(-\frac{1}{2(1-\rho^2)}\frac{z^2}{\sigma^2}(1-\rho^2)\right) du \\ &= \frac{1}{2\pi\sigma\sqrt{1-\rho^2}} \exp\left(\frac{-z^2}{2\sigma^2}\right) \underbrace{\int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2(1-\rho^2)}\right) du}_{\sqrt{2\pi(1-\rho^2)}} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-z^2/(2\sigma^2)}. \end{aligned} \tag{2.6}$$

Das ist die Dichte einer  $N(0, \sigma^2)$ -Verteilung. In völlig analoger Weise stellt sich  $f_Y(y)$  als Dichte einer  $N(0, \tau^2)$ -Verteilung heraus.

Nun sehen wir auch sofort ein, dass  $f(z, y)$  eine zweidimensionale Dichte ist, denn

$$\int_{z=-\infty}^{\infty} \int_{y=-\infty}^{\infty} f(z, y) dy dz = \int_{z=-\infty}^{\infty} f_Z(z) dz = \int_{z=-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sigma} e^{-z^2/(2\sigma^2)} dz = 1.$$

Diese Aussage war im Anschluss an Def. 2.10 noch offen geblieben.

### 2.1.4 Erwartungswerte; n-dim.

Es sei hier nochmals an die Vereinbarung erinnert, dass  $\mathbf{X} = (X_1, \dots, X_n)^T \in \mathbb{R}^n$  und  $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{R}^n$  mit der Indizierung  $x_k, k = 1, \dots, n$ . Wir beginnen mit dem zweidimensionalen Vektor  $(Z, Y)$ . Im diskreten Fall nehme  $Z$  die Werte  $z_i, i \in I$  und  $Y$  die Werte  $y_j, j \in J$  an.

**Def. 2.14 (Erwartungswert von  $g(Z, Y)$ )**

$(Z, Y)$  sei eine zweidimensionale ZV.  $f_{Z,Y}$  bezeichne im diskreten Fall die Wahrscheinlichkeitsfunktion und im stetigen die Dichte von  $(Z, Y)$ .  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$  sei eine (messbare) Funktion. Dann heißt

$$E g(Z, Y) := \int_{\mathbb{R}^2} g(z, y) dF_{Z,Y}(z, y)$$

$$:= \begin{cases} \sum_{i,j} g(z_i, y_j) f_{Z,Y}(z_i, y_j) & , \text{ (diskret)} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(z, y) f_{Z,Y}(z, y) dz dy & , \text{ (stetig)}. \end{cases}$$

der **Erwartungswert (EW)** von  $g(Z, Y)$ , falls  $\int_{\mathbb{R}^2} |g(z, y)| dF_{Z,Y}(z, y) < \infty$ .

**Bemerkung 2.15**

1. Die Definitionen für  $n > 2$  erfolgen analog.
2. Für beliebiges  $n$  sind verschiedene Schreibweisen in Gebrauch;  
z. B. für  $\mathbf{X} = (X_1, \dots, X_n)^T$  und  $\mathbf{x} = (x_1, \dots, x_n)^T$

$$\int_{\mathbb{R}^n} g(\mathbf{x}) dF_{\mathbf{X}}(\mathbf{x}) := \int_{\mathbb{R}^n} g(x_1, \dots, x_n) dF_{X_1, \dots, X_n}(x_1, \dots, x_n) = \int_{\mathbb{R}^n} g(\mathbf{x}) dF(\mathbf{x})$$

usw.

3. Für diskrete ZV schreiben wir auch

$$E g(Z, Y) = \sum_{i,j} g(z_i, y_j) p_{ij}, \tag{2.7}$$

mit  $p_{ij} = f_{Z,Y}(z_i, y_j), i \in I, j \in J$ .

**Satz 2.16 (EW, n-dim.)**

$F_{X_1, \dots, X_n}$  sei Verteilungsfunktion des  $n$ -dim. Zufallsvektors  $\mathbf{X} = (X_1, \dots, X_n)^T$ . Für  $k = 1, \dots, n$  bezeichne  $F_{X_k}$  die Randverteilungsfunktion der  $k$ -ten Komponente  $X_k$ .  $f_{X_1, \dots, X_n}$  und  $f_{X_k}$  seien die entsprechenden Wahrscheinlichkeitsfunktionen bzw. Dichten, dann gilt mit  $\mathbf{x} = (x_1, \dots, x_n)^T$  für  $k = 1, \dots, n$  falls alle EW existieren

$$\begin{aligned} E(X_k) &= \int_{\mathbb{R}^n} x_k dF_{X_1, \dots, X_n}(\mathbf{x}) = \int_{-\infty}^{\infty} x_k dF_{X_k}(x_k) \\ &= \int_{\mathbb{R}^n} x_k f_{X_1, \dots, X_n}(\mathbf{x}) d\mathbf{x} = \int_{-\infty}^{\infty} x_k f_{X_k}(x_k) dx_k \quad \text{falls } \mathbf{X} \text{ stetig.} \end{aligned} \quad (2.8)$$

Bei diskreten zweidimensionalen ZV  $(Z, Y)$  erhält man

$$E(Z) = \sum_{i,j} z_i f_{Z,Y}(z_i, y_j) = \sum_{i,j} z_i p_{ij} = \sum_i z_i f_Z(z_i)$$

und

$$E(Y) = \sum_{i,j} y_j f_{Z,Y}(z_i, y_j) = \sum_{i,j} y_j p_{ij} = \sum_j y_j f_Y(y_j). \quad (2.9)$$

Entsprechende Aussagen gelten für alle Momente der Randverteilungen.

**Def. 2.17 (Kovarianzen, Kovarianzmatrix)**

Für  $\mathbf{X} = (X_1, \dots, X_n)^T$  mögen alle zweiten Momente

$$\tau_{kl} := E(X_k X_l), \quad 1 \leq k, l \leq n,$$

existieren. Mit  $\mu_k := E(X_k)$  nennt man

$$\sigma_{kl} := \text{Cov}(X_k, X_l) := E[(X_k - \mu_k)(X_l - \mu_l)], \quad 1 \leq k, l \leq n,$$

die **Kovarianzen** von  $X_k$  und  $X_l$ .

$$\mathbf{C} := \mathbf{Cov}(\mathbf{X}) := (\sigma_{kl})_{1 \leq k, l \leq n} \in \mathbb{R}^{n,n}$$

heißt **Kovarianzmatrix** von  $\mathbf{X}$ .

$X_k$  und  $X_l$  heißen **unkorreliert**, falls  $\text{Cov}(X_k, X_l) = 0$ .

Die Existenz aller zweiten Momente von  $\mathbf{X}$  sichert also analog zu Satz (1.71), dass alle Erwartungswerte  $\mu_k$ , alle Varianzen  $\sigma_{kk}$  und alle Kovarianzen  $\sigma_{kl}$  der Komponenten  $X_k, X_l, 1 \leq k, l \leq n$  von  $\mathbf{X}$  wohldefiniert sind.

**Folgerung 2.18 ((Ko-)Varianzen, Verschiebungsregel)**

1. Für  $k = l$  gilt

$$\begin{aligned}\sigma_k^2 := \sigma_{kk} = \text{Cov}(X_k, X_k) &= E[(X_k - \mu_k)(X_k - \mu_k)] \\ &= E[(X_k - \mu_k)^2] = \text{Var}(X_k).\end{aligned}\quad (2.10)$$

2. Auch für Kovarianzen gibt es eine **Verschiebungsregel**, nämlich

$$\sigma_{kl} = \text{Cov}(X_k, X_l) = E[(X_k - \mu_k)(X_l - \mu_l)] = E(X_k X_l) - \mu_k \mu_l. \quad (2.11)$$

**Def. 2.19 (Korrelationskoeffizient)**

Für zwei ZV, z. B.  $(Z, Y) = (X_k, X_l)$ , sollen alle zweiten Momente existieren. Ferner gelte  $\sigma_Z^2 := \text{Var}(Z) > 0$  und  $\sigma_Y^2 := \text{Var}(Y) > 0$ . Dann heißt

$$\rho(Z, Y) := \frac{\text{Cov}(Z, Y)}{\sigma_Z \sigma_Y} = \frac{\text{Cov}(Z, Y)}{D(Z) D(Y)} = \frac{\text{Cov}(Z, Y)}{\sqrt{\text{Var}(Z) \text{Var}(Y)}}$$

der **Korrelationskoeffizient** von  $Z$  und  $Y$ .

Für unkorrelierte  $(Z, Y)$  gilt offensichtlich  $\rho(Z, Y) = 0$ . Tatsächlich ist der Korrelationskoeffizient ein Maß für den linearen Zusammenhang zwischen  $Z$  und  $Y$ :

**Satz 2.20 (Korrelationskoeffizient)**

Falls alle zweiten Momente der ZV  $(Z, Y)$  existieren und falls  $\text{Var}(Z) > 0$  und  $\text{Var}(Y) > 0$ , so gilt:

$$-1 \leq \rho(Z, Y) \leq 1$$

und  $\rho(Z, Y) = \pm 1$  genau dann, wenn

$$P(Y = a + bZ) = P(\{\omega \mid Y(\omega) = a + bZ(\omega)\}) = 1$$

für geeignete  $a \in \mathbb{R}$  und  $b \neq 0$ . Ferner gilt für  $|\rho(Z, Y)| = 1$ :

$$\rho(Z, Y) = 1 \iff b > 0 \quad \text{und} \quad \rho(Z, Y) = -1 \iff b < 0.$$

**Beispiel 2.21 (Zweidimensionale Normalverteilung)**

von  $(Z, Y)$  bzgl. Def. 2.10 mit den Parametern  $\boldsymbol{\mu} := (\mu_z, \mu_y)^T \in \mathbb{R}^2$  sowie  $\sigma_z > 0, \sigma_y > 0$  und  $-1 < \rho < 1$  und der positiv definiten Kovarianzmatrix

$$\mathbf{C} := \mathbf{Cov}(Z, Y) = \begin{pmatrix} \sigma_z^2 & \rho \sigma_z \sigma_y \\ \rho \sigma_z \sigma_y & \sigma_y^2 \end{pmatrix}.$$

Dort gilt

$$\begin{aligned} E(Z) &= \mu_z \\ E(Y) &= \mu_y \\ \text{Var}(Z) &= \sigma_z^2 \\ \text{Var}(Y) &= \sigma_y^2 \\ \rho(Z, Y) &= \rho. \end{aligned}$$

**Def. 2.22 (EW von Zufallsvektoren, -matrizen)**

Vorausgesetzt für  $k = 1, \dots, n$  existieren die Erwartungswerte  $\mu_k := E(X_k)$ , so setzt man für  $\mathbf{X} = (X_1, \dots, X_n)^T$

$$E(\mathbf{X}) := \boldsymbol{\mu} := (\mu_1, \dots, \mu_n)^T.$$

Analog (also auch elementweise) wird der Erwartungswert von Matrizen definiert, deren Elemente ZV sind.

**Satz 2.23 (EW von Zufallsvektoren)**

Für  $\mathbf{X} = (X_1, \dots, X_n)^T$  existiere  $\boldsymbol{\mu} = E(\mathbf{X})$ , dann gilt für  $\mathbf{A} \in \mathbb{R}^{m,n}$ ,  $m \in \mathbb{N}$

$$E(\mathbf{A} \mathbf{X}) = \mathbf{A} E(\mathbf{X}) = \mathbf{A} \boldsymbol{\mu}. \quad (2.12)$$



**Satz 2.24 (Kovarianzen von Zufallsvektoren; Varianz von Summen)**

Vorausgesetzt alle zweiten Momente von  $\mathbf{X} = (X_1, \dots, X_n)^T$  existieren, dann gilt mit  $\boldsymbol{\mu} = E(\mathbf{X})$ ,  $\mathbf{A} \in \mathbb{R}^{m,n}$ ,  $\mathbf{a} = (a_1, \dots, a_n)^T \in \mathbb{R}^n$ :

1. Die Kovarianzmatrix  $\mathbf{C}$  ist positiv semidefinit.

2.

$$\begin{aligned} \mathbf{C} = (\sigma_{kl}) = \mathbf{Cov}(\mathbf{X}) &= E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] \\ &= E[\mathbf{X} \mathbf{X}^T] - E[\boldsymbol{\mu} \boldsymbol{\mu}^T] \end{aligned}$$

3.

$$\mathbf{Cov}(\mathbf{A} \mathbf{X}) = \mathbf{A} \mathbf{C} \mathbf{A}^T \quad (2.13)$$

4.

$$\begin{aligned} \text{Var}(\mathbf{a}^T \mathbf{X}) = \text{Var}\left(\sum_{i=1}^n a_i X_i\right) &= \text{Cov}(\mathbf{a}^T \mathbf{X}) = \mathbf{a}^T \mathbf{C} \mathbf{a} = \sum_{k,l} a_k a_l \sigma_{kl} \\ &= \sum_{k=1}^n \sum_{l=1}^n a_k a_l \text{Cov}(X_k, X_l) \quad (2.14) \\ &= \sum_{k=1}^n a_k^2 \text{Var}(X_k) + 2 \sum_{k < l} a_k a_l \text{Cov}(X_k, X_l) \end{aligned}$$

5.

$$\begin{aligned} \text{Var}(\mathbf{a}^T \mathbf{X}) = \text{Var}\left(\sum_{i=1}^n a_i X_i\right) &= \sum_{k=1}^n a_k^2 \text{Var}(X_k), \quad (2.15) \\ &\text{falls } (X_k, X_l) \text{ unkorreliert.} \end{aligned}$$

## 2.2 Unabhängige Zufallsvariable

### Def. 2.25 (Unabhängigkeit)

Die Komponenten  $X_k$  eines Zufallsvektors  $\mathbf{X} = (X_1, \dots, X_n)^T$  heißen genau dann unabhängig, wenn die gemeinsame Verteilungsfunktion in das Produkt der Randverteilungsfunktionen zerfällt, d. h.

$$\begin{aligned} F_{X_1, \dots, X_n}(x_1, \dots, x_n) &= P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \\ &= P(X_1 \leq x_1) \cdot \dots \cdot P(X_n \leq x_n) \\ &= F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n). \end{aligned} \quad (2.16)$$

Für diskrete bzw. stetige ZV ist dazu äquivalent die Aussage dass die gemeinsame Wahrscheinlichkeitsfunktion bzw. Dichte in das Produkt der Randwahrscheinlichkeitsfunktionen bzw. der Randdichten zerfallen, d. h.

$$f_{X_1, \dots, X_n} = f_{X_1} \cdot \dots \cdot f_{X_n}.$$

Äquivalent zu Def. (2.16) ist folgende Charakterisierung:

$X_1, \dots, X_n$  sind genau dann unabhängig, wenn für beliebige Borelmengen  $B_1, \dots, B_n \subset \mathbb{R}$  gilt

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \cdot \dots \cdot P(X_n \in B_n). \quad (2.17)$$

### Satz 2.26 (Unabhängigkeit, diskret)

Beispielsweise gilt für unabhängige zweidimensionale diskrete ZV  $(Z, Y)$

$$\begin{aligned} f_{Z,Y}(z_i, y_j) &= P(Z = z_i, Y = y_j) = p_{ij} \\ &= p_{i \cdot} \cdot p_{\cdot j} = P(Z = z_i) \cdot P(Y = y_j) = f_Z(z_i) \cdot f_Y(y_j) \end{aligned}$$

für alle  $i, j$ .

### Beispiel 2.27

1. **Zweimaliges Würfeln** (siehe Beispiel 2.5): Im gewählten Modell gilt

$$P(Z = i, Y = j) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = P(Z = i)P(Y = j)$$

für  $1 \leq i, j \leq 6$ .

## 2. Unabhängige Münze und Würfel

Im gewählten Modell gilt

$$P(Z = i, Y = j) = \frac{1}{12} = \frac{1}{6} \cdot \frac{1}{2} = P(Z = i)P(Y = j), \quad 1 \leq i \leq 6, j = 0, 1.$$

Tabelle 3: Unabhängige Münze und Würfel $p_{ij} = P(Z = i, Y = j)$							
$i$	1	2	3	4	5	6	$p_{\cdot j}$
$j$							
0	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
$p_{i \cdot}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

3. Bei den beiden vorangegangenen Beispielen sind die  $p_{ij}$  die Produkte der Randwahrscheinlichkeiten  $p_{i \cdot}$  und  $p_{\cdot j}$ . Das trifft übrigens nicht auf das Beispiel 2.5 von der Summe Kopf / Wappen beim zweimaligen Münzwurf zu; vgl. auch Beispiel 2.13.

## 4. Zweidimensionale Gleichverteilung

Die Komponenten der zweidimensionalen ZV mit der Dichte

$$f_{Z,Y}(z, y) = 1_{(0,1) \times (0,1)}(z, y) = 1_{(0,1)}(z) \cdot 1_{(0,1)}(y) = f_Z(z) \cdot f_Y(y)$$

sind unabhängig und jeweils  $U(0, 1)$ -verteilt.

## 5. 2-dim. und n-dim. Normalverteilung

$Z$  und  $Y$  sind genau dann unabhängig, wenn  $\rho = 0$  ist. In diesem Fall gilt mit den Bezeichnungen von Def. 2.10

$$f_{Z,Y}(z, y) = \frac{1}{2\pi\sigma\tau} \exp \left[ -\frac{(z - \mu_Z)^2}{2\sigma^2} - \frac{(y - \mu_Y)^2}{2\tau^2} \right] = f_Z(z) \cdot f_Y(y).$$

Die Komponenten  $X_k$  und  $X_l$  eines n-dim. normalverteilten Vektors  $\mathbf{X}$  sind genau dann unabhängig, wenn  $\rho(X_k, X_l) = 0$ .

Alle Komponenten eines normalverteilten Vektors  $\mathbf{X}$  sind genau dann unabhängig, wenn  $\mathbf{Cov}(\mathbf{X}) = \mathbf{diag}\{Var(X_1), \dots, Var(X_n)\}$ .

**Satz 2.28 (Unabhängig-unkorreliert; Unabhängigkeit von Funktionen)**

Seien  $Z$  und  $Y$  unabhängig mit existierenden 2. Momenten und  $\mu_Z = E(Z)$ ,  $\mu_Y = E(Y)$  sowie  $g, h : \mathbb{R} \rightarrow \mathbb{R}$  Borel-messbare Funktionen, z. B. (stückweise) stetig, dann gilt:

1.  $g(Z)$  und  $h(Y)$  sind unabhängig.
2.  $E[g(Z) \cdot h(Y)] = E[g(Z)] \cdot E[h(Y)]$  oder speziell
3.  $E(ZY) = E(Z)E(Y)$ ; insbesondere sind  $Z$  und  $Y$  unkorreliert, da  $E[(Z - \mu_Z)(Y - \mu_Y)] = E(Z - \mu_Z)E(Y - \mu_Y) = 0$ .
4. Die Umkehrung von 3. gilt nicht.
5. Falls  $(Z, Y)$  gemeinsam (zweidimensional) normalverteilt sind, dann folgt:

$$(Z, Y) \text{ unkorreliert} \iff Z \text{ und } Y \text{ sind unabhängig.}$$

Unkorreliertheit bedeutet, dass kein **linearer Zusammenhang** zwischen den entsprechenden ZV vorliegt. Dies ist i.a. eine **schwächere Bedingung** als *Unabhängigkeit* von ZV. Im Falle einer gemeinsamen Normalverteilung der ZV sind diese Begriffe äquivalent.

Später werden wir für die Erwartungswerte unabhängiger ZV  $Z$  und  $Y$  häufig den Produktsatz

$$E[g(Z) \cdot h(Y)] = E g(Z) \cdot E h(Y)$$

benützen.

**Def. 2.29 (Unabhängig identisch verteilt, iid)**

Die ZV  $X_1, \dots, X_n$  heißen **unabhängig identisch verteilt** wie  $X$ ; man sagt auch  $X, X_1, \dots, X_n$  sind **iid** (*independent identically distributed*), genau dann, wenn die  $X_i$  unabhängig sind und

$$F_X = F_{X_i}, \quad i = 1, \dots, n.$$

*iid*-Variablen  $X_1, \dots, X_n$  heißen auch **mathematische Stichprobe**.

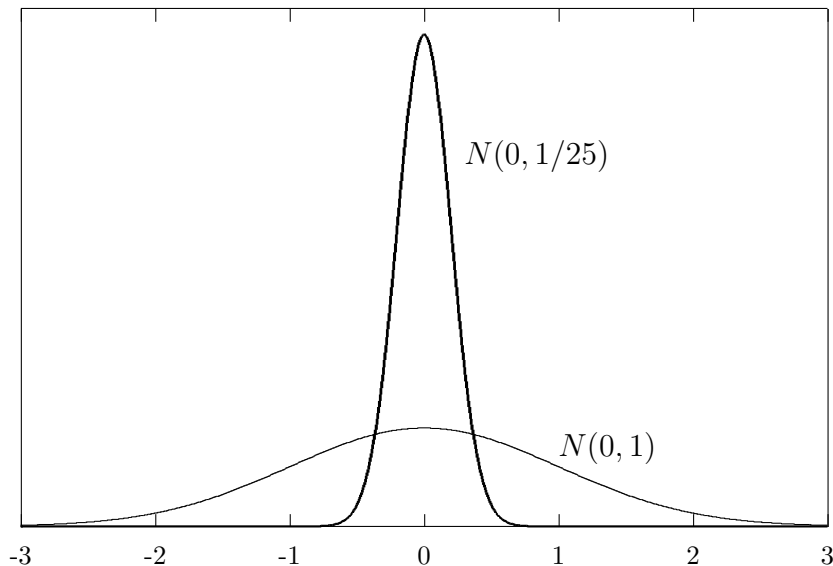
**Satz 2.30 (Stichprobenmittel  $\bar{X}$ )**

$X_1, \dots, X_n$  seien iid mit  $E(X_i) = \mu$  und  $\text{Var}(X_i) = D^2(X_i) = \sigma^2$ ,  $i = 1, \dots, n$ . Für das arithmetische Mittel

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

einer solchen (mathematischen) Stichprobe gilt  $E(\bar{X}) = \mu$  sowie

$$\begin{aligned} \text{Var}(\bar{X}) = D^2(\bar{X}) &= \frac{\sigma^2}{n} \\ D(\bar{X}) &= \frac{\sigma}{\sqrt{n}} \quad (\text{Standardabweichung}). \end{aligned}$$



**Bild 9:** Verteilung von  $\bar{X}$  für  $n = 1$  und  $n = 25$

In der mathematischen Statistik geht man davon aus, dass die eine Verteilung charakterisierenden Parameter, u.a.  $\mu$  und  $\sigma^2$ , unbekannt sind und mithilfe von Stichprobenziehungen geschätzt werden müssen; siehe auch Kap. 8.3.3.  $\bar{X}$  ist die am häufigsten verwendete Schätzung für  $\mu$ . Als Schätzfunktion für die i. a. unbekannte Varianz  $\sigma^2$  wählt man

$$S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (2.18)$$

Der Vorfaktor  $1/(n-1)$  ist nötig, um die sog. *Erwartungstreue*  $E[S^2] = \sigma^2$  zu garantieren; vgl. z. B. [Lehn & Wegmann (1992)], S. 77 oder Bsp. 2.34. Dort wird u.a. auch nachgewiesen, dass für normalverteilte iid-Variable das Stichprobenmittel  $\bar{X}$  und  $S^2$  unabhängige ZV sind.

## 2.3 Transformationen von $n$ Zufallsvariablen

### 2.3.1 Transformationssatz für Dichten

Die  $n$ -dimensionale Zufallsvariable  $\mathbf{X} = (X_1, \dots, X_n)^T$  habe eine Dichte  $f(\mathbf{x})$ ,  $\mathbf{x} = (x_1, \dots, x_n)^T$ . Weiterhin seien  $B := \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) > 0\}$  und  $\mathbf{h} : B \rightarrow D$ ,  $D \subset \mathbb{R}^n$ , differenzierbar und bijektiv. Dann ist  $\mathbf{Y} := \mathbf{h}(\mathbf{X})$  wieder eine  $n$ -dim. ZV. Mit  $\mathbf{y} = (y_1, \dots, y_n)^T$  sei die Umkehrabbildung

$$\mathbf{x} := \mathbf{h}^{-1}(\mathbf{y}) = (x_1(\mathbf{y}), \dots, x_n(\mathbf{y}))^T$$

auf  $D$  (abgesehen von endlich vielen glatten Hyperflächen) stetig differenzierbar.

$$\frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} = \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)}(\mathbf{y}) := \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1}(\mathbf{y}) & \dots & \frac{\partial x_1}{\partial y_n}(\mathbf{y}) \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1}(\mathbf{y}) & \dots & \frac{\partial x_n}{\partial y_n}(\mathbf{y}) \end{pmatrix}$$

sei die Funktionaldeterminante der Umkehrabbildung  $\mathbf{h}^{-1}$ . Dann hat die ZV  $\mathbf{Y} = (Y_1, \dots, Y_n)^T := \mathbf{h}(\mathbf{X})$  wieder eine Dichte  $g(\mathbf{y})$ , die gegeben ist durch

$$g(\mathbf{y}) = g(y_1, \dots, y_n) = f(x_1(\mathbf{y}), \dots, x_n(\mathbf{y})) \left| \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \right|, \quad (2.19)$$

wobei  $|\cdot|$  hier den Betrag einer Determinante bezeichnet.

$(V, W)$  seien stetig mit Dichte  $f_{V,W}$ . Man betrachtet die Transformation

$$X = x(V, W), \quad Y = y(V, W),$$

zu der nach Voraussetzung die Umkehrtransformation

$$V = v(X, Y), \quad W = w(X, Y)$$

mit auf  $D$  (abgesehen von endlich vielen glatten Kurven) stetig differenzierbaren  $v(\cdot, \cdot)$  und  $w(\cdot, \cdot)$  existiert. Dann gilt

$$f_{X,Y}[x, y] = f_{V,W}[v(x, y), w(x, y)] |J(x, y)|, \quad (2.20)$$

wobei die Funktionaldeterminante hier speziell gegeben ist durch

$$J(x, y) = \frac{\partial(v, w)}{\partial(x, y)} = \frac{\partial v}{\partial x} \frac{\partial w}{\partial y} - \frac{\partial v}{\partial y} \frac{\partial w}{\partial x}.$$

Es folgen Beispiele und wichtige Anwendungen.

#### Beispiel 2.31 (Transformationsregel für Polarkoordinaten)

Seien  $x = r \cos \varphi$ ,  $y = r \sin \varphi$ ,  $r \geq 0$ ,  $0 \leq \varphi < 2\pi$ ,  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ ,  $B \subseteq \mathbb{R}^2$ , dann gilt

$$\iint_{B(x,y)} f(x, y) dx dy = \iint_{B(r,\varphi)} f(r \cos \varphi, r \sin \varphi) r dr d\varphi, \quad (2.21)$$

falls eines der beiden Integrale existiert.

Eine wichtige Anwendung ist die **Polar-Transformation** zum Simulieren normalverteilter Zufallsvariablen nach Box-Mutter:

Seien  $(S, T)$  unabhängig gleichverteilt mit Dichte  $f_{S,T}(s, t) = 1_{(0,1) \times (0,1)}(s, t)$ , dann sind

$$\begin{aligned} X &:= \sqrt{-2 \ln S} \sin(2\pi T) \\ Y &:= \sqrt{-2 \ln S} \cos(2\pi T) \end{aligned} \quad (2.22)$$

unabhängig und jeweils  $N(0, 1)$ -verteilt. Die Umkehrtransformation lautet für  $(x, y) \in \mathbb{R}^2 \setminus \{(0, y) \in \mathbb{R}^2 \mid y \geq 0\}$

$$\begin{aligned} s &= e^{-(x^2+y^2)/2} \\ t &= \begin{cases} \frac{1}{2\pi} \arccos \frac{y}{\sqrt{x^2+y^2}} & , \text{ falls } x \geq 0 \\ \frac{1}{2\pi} \left( 2\pi - \arccos \frac{y}{\sqrt{x^2+y^2}} \right) & , \text{ falls } x < 0. \end{cases} \end{aligned}$$

Die Funktionaldeterminante erhält man elegant mit einem kleinen Umweg über die Relation

$$\frac{\partial(s, t)}{\partial(x, y)} = 1 / \frac{\partial(x, y)}{\partial(s, t)}$$

und

$$\begin{aligned} \frac{\partial(x, y)}{\partial(s, t)} &= \det \begin{pmatrix} -\frac{\sin(2\pi t)}{s\sqrt{-2 \ln s}} & 2\pi \sqrt{-2 \ln s} \cos(2\pi t) \\ -\frac{\cos(2\pi t)}{s\sqrt{-2 \ln s}} & -2\pi \sqrt{-2 \ln s} \sin(2\pi t) \end{pmatrix} \\ &= \frac{2\pi}{s} \sin^2(2\pi t) + \frac{2\pi}{s} \cos^2(2\pi t) = \frac{2\pi}{s}. \end{aligned}$$

Damit gilt

$$\frac{\partial(s, t)}{\partial(x, y)} = \frac{s}{2\pi} = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$$

und man erhält mit dem Transformationssatz für Dichten

$$f_{X,Y}(x, y) = f_{S,T}[s(x, y), t(x, y)] \left| \frac{\partial(s, t)}{\partial(x, y)} \right| = 1 \cdot \frac{1}{2\pi} e^{-(x^2+y^2)/2}.$$

### 2.3.2 Lineare Transformationen

Wir betrachten nun eine weitere wichtige Klasse linearer Transformationen und folgen in der Darstellung [Stirzaker (1994)], pp. 287.

**Satz 2.32 (Lineare Transformation)**

Zur Matrix  $\mathbf{A} = (a_{ij})$  existiere die Inverse  $\mathbf{A}^{-1} = (b_{ij}) = \mathbf{B}$ . Weiterhin betrachten wir  $\mathbf{X} = (X_1, \dots, X_n)^T$ ,  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$  und

$$\mathbf{Y} = \mathbf{A} \mathbf{X}, \quad \mathbf{X} = \mathbf{B} \mathbf{Y} \quad \text{bzw. für } i = 1, \dots, n : \quad Y_i = \sum_{j=1}^n a_{ij} X_j, \quad X_i = \sum_{j=1}^n b_{ij} Y_j.$$

Aus der gemeinsamen Dichte  $f_X(x_1, \dots, x_n)$  der  $(X_1, \dots, X_n)$  kann dann wegen  $\det(\mathbf{A}) = 1/\det(\mathbf{B}) \neq 0$  die gemeinsame Dichte  $f_Y(y_1, \dots, y_n)$  von  $(Y_1, \dots, Y_n)$  berechnet werden. Es gilt gemäß Gleichung (2.19) mit  $\mathbf{y} = (y_1, \dots, y_n)^T$

$$\begin{aligned} f_Y(y_1, \dots, y_n) &= \frac{1}{|\det(\mathbf{A})|} f_X(x_1(y_1, \dots, y_n), \dots, x_n(y_1, \dots, y_n)) \\ &= |\det(\mathbf{B})| f_X(x_1, \dots, x_n) = |\det(\mathbf{A}^{-1})| f_X(\mathbf{A}^{-1} \mathbf{y}). \end{aligned}$$

Für affin lineare Transformationen  $\mathbf{Y} = \mathbf{A} \mathbf{X} + \mathbf{b}$  mit  $\det(\mathbf{A}) \neq 0$  gilt entsprechend

$$f_Y(\mathbf{y}) = |\det(\mathbf{A}^{-1})| f_X(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})). \quad (2.23)$$

**Beispiel 2.33 (n-dimensionale Normalverteilung)**

Nun können wir zeigen, dass das bereits in (2.3) eingeführte

$$f_X(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C})}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \quad (2.24)$$

mit  $\mathbf{x} \in \mathbb{R}^n$  und den Parametern  $\boldsymbol{\mu} \in \mathbb{R}^n$  sowie  $\mathbf{C} \in \mathbb{R}^{n,n}$ , positiv definit, eine n-dim. Dichte ist.

Da  $\mathbf{C}$  per Definition auch symmetrisch ist, hat es nur reelle Eigenwerte  $\lambda_i$ , die wir zu der Matrix  $\boldsymbol{\Lambda} := \mathbf{diag}\{\lambda_1, \dots, \lambda_n\}$  zusammenfassen. Wegen der positiven Definitheit folgt  $\lambda_i > 0$ ,  $i = 1, \dots, n$ . Außerdem gibt es ein System von  $n$  orthonormierten Eigenvektoren, die wir spaltenweise zu einer Matrix  $\mathbf{U}$  zusammenfassen. Dann lautet die Eigenwert-Eigenvektorzerlegung

$$\mathbf{C} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T, \quad \mathbf{U} \mathbf{U}^T = \mathbf{U}^T \mathbf{U} = \mathbf{I}_n = \mathbf{diag}\{1, \dots, 1\}.$$

Nun seien mit  $\boldsymbol{\Lambda}^{1/2} := \mathbf{diag}\{\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}\}$

$$\mathbf{C}^{1/2} := \mathbf{U} \boldsymbol{\Lambda}^{1/2} \mathbf{U}^T \in \mathbb{R}^{n,n} \quad \text{sowie} \quad \mathbf{y} := \mathbf{C}^{-1/2} (\mathbf{x} - \boldsymbol{\mu}) \in \mathbb{R}^n.$$

Wegen  $\det(\mathbf{C}) = \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_n > 0$  gilt  $\det(\mathbf{C}^{1/2}) = \sqrt{\det(\mathbf{C})} > 0$  sowie  $\det(\mathbf{C}^{-1/2}) = 1/\sqrt{\det(\mathbf{C})}$ . Nach Satz 2.32 lautet dann die Dichte von  $\mathbf{Y} := \mathbf{C}^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \in \mathbb{R}^n$  mit



$$\mathbf{x} - \boldsymbol{\mu} = \mathbf{C}^{1/2} \mathbf{y}$$

$$\begin{aligned} f_Y(y_1, \dots, y_n) &= \frac{1}{|\det(\mathbf{C}^{-1/2})|} f_X(\boldsymbol{\mu} + \mathbf{C}^{1/2} \mathbf{y}) \\ &= \frac{\det(\mathbf{C}^{1/2})}{\sqrt{(2\pi)^n \det(\mathbf{C})}} \exp \left[ -\frac{1}{2} \mathbf{y}^T \mathbf{C}^{1/2} \mathbf{C}^{-1} \mathbf{C}^{1/2} \mathbf{y} \right] \\ &= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2} \mathbf{y}^T \mathbf{y}} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-y_i^2/2}. \end{aligned}$$

Nun erhält man

$$\int_{\mathbb{R}^n} f_Y(y_1, \dots, y_n) d(y_1, \dots, y_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y_i^2/2} dy_i = 1.$$

Damit sind  $f_Y$  und  $f_X$  Dichten.

### Satz 2.34 (Orthogonale Transformation bei Normalverteilung)

Seien  $(X_1, \dots, X_n)$   $N(0, 1)$ -iid Zufallsvariablen und  $\mathbf{A} = (a_{ij})$  eine orthogonale Matrix mit  $\det(\mathbf{A}) = \pm 1$  sowie  $\mathbf{A}^{-1} = \mathbf{A}^T$ . Weiterhin gelte mit  $\mathbf{X} = (X_1, \dots, X_n)^T$  und  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$  die Beziehung  $\mathbf{Y} = \mathbf{A} \mathbf{X}$ , d. h.

$$Y_i = \sum_{j=1}^n a_{ij} X_j, \quad 1 \leq i \leq n. \quad (2.25)$$

Dann gilt:

1.  $(Y_1, \dots, Y_n)$  sind unabhängige  $N(0, 1)$ -verteilte Zufallsvariablen.

2. Das Stichprobenmittel  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  und die

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

3. Für  $N(\boldsymbol{\mu}, \sigma^2)$ -iid ZV  $(X_1, \dots, X_n)$  gilt  $E(S^2) = \sigma^2$ .

Beweis:

1. Es gilt  $\mathbf{X} = \mathbf{A}^T \mathbf{Y}$  und

$$\sum_{i=1}^n X_i^2 = \mathbf{X}^T \mathbf{X} = \mathbf{Y}^T \mathbf{A} \mathbf{A}^T \mathbf{Y} = \mathbf{Y}^T \mathbf{Y} = \sum_{i=1}^n Y_i^2.$$

Die unabhängigen  $N(0, 1)$ -verteilten  $X_i$  haben die Dichte  $(2\pi)^{-n/2} \exp\left(-\frac{1}{2} \sum x_i^2\right)$ . Somit gilt für die Dichte von  $(Y_1, \dots, Y_n)$  nach Satz 2.32

$$f_Y(y_1, \dots, y_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2} \sum_{i=1}^n y_i^2\right).$$

Deshalb sind  $(Y_1, \dots, Y_n)$  unabhängig  $N(0, 1)$ -verteilt.

2. Nun sei  $\mathbf{A} = (a_{ij})$  orthogonal mit spezieller erster Zeile der Form  $a_{1j} = 1/\sqrt{n}$ , was

$$Y_1 = \sum_{j=1}^n \frac{1}{\sqrt{n}} X_j = \sqrt{n} \bar{X}$$

zur Folge hat. Weiterhin gilt

$$\begin{aligned} (n-1)S^2 &= \sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + n\bar{X}^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2 \\ &= \sum_{i=1}^n Y_i^2 - Y_1^2 = \sum_{i=2}^n Y_i^2. \end{aligned} \quad (2.26)$$

$S^2$  ist unabhängig von  $\bar{X}$ , da  $Y_1$  gemäß 1. unabhängig von  $(Y_2, \dots, Y_n)$  ist.

3. Ausgehend von  $N(\mu, \sigma^2)$ -iid  $\tilde{X}_i$  führt man zunächst eine Translation der Form  $\tilde{X}_i - \mu \rightarrow X_i$ ,  $i = 1, \dots, n$  durch. Dann sind die  $X_i$  iid und  $N(0, \sigma^2)$ . Durch Einsetzen gemäß 1. verifiziert man sofort, dass die orthogonale Transformation (2.25)  $N(0, \sigma^2)$  iid  $Y_i$  liefert. Damit folgt gemäß 2. und (2.26)

$$E[(n-1)S^2] = E\left[\sum_{i=2}^n Y_i^2\right] = (n-1)\sigma^2. \quad \square$$

## Lineare Transformationen normalverteilter ZV

Aus Satz 2.32 folgt noch, dass allgemeine lineare Transformationen normalverteilter ZV wieder normalverteilt sind:

Seien  $\mathbf{X} \sim N(\boldsymbol{\mu}, \mathbf{C})$  und  $\mathbf{A} \in \mathbb{R}^{m,n}$ ,  $\text{Rang}(\mathbf{A}) = m > 0$  gegeben. Dann gilt

$$\mathbf{Y} := \mathbf{A} \mathbf{X} \sim N(\mathbf{A} \boldsymbol{\mu}, \mathbf{A} \mathbf{C} \mathbf{A}^T). \quad (2.27)$$

Zum Beweis ergänze man  $\mathbf{A}$  zu einer nichtsingulären Matrix.

### 2.3.3 Summen, Quotienten

Wir betrachten einfache Funktionen von zwei ZV  $(X, Y)$  mit Wahrscheinlichkeitsfunktion bzw. Dichte  $f = f_{X,Y}$ . Für die Summe  $V := X + Y$  erhält man

$$F_{X+Y}(v) = P(V \leq v) = P(X + Y \leq v),$$

was im diskreten Fall

$$f_V(v) = f_{X+Y}(v) = P(X + Y = v) = P(X = v - Y) = \sum_y f(v - y, y) = \sum_x f(x, v - x)$$

nach sich zieht. Für stetige  $(X, Y)$  lautet die Dichte

$$f_V(v) = f_{X+Y}(v) = \int_{-\infty}^{\infty} f(x, v - x) dx = \int_{-\infty}^{\infty} f(v - y, y) dy.$$

Sind  $X$  und  $Y$  zusätzlich unabhängig, so kann die Verteilung der Summe elegant mit den erzeugenden bzw. momenterzeugenden Funktionen von Kapitel 3 gewonnen werden. Der hier gewählte Ansatz führt auf sogenannte **Faltungen** (siehe auch Def. 3.6):

$$f_{X+Y}(v) = \sum_y f_X(v - y) f_Y(y) = \sum_x f_X(x) f_Y(v - x) \quad (2.28)$$

bzw. für Dichten

$$f_{X+Y} = \int_{-\infty}^{\infty} f_X(v - y) f_Y(y) dy = \int_{-\infty}^{\infty} f_X(x) f_Y(v - x) dx. \quad (2.29)$$

Häufig benötigt man die Dichte des Quotienten stetiger ZV. Diese lautet

$$f_{X/Y}(v) = \int_{-\infty}^{\infty} f(x, xv) |x| dx = \int_{-\infty}^{\infty} f(yv, y) |y| dy. \quad (2.30)$$

#### Beispiel 2.35 (Summen bei zweidimensionaler Normalverteilung)

Gemäß (2.5) haben  $X$  und  $Y$  eine zweidimensionale Normalverteilung mit  $E(X) = E(Y) = 0$  sowie  $D(X) = \sigma > 0$  und  $D(Y) = \tau > 0$ , falls die gemeinsame Dichte mit  $-1 < \rho < 1$  lautet

$$f(x, y) = \frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma^2} - \frac{2\rho xy}{\sigma\tau} + \frac{y^2}{\tau^2}\right)\right].$$

Gesucht ist die Dichte von  $aX + bY$  für Konstanten  $a$  und  $b$ .

Lösung (gemäß [Stirzaker (1994)], p. 269, 275):

Zunächst zeigt Bsp. 2.13 (2), dass sich für  $a = 0$  und  $b = 1$  die Randdichte einer

$N(0, \tau^2)$ -Verteilung ergibt. Analog erhält man für  $a = 1$  und  $b = 0$  eine  $N(0, \sigma^2)$ -Verteilung. Nun seien  $a \neq 0$  und  $b \neq 0$ . Die gemeinsame Dichte von  $U = aX$  und  $V = bY$  lautet

$$g(u, v) = \frac{1}{|ab|} f\left(\frac{u}{a}, \frac{v}{b}\right).$$

Die Dichte von  $Z = U + V = aX + bY$  ist dann

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|ab|} f\left(\frac{u}{a}, \frac{z-u}{b}\right) du.$$

Neuordnung des Exponenten unter dem Integral liefert

$$\frac{-1}{2(1-\rho^2)} \left( \frac{u^2}{a^2\sigma^2} - \frac{2\rho u(z-u)}{ab\sigma\tau} + \frac{(z-u)^2}{b^2\tau^2} \right) = \frac{-\alpha}{2(1-\rho^2)} \left[ \left( u - \frac{\beta z}{\alpha} \right)^2 + \frac{z^2}{\alpha^2} \frac{1-\rho^2}{a^2b^2\sigma^2\tau^2} \right],$$

wobei  $\alpha = \frac{1}{a^2\sigma^2} + \frac{2\rho}{ab\sigma\tau} + \frac{1}{b^2\tau^2}$  und  $\beta = \frac{\rho}{ab\sigma\tau} + \frac{1}{b^2\tau^2}$ .

Setzt man  $u = \beta z/\alpha$  im Integranden, so folgt

$$\int_{-\infty}^{\infty} \exp\left(-\frac{\alpha v^2}{2(1-\rho^2)}\right) dv = \sqrt{\frac{2\pi(1-\rho^2)}{\alpha}}.$$

Nach einigem Rechnen ergibt sich dann

$$f_Z(z) = \frac{1}{\sqrt{2\pi\xi^2}} \exp\left(-\frac{z^2}{2\xi^2}\right)$$

mit  $\xi^2 = a^2\sigma^2 + 2\rho ab\sigma\tau + b^2\tau^2$ . Somit gilt  $Z \sim N(0, \xi^2)$ .

Dieses Ergebnis hätte man übrigens auch durch spezielle Wahl von

$$\mathbf{A} = \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} Z \\ W \end{pmatrix} = \mathbf{A} \begin{pmatrix} X \\ Y \end{pmatrix}$$

in Formel (2.27) erhalten, wobei die zweite Zeile der Matrix nicht eindeutig festgelegt, aber so wählen ist, dass  $\mathbf{A}$  nichtsingulär ist.

Einen interessanten Spezialfall erhält man für  $\rho = 0$ . Mit dem obigen Ergebnis ist dann folgendes bewiesen:

Seien  $X$  bzw.  $Y$  unabhängig  $N(0, \sigma^2)$  bzw.  $N(0, \tau^2)$ -verteilte Zufallsvariablen. Dann hat die Summe  $Z = aX + bY$  eine  $N(0, a^2\sigma^2 + b^2\tau^2)$ -Verteilung.

Durch Anwenden der Transformationen  $X - \mu \rightarrow \tilde{X}$  und  $Y - \eta \rightarrow \tilde{Y}$  folgt zunächst für unabhängige  $N(\mu, \sigma^2)$  bzw.  $N(\eta, \tau^2)$ -verteilte  $X$  bzw.  $Y$ , dass die Summe  $Z := c + aX + bY$  nach  $N(c + \mu + \eta, a^2\sigma^2 + b^2\tau^2)$  verteilt ist. Oder noch allgemeiner:

**Satz: Summe von normalverteilten Zufallsvariablen**

Sind  $(X, Y)$  gemeinsam normalverteilt mit  $E(X) = \mu$ ,  $E(Y) = \eta$ ,  $\text{Var}(X) = \sigma^2 > 0$ ,  $\text{Var}(Y) = \tau^2 > 0$  und Korrelationskoeffizient  $-1 < \rho < 1$ . Für Koeffizienten  $a, b$  mit  $a^2 + b^2 > 0$ , ist  $Z := c + aX + bY$  wieder normalverteilt mit

$$E(Z) = c + a\mu + b\eta \quad \text{und} \quad \text{Var}(Z) = a^2\sigma^2 + 2ab\rho\sigma\tau + b^2\tau^2.$$

**Beispiel 2.36 (Summen und Quotienten von ZV)**

1.  $X \sim B(n_X, p)$  und  $Y \sim B(n_Y, p)$  unabhängig  $\implies X + Y \sim B(n_X + n_Y, p)$ .
2.  $X \sim \text{Poi}(\lambda_X)$ ,  $Y \sim \text{Poi}(\lambda_Y)$  unabhängig  $\implies X + Y \sim \text{Poi}(\lambda_X + \lambda_Y)$ .
3.  $X \sim N(\mu_X, \sigma_X^2)$ ,  $Y \sim N(\mu_Y, \sigma_Y^2)$  unabh.  $\implies X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$ .
4.  $X \sim \Gamma(q_X, \lambda)$ ,  $Y \sim \Gamma(q_Y, \lambda)$  unabhängig  $\implies X + Y \sim \Gamma(q_X + q_Y, \lambda)$ .
5.  $X$  und  $Y$  unabhängig  $N(0, 1)$   $\implies Z := X/Y$  ist Cauchy-verteilt mit Dichte

$$f_{X/Y}(z) = \frac{1}{\pi} \frac{1}{1 + z^2}.$$

Siehe auch die Beispiele 3.9 und 3.14.

**2.3.4 Minimum, Maximum****Satz 2.37 (Minimum und Maximum von ZV)**

$X, X_1, \dots, X_n$  seien iid-ZV mit Verteilungsfunktion  $F_X$  und  $\bar{F}_X(x) := 1 - F_X(x)$  bezeichne wiederum die sog. tail probability. Dann lauten die Verteilungsfunktionen  $F_Z$  von  $Z := \max_i \{X_i\}$  bzw.  $F_Y$  von  $Y := \min_i \{X_i\}$

$$\begin{aligned} F_Z(z) = F_{\max\{X_i\}}(z) &= [F_X(z)]^n \\ 1 - F_Y(y) = 1 - F_{\min\{X_i\}} &= [1 - F_X(y)]^n \end{aligned} \quad (2.31)$$

oder

$$\bar{F}_Y(y) = \bar{F}_{\min\{X_i\}}(z) = [\bar{F}_X(y)]^n. \quad (2.32)$$

Beweis: Wegen der Unabhängigkeit der  $X_i$  gilt

$$\begin{aligned} F_{\max\{X_i\}}(z) &= P(\max_i \{X_i\} \leq z) \\ &= P(X_1 \leq z, \dots, X_n \leq z) \\ &= P(X_1 \leq z) \cdot \dots \cdot P(X_n \leq z) = [F_X(z)]^n. \end{aligned}$$

$$\begin{aligned} \bar{F}_{\min\{X_i\}}(y) &= P(\min_i \{X_i\} > y) \\ &= P(X_1 > y, \dots, X_n > y) \\ &= P(X_1 > y) \cdot \dots \cdot P(X_n > y) = [\bar{F}_X(y)]^n. \quad \square \end{aligned}$$

**Beispiel 2.38 (Gleich- und Exponentialverteilung)**

$X, X_1, \dots, X_n$  seien iid-ZV mit Verteilungsfunktion  $F_X$ .

1. Gleichverteilung:  $X \sim U(0, a)$ ,  $a > 0$ ,  $Z := \max_i \{ X_i \}$

Man hat für  $0 < x < a$  die Darstellung  $F_X(x) = \frac{x}{a}$ , woraus sich nach (2.31)

$F_Z(z) = \left(\frac{z}{a}\right)^n$  für  $0 < z < a$  ableitet. Damit lautet die Dichte  $f_Z$  von  $Z$

$$f_Z(z) = \frac{n z^{n-1}}{a^n} 1_{0,a}(z).$$

2. Exponentialverteilung:  $X \sim ED(\lambda)$ ,  $\lambda > 0$ ,  $Y := \min_i \{ X_i \}$

Es gilt  $\bar{F}_X(x) = e^{-\lambda x}$  für  $x > 0$  und gemäß (2.32) ergibt sich

$$\bar{F}_Y(y) = \left(e^{-\lambda y}\right)^n 1_{(0,\infty)}(y) = e^{-n\lambda y} 1_{(0,\infty)}(y).$$

Damit ist  $Y$  auch  $ED(n\lambda)$ -exponentialverteilt und

$$E(Y) = \frac{1}{n\lambda} = \frac{E(X)}{n}.$$

## 2.4 Bedingte Verteilungen und Erwartungswerte

Schon bei einfachen kombinatorischen Aufgaben hat sich der Übergang zu bedingten Wahrscheinlichkeiten als vorteilhaft erwiesen. In vielen Fragestellungen der Stochastik kommt man ohne das sog. *Conditioning* überhaupt nicht weiter. Deswegen soll hier der Kalkül der **bedingten Erwartung** am Beispiel diskreter Zufallsvariabler (ZV) eingeführt werden. Der Vollständigkeit halber werden gelegentlich auch Formeln für Dichten mit angegeben. Wir folgen in der Darstellung grob dem Konzept von [Stirzaker (1994)], p.94 und p. 138.

### 2.4.1 Bedingung bzgl. Ereignis $B$

**Def. 2.39 (Bedingte Verteilung bzgl. Ereignis  $B$ )**

$X$  sei eine diskrete ZV bzgl.  $(\Omega, \mathfrak{F}, P)$  und  $B \in \mathfrak{F}$  ein Ereignis mit  $P(B) > 0$ . Bezeichne weiterhin  $A_x := \{\omega \mid X(\omega) = x\}$  für die höchstens abzählbar vielen Werte von  $X$ . Dann definiert

$$f(x|B) := P(A_x|B) = \frac{P(A_x \cap B)}{P(B)} \quad (2.33)$$

die Wahrscheinlichkeitsverteilung von  $X$  unter der Bedingung  $B$  oder von  $X$  gegeben  $B$ .

Wegen  $\bigcup_x A_x = \Omega$  und  $A_x \cap A_y = \emptyset$ ,  $x \neq y$ , ist  $f(x|B)$  wohldefiniert, denn

$$\sum_x P(A_x|B) = \sum_x \frac{P(A_x \cap B)}{P(B)} = \frac{P(\Omega \cap B)}{P(B)} = 1.$$

**Def. 2.40 (Bedingter Erwartungswert bzgl.  $B$ )**

Falls  $\sum_x |x| f(x|B) < \infty$ , nennt man

$$E(X|B) := \sum_x x f(x|B)$$

den **bedingten Erwartungswert** von  $X$  bzgl.  $B$  oder den Erwartungswert von  $X$  gegeben  $B$ .

Es folgt ein für spätere Rechnungen äußerst nützlicher Satz:

**Satz 2.41 (Konditionieren des Erwartungswertes bzgl.  $B$ )**

$X$  sei eine (diskrete) ZV mit Erwartungswert  $E(X)$  und  $B$  ein Ereignis, so dass  $P(B)P(B^c) > 0$ , dann gilt

$$E(X) = E(X|B)P(B) + E(X|B^c)P(B^c). \quad (2.34)$$

Für vollständige Ereignismengen  $(B_i)_{i=1,2,\dots}$  der Form

1.  $\bigcup_i B_i = \Omega$

2.  $B_i \cap B_j = \emptyset, \quad i \neq j$

3.  $P(B_i) > 0, i = 1, 2, \dots$

gilt sogar allgemeiner

$$E(X) = \sum_i E(X|B_i)P(B_i). \quad (2.35)$$

Schließlich sei noch erwähnt, dass für  $A \subset B$  gilt

$$E(X | A \cap B) = E(X|A). \quad (2.36)$$

Diese Resultate legen ein Approximationsprinzip für  $X$  durch einfachere ZV (mit kleinerem Wertebereich) nahe, bei dem der Erwartungswert invariant bleibt.

**Def. 2.42 (Bedingte Erwartung bzgl.  $A, A^c$ )**

$X$  sei eine (diskrete) ZV mit Erwartungswert  $E(X)$ ,  $A$  ein Ereignis mit  $0 < P(A) < 1$  und  $\mathfrak{A} := \{\Omega, \emptyset, A, A^c\}$ . Dann nennt man die ZV  $Z : \Omega \rightarrow \mathbb{R}$  mit

$$Z(\omega) := \begin{cases} E(X|A) & , \text{ falls } \omega \in A \\ E(X|A^c) & , \text{ falls } \omega \in A^c \end{cases}$$

die **bedingte Erwartung** von  $X$  gegeben  $\mathfrak{A}$  bzw. von  $X$  unter der **Bedingung**  $\mathfrak{A}$ . Man schreibt auch  $Z = E(X|\mathfrak{A})$ . In diesem Fall nimmt  $Z$  genau zwei Werte mit den Wahrscheinlichkeiten  $P(A)$  und  $1 - P(A)$  an.

**Satz 2.43 (Invarianz des Erwartungswerts)**

Unter den Voraussetzungen von Def. 2.42 gilt mit Satz 2.41

$$E[Z] = E[E(X|\mathfrak{A})] = E(X|A)P(A) + E(X|A^c)P(A^c) = E(X).$$



Die bedingte Erwartung stellt also in der Regel eine Vergrößerung, und damit oft eine leichter handhabbare, vereinfachte Version der ursprünglichen ZV dar, ohne den Erwartungswert zu verändern. Zur Eingewöhnung der abstrakten Begriffe zunächst ein einfaches Beispiel.

### Beispiel 2.44 (Diskrete Gleichverteilung)

$X$  sei diskret gleichverteilt auf  $\{1, 2, \dots, 11, 12\}$  und  $A = \{9, 10, 11, 12\}$ . Damit ist  $P(A) = 1/3$ ,  $E(X) = 6.5$  und  $f(x) = 1/12$  für alle  $x$ . Man erhält weiter

$$f(x|A) = \begin{cases} 0 & , \text{ falls } 1 \leq x \leq 8 \\ \frac{1}{4} & , \text{ falls } 9 \leq x \leq 12. \end{cases}$$

Dies ergibt

$$E(X|A) = 10.5, \quad \text{und} \quad E(X|A^c) = 4.5$$

und die bedingte Erwartung  $Z$  bezüglich  $\mathfrak{A} := \{\Omega, \emptyset, A, A^c\}$

$$Z(\omega) = E(X|\mathfrak{A}) = \begin{cases} 10.5 & , \text{ falls } \omega \in A \\ 4.5 & , \text{ falls } \omega \in A^c. \end{cases}$$

Damit gilt

$$E[Z] = E[E(X|\mathfrak{A})] = E(X|A)P(A) + E(X|A^c)P(A^c) = 10.5 \cdot \frac{1}{3} + 4.5 \cdot \frac{2}{3} = \frac{13}{2} = E(X).$$

## 2.4.2 Bedingte Verteilungen

### Def. 2.45 (Diskrete bedingte Verteilung)

$f(x, y) = P(X = x, Y = y)$  sei die gemeinsame Wahrscheinlichkeitsfunktion von  $(X, Y)$ . Dann lautet die bedingte Verteilung von  $X$  **gegeben**  $Y = y$  oder von  $X$  **unter der Bedingung**  $Y$

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}, \quad \text{für alle } y \text{ mit } f_Y(y) > 0.$$

$f_Y$  ist die Randwahrscheinlichkeitsfunktion bzgl.  $Y$ .

**Def. 2.46 (Stetige bedingte Verteilung)**

$f(x, y)$  sei die gemeinsame Dichte der zweidimensionalen, stetigen ZV  $(X, Y)$ .  
Mit

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

wird die Randdichte von  $Y$  bezeichnet. Dann ist für alle  $y$  mit  $f_Y(y) > 0$

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

die Dichte von  $X$  unter der Bedingung  $Y$  bzw. von  $X$  gegeben  $Y = y$ .

**Satz 2.47 (Bedingte Verteilung)**

$f_{X|Y}$  sei eine Wahrscheinlichkeitsfunktion bzw. eine Dichte. Dann gilt

1.  $f_{X|Y} \geq 0$
  2.  $\sum_x f_{X|Y}(x) = 1$ , falls  $X$  diskret
- $$\int_{-\infty}^{\infty} f_{X|Y}(x) dx = 1$$
- , falls
- $X$
- stetig.

**Beispiel 2.48 (Bedingte Normalverteilung)**

Wir betrachten wie schon in Def. 2.10 und den Beispielen 2.13, 2.21 und 2.35 die zweidimensionale Normalverteilung von  $X$  und  $Y$  mit  $E(X) = \mu_x$ ,  $E(Y) = \mu_y$  sowie  $Var(X) = \sigma_x^2 > 0$  und  $Var(Y) = \sigma_y^2 > 0$ . Für den Korrelationskoeffizienten  $-1 < \rho < 1$  lautet dann die gemeinsame Dichte

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left[ -\frac{1}{2(1-\rho^2)} \left( \frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right) \right].$$

Nach Bsp. 2.13 (2) hat  $Y$  eine  $N(\mu_y, \sigma_y^2)$ -Dichte. Deshalb lautet die bedingte Dichte von  $X$  bzgl. gegebenem  $Y = y$

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{1}{\sigma_x\sqrt{2\pi(1-\rho^2)}} \exp \left( -\frac{1}{2(1-\rho^2)} \left( \frac{x-\mu_x}{\sigma_x} - \frac{\rho(y-\mu_y)}{\sigma_y} \right)^2 \right).$$

Damit ist die bedingte Dichte von  $X$  gegeben  $Y = y$  vom Typ einer eindimensionalen Normalverteilung  $N(\mu_x + \rho \frac{\sigma_x}{\sigma_y} (y - \mu_y), \sigma_x^2 (1 - \rho^2))$ .

Der Erwartungswert von  $X|Y = y$  bei bedingten Normalverteilungen hängt also ab von den Standardabweichungen  $\sigma_x$ ,  $\sigma_y$ , dem Korrelationskoeffizienten  $\rho$  und den Erwartungswerten  $\mu_x$  und  $\mu_y$ . Nach diesen expliziten Rechnungen suchen wir wieder einen abstrakten Zugang.

### 2.4.3 Bedingte Erwartung

#### Def. 2.49 (Bedingte Erwartung)

$f_{X|Y}$  sei eine bedingte Wahrscheinlichkeitsfunktion bzw. bedingte Dichte und  $F_{X|Y}$  die zugehörige Verteilungsfunktion. Man setzt

$$E(X|Y) = E(X|Y = y) = \int_{-\infty}^{\infty} x dF_{X|Y}(x) = \begin{cases} \sum_x x f_{X|Y}(x) & , \text{ diskret} \\ \int_{-\infty}^{\infty} x f_{X|Y}(x) dx & , \text{ stetig.} \end{cases}$$

Die **bedingte Erwartung**  $Z := E(X|Y)$  ist eine Zufallsvariable  $Z : \Omega \rightarrow \mathbb{R}$  mit dem Wertebereich  $W = \{E(X|Y = y) \mid f_Y(y) > 0\}$ . Im diskreten Fall kann also die bedingte Erwartung höchstens so viele Werte annehmen wie  $Y$  selbst (abgesehen von einer sog. Nullmenge  $N \subset \Omega$  mit  $P(N) = 0$ ).

#### Satz 2.50 (Bedingter Erwartungswert)

Falls beide Seiten existieren, gilt

$$E[E(X|Y)] = E(X).$$

Obige Sätze können zum Beispiel zur elementaren Analyse der einfachen Irrfahrt (random walk) herangezogen werden; vgl. [Stirzaker (1994)], pp. 145. Wir fahren zur Illustration mit einem abstrakten Beispiel fort.

#### Beispiel 2.51 (Der einfachste Fall)

$\Omega$  sei abzählbar,  $X : \Omega \rightarrow \mathbb{R}$  eine ZV mit Wahrscheinlichkeitsfunktion  $f(\omega) = P(\{\omega\})$  und existierendem Erwartungswert  $E(X) = \sum_{\omega} X(\omega) f(\omega)$ .

Seien weiter  $A \subset \Omega$  mit  $0 < p := P(A) < 1$ ,  $Y := 1_A$  die Indikatorfunktion von  $A$  sowie  $E_A(X) = \sum_{\omega \in A} X(\omega) f(\omega)$  und  $E_{A^c}(X) = \sum_{\omega \in A^c} X(\omega) f(\omega)$ . Wir definieren

$$Z(\omega) := E(X|Y)(\omega) = \begin{cases} \frac{1}{p} E_A(X) & , \text{ falls } \omega \in A \\ \frac{1}{1-p} E_{A^c}(X) & , \text{ falls } \omega \in A^c, \end{cases}$$

das für  $y = 1$  (falls  $\omega \in A$ ) bzw.  $y = 0$  (falls  $\omega \in A^c$ ) zwei verschiedene Werte mit Wahrscheinlichkeit  $p$  bzw.  $1 - p$  annimmt. Nun gilt

$$\begin{aligned} E(Z) = E(X|Y) &= p \cdot \frac{1}{p} E_A(X) + (1-p) \frac{1}{1-p} E_{A^c}(X) = E_A(X) + E_{A^c}(X) \\ &= \sum_{\omega \in A} X(\omega) f(\omega) + \sum_{\omega \in A^c} X(\omega) f(\omega) = \sum_{\omega} X(\omega) f(\omega) = E(X). \end{aligned}$$

Unter anderem verwendet man auch die Schreibweise

$$E(X|Y=1) = \frac{1}{p} E_A(X) \quad \text{bzw.} \quad E(X|Y=0) = \frac{1}{1-p} E_{A^c}(X).$$

### Satz 2.52 (Eigenschaften bedingter Erwartungen)

Seien  $a$  und  $b$  Konstanten,  $g(\cdot)$  eine beliebige Funktion und  $X, Y$  und  $Z$  gemeinsam verteilt. Dann gilt

1.  $E(a|Y) = a$
2.  $E(aX + bZ|Y) = aE(X|Y) + bE(Z|Y)$
3.  $E(X|Y) \geq 0$ , falls  $X \geq 0$
4.  $E(X|Y) = E(X)$ , falls  $X$  und  $Y$  unabhängig sind
5.  $E(Xg(Y)|Y) = g(Y)E(X|Y)$
6.  $E[X|Y; g(Y)] = E(X|Y)$
7.  $E[E(X|Y; Z)|Y] = E(X|Y)$ .

Bezeichnet etwa im diskreten Fall  $f_{Y,Z}(y, z) = \sum_x f(x, y, z)$  die Randwahrscheinlichkeitsfunktion, so definiert man

$$E(X|Y; Z) = \sum_x \frac{x f(x, y, z)}{f_{Y,Z}(y, z)}.$$

Die Eigenschaft 7. ermöglicht es, Bedingungen in beliebiger Ordnung hintereinanderschalten.

**Satz 2.53 (Projektionseigenschaft)**

Für jede (messbare) Funktion  $h(Y)$  mit  $E[h(Y)^2] < \infty$  gilt

$$E[(X - h(Y))^2] \geq E[(X - E(X|Y))^2].$$

Falls  $h(Y)$  eine Funktion von  $Y$  ist, so dass  $E(X - h(Y))^2 = E(X - E(X|Y))^2$ , dann gilt

$$E[(h(Y) - E(X|Y))^2] = 0.$$

Weitere Details entnehme man z. B. [Stirzaker (1994)], pp. 144.

**Lineare Prognose**

Aus der Projektionseigenschaft von Satz 2.53 leitet man ab, dass die bedingte Erwartung  $E(X|Y)$  bei Kenntnis von  $Y$  die (unbekannte) ZV  $X$  mindestens ebenso gut approximiert, wie die lineare Bestapproximation

$$h_{lin}(Y) := a^* + b^* Y, \quad (2.37)$$

wobei

$$E[(X - a^* - b^* Y)^2] \leq E[(X - a - b Y)^2] \quad \text{für alle } a, b \in \mathbb{R},$$

und

$$b^* = \frac{Cov(X, Y)}{Var(Y)}, \quad a^* = E(X) - b E(Y). \quad (2.38)$$

Wie erwähnt, liefert die bedingte Erwartung stets eine optimale Approximation im Sinne von Satz 2.53. Im Falle einer zweidimensionalen Normalverteilung von  $(X, Y)$  stimmt die bedingte Erwartung mit der linearen Bestapproximation von (2.37) überein; vgl. Beispiel 2.48.

Die Resultate von (2.38) lassen sich elementar, etwa mit den Techniken zum Beweis von  $|\rho(X, Y)| \leq 1$ , herleiten. Interessant ist der formale Zusammenhang zur empirischen Regression von Abschnitt 7.9. Dort sind lediglich die Rollen von  $X$  und  $Y$  zu vertauschen und die hier vorkommenden theoretischen Größen durch die entsprechenden empirischen Schätzungen zu ersetzen.

## 2.5 Ordnungsstatistiken

In diesem Abschnitt seien  $X, X_1, \dots, X_n$  iid und stetig mit Dichte  $f$  und Verteilungsfunktion  $F$ . Die Notation ist an [Stirzaker (1994)], pp. 288 angelehnt.

Die Ordnung der  $X_1, \dots, X_n$  in aufsteigender Reihenfolge ist eine weitere interessante Anwendung linearer Transformationen von ZV, die bereits in Satz 2.32 und Bsp. 2.34 einmal behandelt wurden. Seien also

$$\begin{aligned} Y_1 &= \min \{ X_1, \dots, X_n \} \\ Y_2 &= \text{zweitkleinstes Element der } X_1, \dots, X_n \\ &\dots \quad \dots \\ Y_n &= \max \{ X_1, \dots, X_n \}. \end{aligned} \quad (2.39)$$

Weiterhin verwenden wir die Bezeichnung  $Y_k = X_{(k)}$ , daneben ist auch  $Y_k = X_{k:n}$  gebräuchlich.

### Def. 2.54 (Ordnungsstatistiken)

$X_1, \dots, X_n$  seien identisch verteilt. Man nennt dann

$$Y_1, \dots, Y_n = X_{(1)}, \dots, X_{(n)} = X_{1:n}, \dots, X_{n:n}$$

mit den Eigenschaften (2.39) die **Ordnungsstatistiken** von  $X_1, \dots, X_n$ .

Die  $Y_1, \dots, Y_n$  sind wohldefiniert, da bei stetigen ZV gleiche Werte  $X_k(\omega) = X_j(\omega)$ ,  $k \neq j$  nur mit Wahrscheinlichkeit 0 auftreten.

Die Transformation (2.39) ist linear, aber nicht eineindeutig. Um das einzusehen, nehmen wir  $y_1 < y_2 < \dots < y_n$  an. Die Ergebnisse des Zufallsexperiments

$$X_1 = y_1, X_2 = y_2, \dots, X_n = y_n \quad \text{und} \quad X_2 = y_1, X_1 = y_2, \dots, X_n = y_n$$

ergeben die gleiche statistische Ordnung, nämlich

$$X_{(1)} = y_1, X_{(2)} = y_2, \dots, X_{(n)} = y_n.$$

Falls  $(\pi(1), \dots, \pi(n))$  eine der  $n!$  verschiedenen Permutationen der ersten  $n$  natürlichen Zahlen und  $R_\pi$  das Gebiet  $x_{\pi(1)} < \dots < x_{\pi(n)}$  ist, dann ist die Transformation

$$x_{(k)} = x_{\pi(k)}, \quad 1 \leq k \leq n$$

eindeutig und linear. In diesem Fall gilt mit  $\mathbf{A} = (a_{ij})$  und

$$a_{ij} = \begin{cases} 1 & , \text{ falls } i = \pi(j) \\ 0 & , \text{ sonst,} \end{cases}$$

$|\det(\mathbf{A})| = 1$ . Man sieht, dass  $(X_1, \dots, X_n)$  genau in einem der  $n!$  Gebiete von  $R_\pi$  liegt.

Damit haben die Ordnungsstatistiken für iid  $X_1, \dots, X_n$  mit Dichte  $f$  gemäß Satz 2.32 eine gemeinsame Dichte  $g$  der Form

$$g(y_1, \dots, y_n) = \begin{cases} n! \prod_{i=1}^n f(y_i) & , \text{ falls } y_1 < y_2 < \dots < y_n \\ 0 & , \text{ sonst.} \end{cases} \quad (2.40)$$

### Beispiel 2.55 (Randverteilungen, Gleichverteilung)

#### 1. Randverteilung der $X_{(k)}$

Die Dichten  $g_{(k)}$  der Randverteilungen von  $X_{(k)}$  lauten folgendermaßen

$$g_{(k)}(y) = k \binom{n}{k} f(y) [1 - F(y)]^{n-k} [F(y)]^{k-1} , \quad (2.41)$$

wobei  $f$  bzw.  $F$  Dichte bzw. Verteilungsfunktion der iid  $X_i$  bezeichnen. Das Ereignis  $X_{(k)} \leq y$  ereignet sich genau dann, wenn mindestens  $k$  der  $X_j$  in  $(-\infty, y]$  liegen. Die Wahrscheinlichkeit, dass dieses Ereignis für genau eines der  $X_j$  auftritt, ergibt sich über die Binomialverteilung mit  $p := P(X \leq y) = F(y)$  zu  $\binom{n}{j} F(y)^j [1 - F(y)]^{n-j}$ . Dann gilt für mindestens  $k$  solche Ereignisse

$$G_{(k)}(y) = P(X_{(k)} \leq y) = \sum_{j=k}^n \binom{n}{j} F(y)^j [1 - F(y)]^{n-j} .$$

Differenzieren liefert die Dichte

$$\begin{aligned} g_{(k)}(y) &= f(y) \sum_{j=k}^n j \binom{n}{j} F^{j-1} [1 - F]^{n-j} - (j+1) \binom{n}{j+1} F^j [1 - F]^{n-j-1} \\ &= f(y) k \binom{n}{k} F(y)^{k-1} [1 - F(y)]^{n-k} , \end{aligned}$$

wobei sich die einzelnen Summenglieder der Reihe nach wegheben,  $\binom{n}{n+1} = 0$  und  $F^k = [F(y)]^k$ .

#### 2. Gleichverteilung

Gemäß Korollar 1.56 ist bei stetigem  $X$  mit Verteilungsfunktion  $F$  die Zufallsvariable  $U := F(X)$  gleichverteilt in  $(0, 1)$ . Die Transformation  $U = F(X)$  ermöglicht in manchen Bereichen einen einheitlichen Zugang, indem zunächst die Aussagen für gleichverteilte ZV hergeleitet werden und danach eine Rücktransformation für die ursprüngliche ZV erfolgt. Diese Vorgehensweise bietet sich insbesondere bei Ordnungsstatistiken an.

$(X_1, \dots, X_n)$  seien unabhängig und gleichverteilt in  $(0, a)$ ,  $a > 0$ . Dann lautet die Dichte der Ordnungsstatistiken

$$g(y_1, \dots, y_n) = \frac{n!}{a^n} \quad \text{für } y_1 < y_2 < \dots < y_n . \quad (2.42)$$

O.B.d.A. setzen wir jetzt  $a = 1$  (Übergang von  $X$  zu  $X/a$ ). Für  $(0, 1)$ -gleichverteiltes  $X$  lauten Verteilungsfunktion  $F(x) = x$  für  $x \in (0, 1)$  und Dichte  $f(x) = 1_{(0,1)}(x)$ . Gemäß (2.41) erhält man dann für die Ordnungsstatistiken die Randdichten

$$g_{(k)}(y) = k \binom{n}{k} y^{k-1} (1-y)^{n-k} \quad \text{für } y \in (0, 1).$$

Zur Berechnung der Erwartungswerte benötigen wir noch

**Lemma 2.56**

Für  $k = 0, 1, \dots, n$  gilt

$$I_k := \int_0^1 \binom{n}{k} y^k (1-y)^{n-k} dy = \frac{1}{n+1}.$$

Beweis z. B. mit Induktion.

Nun kann der Erwartungswert der  $k$ -ten Ordnungsstatistik angegeben werden

$$\begin{aligned} E[X_{(k)}] &= \int_0^1 y k \binom{n}{k} (1-y)^{n-k} y^{k-1} dy \\ &= \int_0^1 k \binom{n}{k} (1-y)^{n-k} y^k dy = k I_k = \frac{k}{n+1}. \end{aligned}$$

Für  $(0, a)$ -gleichverteilte ZV erhält man entsprechend

$$E[X_{(k)}] = \frac{k a}{n+1}. \quad (2.43)$$



# Kapitel 3

## Erzeugende Funktionen

Wir beschränken uns in diesem Abschnitt auf die praktisch besonders wichtigen Fälle stetiger Zufallsvariablen (ZV) und diskreter mit dem Wertebereich  $\mathbb{N}_0 := \{0, 1, 2, \dots\}$ . Aus technischen Gründen werden nicht angenommene Werte mit Wahrscheinlichkeit 0 belegt; also bei  $B(n, p)$ -verteiltem  $Y$  setzt man  $P(Y = m) = 0$  für  $m = n + 1, n + 2, \dots$ . Generell seien die Komponenten  $X_k$  des Zufallsvektors  $\mathbf{X} = (X_1, \dots, X_n)^T$  unabhängig.

Die folgenden Transformationen sind besonders hilfreich bei

1. der Festlegung von Wahrscheinlichkeitsfunktionen für diskrete ZV mit nichtnegativen ganzzahligen Werten
2. der einfachen Berechnung von Momenten (differenzieren anstatt summieren und integrieren); Momente spielen in vielen stochastischen Modellen eine Rolle und liefern einfache Schätzmethoden für unbekannte Parameter
3. der Berechnung von Grenzverteilungen
4. der Lösung von Differenzen- und Differentialgleichungen im Zusammenhang mit stochastischen Prozessen; vgl. u.a. [Resnick (1994)]

### 3.1 Nichtnegative, ganzzahlige Zufallsvariable

Zunächst behandeln wir eine elegante Methode zur kompakten Beschreibung von diskreten Wahrscheinlichkeitsgesetzen. Aus der Theorie der analytischen Funktionen ist der eineindeutige Zusammenhang zwischen der Funktion und den Koeffizienten der entsprechenden Potenzreihe bekannt. Dies nützen wir hier aus. Sei zunächst  $(a_i) = (a_i)_{i \in \mathbb{N}_0}$  eine beliebige reelle Zahlenfolge. Später werden die  $a_i$  durch Wahrscheinlichkeiten  $p_i$  ersetzt. Falls nichts anderes gesagt wird, sei stets  $i \in \mathbb{N}_0$ .

**Def. 3.1 (Erzeugende Funktion von  $(a_i)$ )**

$$g_a(s) := \sum_{i=0}^{\infty} a_i s^i$$

heißt **erzeugende Funktion** (*generating function*) der Folge  $(a_i)$ , falls es ein  $s_0 > 0$  gibt, so dass  $g_a(|s|) < \infty$  für alle  $|s| < s_0$ .

Bei diskreten Wahrscheinlichkeitsverteilungen ist der Konvergenzradius wegen  $\sum p_i = 1$  mindestens gleich 1.

**Satz 3.2 (Eindeutigkeit)**

Falls  $s_0, s_1$  existieren, so dass

$$-\infty < g_a(s) = g_b(s) < \infty \quad \text{für alle } s_0 < s < s_1,$$

so folgt aus dem Identitätssatz für Potenzreihen

$$a_i = b_i \quad \text{für alle } i \in \mathbb{N}_0.$$

Erzeugende Funktionen sind ein Standardhilfsmittel zum Lösen linearer Differenzgleichungen; siehe u.a. [Stirzaker (1994)], pp. 58-60 oder [Resnick (1994)], pp. 7-17 and pp. 33.

Wir wenden uns jetzt dem Spezialfall zu, dass die Folge der  $a_i$  eine diskrete Wahrscheinlichkeitsverteilung charakterisiert.

**Def. 3.3 (Erzeugende Funktion; generating function)**

Die diskrete Zufallsvariable  $Y$  nehme die Werte  $i = 0, 1, 2, \dots$  mit Wahrscheinlichkeiten  $p_i := P(Y = i) = f_Y(i) = f(i) \geq 0$  an. Dann heißt

$$G(s) = G_Y(s) := E(s^Y) = \sum_i p_i s^i = \sum_i f(i) s^i \quad (3.1)$$

**erzeugende Funktion** von  $Y$ .

**Satz 3.4 (Erzeugende Funktion)**

Unter den Voraussetzungen von Definition 3.3 gilt:

1. Der Konvergenzradius von (3.1) ist mindestens 1, da  $G_Y(1) = \sum_i p_i \leq 1$ .
2.  $G_Y$  bestimmt die Verteilung  $p_i$  eindeutig; d. h. falls  $G_Y = G_Z$ , so sind  $Y$  und  $Z$  identisch verteilt.
3.  $G_Y$  ist analytisch mit den  $m$ -ten Ableitungen

$$G_Y^{(m)}(s) = \sum_{i=m}^{\infty} \frac{i!}{(i-m)!} p_i s^{i-m}, \quad |s| < 1.$$

4. In  $s = 1$  gilt

$$G_Y^{(m)}(1) = \sum_{i=m}^{\infty} \frac{i!}{(i-m)!} p_i,$$

falls die rechte Seite konvergiert.

5. Falls auch  $\infty$  als Erwartungswert zugelassen wird, so gilt

$$G_Y^{(m)}(1) = E[Y(Y-1) \cdot \dots \cdot (Y-m+1)].$$

Bei endlichen faktoriellen Momenten lassen sich die gewöhnlichen Momente einfach berechnen, z. B. mit  $G = G_Y$

$$\begin{aligned} E(Y) &= G'(1) \\ \text{Var}(Y) &= G''(1) + G'(1) - [G'(1)]^2 \end{aligned}$$

usw.

**Beispiel 3.5 (Erzeugende Funktionen)**

1. **Binomialverteilung**  $B(n, p)$ ,  $q := 1 - p$ :

$$G(s) = (q + ps)^n \tag{3.2}$$

2. **Poisson-Verteilung**  $P(\lambda)$ :

$$G(s) = e^{\lambda(s-1)} \tag{3.3}$$

3. **Geometrische Verteilung**  $p_i = pq^i$ ,  $i = 0, 1, 2, \dots$ ;  $q := 1 - p$ :

$$G(s) = \frac{p}{1 - qs} \tag{3.4}$$

Bevor wir uns Summen von unabhängigen diskreten Zufallsvariablen zuwenden, zunächst eine formale Definition.

**Def. 3.6 (Faltung)**

$a_0, a_1, a_2, \dots$  und  $b_0, b_1, b_2, \dots$  seien reelle Folgen. Die durch

$$c_k = \sum_{i=0}^k a_i b_{k-i}, \quad k = 0, 1, 2, \dots \quad (3.5)$$

definierte Folge  $(c_k)$  heißt **Faltung** (convolution) der  $(a_i)$  und  $(b_j)$ .

Faltungen sind u.a. kommutativ, denn für  $k = 0, 1, 2, \dots$  gilt

$$c_k = \sum_{i=0}^k a_i b_{k-i} = \sum_{i+j=k} a_i b_j = \sum_{j=0}^k a_{k-j} b_j.$$

Faltungen sind abstrakte Hilfsmittel zur Darstellung der Wahrscheinlichkeitsverteilungen von Summen unabhängiger ZV.

$Z$  bzw.  $Y$ , jeweils mit Wertebereich  $\mathbb{N}_0$  und Wahrscheinlichkeitsfunktionen  $f_Z$  bzw.  $f_Y$  seien unabhängig. Gesucht ist die Wahrscheinlichkeitsfunktion  $f_W$  und später die erzeugende Funktion der Summe  $W := Z + Y$ . Es gilt

$$\begin{aligned} c_k := f_W(k) &= P(W = k) = P(Z + Y = k) \\ &= \sum_{i=0}^k P(Z = i) P(Y = k - i) \quad \text{vgl. (3.7)} \\ &= \sum_{i=0}^k f_Z(i) f_Y(k - i) = \sum_{i=0}^k a_i b_{k-i}, \end{aligned} \quad (3.6)$$

wobei  $a_i := f_Z(i)$  und  $b_j := f_Y(j)$ ,  $i, j \in \mathbb{N}_0$  und wieder der Satz von der totalen Wahrscheinlichkeit angewendet wurde, nämlich:

Unter der Bedingung  $Z = i$  gilt  $Z + Y = k$  genau dann, wenn  $Y = k - i$ ; also mit  $P(A) = P(A|B)$  für unabhängige  $A$  und  $B$

$$\begin{aligned} P(Z + Y = k) &= \sum_{i=0}^{\infty} P(Z = i) P(Z + Y = k | Z = i) \\ &= \sum_{i=0}^{\infty} P(Z = i) P(Y = k - i | Z = i) \\ &= \sum_{i=0}^k P(Z = i) P(Y = k - i); \end{aligned} \quad (3.7)$$

für  $i > k$  gilt  $k - i < 0$  und somit  $P(Y = k - i) = 0$ .

**Beispiel 3.7 (Summe von Poisson-Variablen)**

In Spezialfällen läßt sich die Faltung direkt bestimmen. Seien  $Z \sim Poi(\lambda)$  und  $Y \sim$

$Poi(\mu)$  unabhängig, dann ist  $W := Z + Y$  Poisson-verteilt mit Parameter  $\lambda + \mu$ , denn

$$\begin{aligned} f_W(k) = c_k &= \sum_{i=0}^k f_Z(i) f_Y(k-i) = \sum_{i=0}^k \frac{\lambda^i}{i!} e^{-\lambda} \frac{\mu^{k-i}}{(k-i)!} e^{-\mu} \\ &= e^{-\lambda-\mu} \frac{1}{k!} \sum_{i=0}^k \frac{k!}{i!(k-i)!} \lambda^i \mu^{k-i} \\ &= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{i=0}^k \binom{k}{i} \lambda^i \mu^{k-i} = e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^k}{k!}. \end{aligned}$$

Diese eher mühsame Herleitung läßt sich elegant umgehen. Dazu beachte man, dass für die Multiplikation von  $g_a$ ,  $g_b$ ,  $g_c$ , der erzeugenden Funktionen bzgl.  $(a_i)$ ,  $(b_j)$ ,  $(c_k)$  mit der üblichen Umordnung von Doppelsummen und der Faltung (3.5) gilt

$$\begin{aligned} g_a(s) \cdot g_b(s) &= \left( \sum_{i=0}^{\infty} a_i s^i \right) \left( \sum_{j=0}^{\infty} b_j s^j \right) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_i s^i b_j s^j \\ &= \sum_{k=0}^{\infty} \sum_{i=0}^k a_i b_{k-i} s^k = \sum_{k=0}^{\infty} c_k s^k \\ &= g_c(s). \end{aligned}$$

Dies ergibt den wichtigen

**Satz 3.8 (Summen unabhängiger ZV)**

$(Z, Y)$  bzw.  $(X_1, \dots, X_n)$  seien unabhängige diskrete ZV mit Wertebereich  $\mathbb{N}_0$ . Dann gilt für die erzeugenden Funktionen der Summen

$$\begin{aligned} G_{Z+Y}(s) &= G_Z(s) G_Y(s) \\ G_{X_1+\dots+X_n}(s) &= G_{X_1}(s) \cdot \dots \cdot G_{X_n}(s). \end{aligned}$$

Beweis:

In den vorangegangenen Überlegungen wurde die Aussage des Satzes konstruktiv hergeleitet. Formal kann der Beweis in einer Zeile geführt werden. Da  $Z$  und  $Y$  unabhängig sind, gilt dies auch für  $t^Z$  und  $t^Y$ ; also

$$G_{Z+Y}(s) = E[s^{Z+Y}] = E[s^Z s^Y] = E[s^Z] E[s^Y] = G_Z(s) G_Y(s),$$

wobei die Multiplikationsregel für Erwartungswerte bei unabhängigen ZV benutzt wurde. Der Beweis im  $n$ -dim. Fall geht völlig analog.

**Beispiel 3.9**

**1. Binomialverteilung**

Seien  $Z$  bzw.  $Y$  unabhängig  $B(n, p)$ - bzw.  $B(m, p)$ -verteilt, dann ist die Summe  $Z + Y$  wie  $B(n + m, p)$ -verteilt, denn mit  $n, m \in \mathbb{N}$ ,  $0 < p < 1$  und  $q := 1 - p$  gilt:

$$G_{Z+Y}(s) = G_Z(s) \cdot G_Y(s) = (q + ps)^n \cdot (q + ps)^m = (q + ps)^{n+m}$$

## 2. Poisson-Verteilung

Für unabhängig  $Poi(\lambda_Z)$ - bzw.  $Poi(\lambda_Y)$ -verteilte ZV  $Z$  bzw.  $Y$  gilt

$$G_{Z+Y}(s) = G_Z(s) \cdot G_Y(s) = e^{\lambda_Z(s-1)} \cdot e^{\lambda_Y(s-1)} = e^{(\lambda_Z+\lambda_Y)(s-1)}.$$

Am Beispiel der Poisson-Verteilung wird die Arbeitserleichterung bei Verwendung der erzeugenden Funktion im Vergleich zur direkten Berechnung der Faltung deutlich.

## 3.2 Momenterzeugende Funktion

Nun wird die Beschränkung auf ganzzahlige ZV fallengelassen. Ein allgemeiner Zugang zur *Charakterisierung* der Verteilung einer ZV  $Y$  führt über die **charakteristische Funktion**  $\phi_Y(s) := E[e^{isY}]$ ,  $i^2 = -1$ . Diese *Fouriertransformierte* existiert für beliebige ZV und wird in weiterführenden Vorlesungen näher untersucht.

Hier behandeln wir eine einfachere, ähnliche Transformation, die für die meisten wichtigen ZV definiert ist (leider nicht für alle!); vgl. auch [Stirzaker (1994)], pp. 239.

### Def. 3.10 (Momenterzeugende Funktion)

Falls für die ZV  $Y$  mit VF  $F_Y$  gilt:  $E(e^{sY}) < \infty$ , für  $|s| < \epsilon$ ,  $\epsilon > 0$ , so heißt

$$M(s) = M_Y(s) := E(e^{sY}) = \int_{-\infty}^{\infty} e^{sy} dF_Y(y)$$

**momenterzeugende Funktion** von  $Y$ .

Ein endlicher Wert  $M_Y(s)$  für  $s = t$  und für  $s = -t$ ,  $t \neq 0$ , sichert bereits die Existenz aller absoluten Momente  $E(|Y|^k)$ ,  $k \in \mathbb{N}$ . Denn das schnell wachsende  $e^{sy}$  dominiert für ausreichend großes  $y$  ( $s > 0$ ) bzw. für genügend kleines negatives  $y$  ( $s < 0$ ) jede Potenzfunktion  $|y|^k$ ,  $k \in \mathbb{N}$ . Umgekehrt kann man aus der Existenz aller Momente noch nicht auf die Existenz der momenterzeugenden Funktion schließen (vgl. Lognormalverteilung).

Die Wahl des Namens für  $M_Y$  wird durch folgenden Satz klar.

**Satz 3.11 (Eindeutigkeit und Momente)**

Falls für die ZV  $Y$  die momenterzeugende Funktion  $M$  in einer Umgebung der 0 existiert, so gilt:

1. Durch  $k$ -maliges Differenzieren erhält man die Momente von  $Y$

$$\mu_k = E[Y^k] = M^{(k)}(0) = \frac{\partial^k M(0)}{\partial s^k} \quad k = 0, 1, 2, \dots$$

2.  $M_Y$  bestimmt die Verteilung von  $Y$  eindeutig und es gilt

$$M_Y(s) = \sum_{k=0}^{\infty} \mu_k \frac{s^k}{k!}.$$

**Satz 3.12 (Summen unabhängiger ZV)**

$(Z, Y)$  bzw.  $(X_1, \dots, X_n)$  seien unabhängige ZV mit existierenden momenterzeugenden Funktionen. Dann existieren diese Transformierten auch für die Summen und es gilt

$$\begin{aligned} M_{Z+Y}(s) &= M_Z(s) M_Y(s) \\ M_{X_1+\dots+X_n}(s) &= M_{X_1}(s) \cdot \dots \cdot M_{X_n}(s). \end{aligned}$$

Der Beweis ist völlig analog zu dem für erzeugende Funktionen.

**Beispiel 3.13 (Momenterzeugende Funktion)**

1. **Poisson-Verteilung:**  $Y \sim Poi(\lambda)$

$$M_Y(s) = \exp[\lambda(e^s - 1)] \quad (3.8)$$

2. **Normalverteilung:**  $Z \sim N(\mu, \sigma^2)$

$$M_Z(s) = \exp\left(\mu s + \frac{1}{2} \sigma^2 s^2\right) \quad (3.9)$$

3. **Gleichverteilung:**  $W \sim U(0, a), a > 0$

$$M_W(s) = \frac{e^{as} - 1}{as} \quad (3.10)$$

4. **Gammaverteilung:**  $Y \sim \Gamma(q, \lambda), q > 0, \lambda > 0$

$$M_Y(s) = \left(\frac{\lambda}{\lambda - s}\right)^q, \quad (3.11)$$

woraus sich für den Spezialfall der Exponentialverteilung  $V \sim ED(\lambda) = \Gamma(1, \lambda)$

$$M_V(s) = \frac{\lambda}{\lambda - s}$$

ergibt.

Die Herleitungen der genannten Formeln findet man in [Stirzaker (1994)], p. 189 sowie pp. 239-240.

### Beispiel 3.14 (Summen unabhängiger ZV)

#### 1. Normalverteilung

Seien  $Z$  bzw.  $Y$  unabhängig  $N(\mu_Z, \sigma_Z^2)$ - bzw.  $N(\mu_Y, \sigma_Y^2)$ -verteilt, dann ist die Summe wegen

$$\begin{aligned} M_{Z+Y}(s) &= M_Z(s) \cdot M_Y(s) \\ &= \exp\left(\mu_Z s + \frac{1}{2} \sigma_Z^2 s^2\right) \cdot \exp\left(\mu_Y s + \frac{1}{2} \sigma_Y^2 s^2\right) \\ &= \exp\left((\mu_Z + \mu_Y) s + \frac{1}{2} (\sigma_Z^2 + \sigma_Y^2) s^2\right) \end{aligned}$$

$N(\mu_Z + \mu_Y, \sigma_Z^2 + \sigma_Y^2)$ -verteilt.

#### 2. Gammaverteilung

Die Summe  $Z + Y$  unabhängiger  $\Gamma(r, \lambda)$ - bzw.  $\Gamma(t, \lambda)$ -verteilter ZV  $Z$  bzw.  $Y$  ist  $\Gamma(r + t, \lambda)$ -verteilt, wobei  $r, t, \lambda > 0$ , denn

$$M_{Z+Y}(s) = M_Z(s) \cdot M_Y(s) = \left(\frac{\lambda}{\lambda - s}\right)^r \cdot \left(\frac{\lambda}{\lambda - s}\right)^t = \left(\frac{\lambda}{\lambda - s}\right)^{r+t}.$$

#### 3. Erlangverteilung *Erlang*( $n, \lambda$ )

Ein wichtiger Spezialfall der Gammaverteilung ergibt sich für die Summe von iid-exponentialverteilten ZV  $X_1, \dots, X_n$

$$V = \sum_{i=1}^n X_i.$$

Man nennt  $V$  Erlang-verteilt. Die zugehörige momenterzeugende Funktion lautet

$$M_V(s) = \left(\frac{\lambda}{\lambda - s}\right)^n.$$

Mit (1.30) hat  $V$  also eine Dichte der Form

$$f_V(v) = \frac{\lambda^n}{(n-1)!} v^{n-1} e^{-\lambda v} \mathbf{1}_{(0, \infty)}(v). \quad (3.12)$$



Am Beispiel der Normalverteilung sieht man wieder die Vorteile des Zugangs über momenterzeugende Funktionen im Vergleich zur direkten Berechnung der Faltung gemäß Bsp. 2.35.

Alle Formeln (3.8) bis (3.11) haben exponentielle Form und vereinfachen sich durch Logarithmieren. Diese neue *erzeugende Funktion* definiert weitere charakteristische Größen der zugehörigen Verteilung; vgl. z. B. [Stirzaker (1994)], pp. 189.

**Def. 3.15 (Kumulantenfunktion)**

Falls die momenterzeugende Funktion der ZV  $Y$  existiert, so nennt man

$$K(s) = K_Y(s) := \ln [E(e^{sY})]$$

**Kumulantenfunktion** oder *Kumulanten erzeugende Funktion (generating function of the cumulants)* von  $Y$ , wenn eine Entwicklung der Form

$$K(s) = \sum_{k=0}^{\infty} \kappa_k \frac{s^k}{k!}$$

in einer Umgebung der 0 existiert.

Der Name nimmt bereits folgende Aussage vorweg.

**Satz 3.16 (Kumulanten)**

Falls die Kumulantenfunktion  $K$  zur ZV  $Y$  existiert, so erhält man durch  $k$ -maliges Differenzieren die Kumulanten von  $Y$ :

$$\kappa_k = K^{(k)}(0) \quad k = 0, 1, 2, \dots$$

Zum Beweis beachtet man die Beziehung:

$$M(s) = e^{K(s)},$$

wobei  $M$  die momenterzeugende Funktion bezeichnet. Differenzieren liefert

$$M'(s) = e^{K(s)} K'(s) = M(s) K'(s) = M K',$$

$$M'' = M K'' + M' K',$$

usf., woraus sich mit  $M(0) = 1$ ,  $M'(0) = EY$  und  $M''(0) = EY^2$  die beiden ersten Kumulanten ergeben. Die höheren Kumulanten folgen gemäß Def. 1.74 analog. Zur Erinnerung:

$$\kappa_1 = \mu = E(Y), \quad \kappa_2 = \sigma^2 = \text{Var}(Y), \quad \kappa_3 = E(Y - \mu)^3, \quad \kappa_4 = E(Y - \mu)^4 - 3\sigma^4, \dots$$

**Beispiel 3.17 (Kumulantenfunktion)**

1. Poisson-Verteilung:
- $Y \sim Poi(\lambda)$

$$K_Y(s) = \lambda(e^s - 1), \quad \kappa_k = \lambda, \quad k = 1, 2, 3, \dots$$

2. Normalverteilung:
- $Z \sim N(\mu, \sigma^2)$

$$K_Z(s) = \mu s + \frac{1}{2} \sigma^2 s^2$$

$$\kappa_1 = \mu, \quad \kappa_2 = \sigma^2, \quad \kappa_k = 0, \quad k = 3, 4, \dots$$

Die Normalverteilung ist übrigens die einzige Verteilung mit der Eigenschaft, dass alle Kumulanten ab der Ordnung 3 verschwinden.

Als Vorbereitung auf den Zentralen Grenzwertsatz formulieren wir am Ende dieses Abschnitts noch zwei wichtige Aussagen für momenterzeugende Funktionen.

**Satz 3.18 (Lineare Transformation)**

Falls  $Y$  die momenterzeugende Funktion  $M_Y$  besitzt, so existiert diese Transformation auch für  $Z := a + bY$ ,  $a, b \in \mathbb{R}$ , und lautet

$$M_{a+bY} = e^{as} M_Y(bs).$$

Beweis:

$$M_{a+bY} = E[e^{s(a+bY)}] = E[e^{sa} e^{sbY}] = E[e^{sa}] E[e^{(sb)Y}] = e^{as} M_Y(bs).$$

Weiterhin gilt ein wichtiges Stetigkeitsresultat für Folgen von Verteilungsfunktionen; siehe [Stirzaker (1994)], p. 241.

**Satz 3.19 (Momenterzeugende Funktionen für Folgen  $(F_n)$ )**

$(Y_n)$  sei eine Folge von ZV mit Verteilungsfunktionen  $(F_n)$  und momenterzeugenden Funktionen  $(M_n)$ , die für  $|s| < b$ ,  $b > 0$  existieren. Falls für ein  $0 < a < b$  und für alle  $|t| \leq a$  gilt

$$\lim_{n \rightarrow \infty} M_n(t) = M(t)$$

und  $M$  momenterzeugende Funktion einer ZV  $Y$  mit Verteilungsfunktion  $F$  ist, so folgt

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

in allen Stetigkeitspunkten  $x$  von  $F$ .

# Kapitel 4

## Gesetze der großen Zahlen

Durch Probieren läßt sich feststellen, dass beim Wurf einer symmetrischen Münze die relativen Häufigkeiten für "Kopf" gegen  $\frac{1}{2}$  konvergieren, wenn man das Experiment nur oft genug wiederholt. Die **Gesetze der großen Zahlen**, von denen die einfachsten hier hergeleitet werden, sichern unter sehr schwachen Bedingungen die Konvergenz des Stichprobenmittels gegen den Erwartungswert. Das Verhalten der relativen Häufigkeiten beim wiederholten Münzwurf kann man als Spezialfall des sog. starken Gesetzes der großen Zahlen deuten.

Es sei daran erinnert, dass die moderne Wahrscheinlichkeitstheorie auf den Axiomen von Kolmogorov basiert; vgl. Def. 1.10. Die Gesetze der großen Zahlen, und deren Übereinstimmung mit praktischen Experimenten, rechtfertigen den axiomatischen Zugang im nachhinein. In früherer Zeit hatte man versucht, umgekehrt Wahrscheinlichkeiten als Grenzwerte von relativen Häufigkeiten zu definieren. Dies führte nicht zum gewünschten Ziel einer widerspruchsfreien Theorie. Die axiomatische Fundierung hat sich als überlegen erwiesen.

### 4.1 Ungleichungen

Da die exakte Berechnung von Wahrscheinlichkeiten nicht immer möglich ist, greift man oft auf Ungleichungen zurück. Hier sollen einige davon behandelt werden. Für eine weitere Diskussion, z. B. der Ungleichung von Jensen, siehe z. B. [Stirzaker (1994)], pp. 98.

**Satz 4.1 (Basisungleichung)**

Falls  $h(x)$  eine nicht-negative Funktion ist und  $E[h(X)] < \infty$ , dann gilt für alle  $a > 0$

$$P(h(X) \geq a) \leq E[h(X)]/a. \quad (4.1)$$

Beweis:

Sei  $A := \{\omega \mid h[X(\omega)] \geq a\}$  und  $1_A$  die zugehörige Indikatorfunktion mit  $E(1_A) = P(h(X) \geq a)$ . Nach dieser Konstruktion gilt  $h(X) - a 1_A \geq 0$  und  $E[h(X) - a 1_A] \geq 0$ , womit folgt:

$$E h(X) \geq a E(1_A) = a P(h(X) \geq a).$$

□

**Satz 4.2 (Ungleichung von Markov)**

Aus der Basisungleichung ergibt sich mit  $h = |\cdot|$  für beliebige Zufallsvariable (ZV) mit existierendem Erwartungswert und  $a > 0$

$$P(|X| \geq a) \leq E(|X|)/a. \quad (4.2)$$

Von noch größerer Bedeutung ist die

**Satz 4.3 (Ungleichung von Tschebyschow,  $E(X) = 0$ )**

Falls  $E(X^2)$  existiert und  $E(X) = 0$ , so gilt für beliebige  $t > 0$

$$P(|X| \geq t) \leq E(X^2)/t^2. \quad (4.3)$$

Beweis:

Wir verwenden hier die Riemann-Stieltjes-Notation. Zur Übung sollte dieser wichtige Beweis für diskrete bzw. stetige ZV wiederholt werden.

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 dF(x) = \int_{|x| < t} x^2 dF(x) + \int_{|x| \geq t} x^2 dF(x) \\ &\geq \int_{|x| \geq t} x^2 dF(x) \geq t^2 \int_{|x| \geq t} dF(x) = t^2 P(|X| \geq t). \end{aligned} \quad \square$$

Durch die Translation  $X - \mu \rightarrow \tilde{X}$  erhält man die bekanntere Version

**Satz 4.4 (Ungleichung von Tschebyschow)**

Für beliebige Zufallsvariable mit  $E(X) = \mu$ ,  $Var(X) = \sigma^2$  gilt

$$P(|X - \mu| < t\sigma) \geq 1 - \frac{1}{t^2}, \quad \forall t > 0, \quad (4.4)$$

oder äquivalent dazu

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}, \quad \forall \epsilon > 0. \quad (4.5)$$

Diese Ungleichung gilt für beliebige ZV mit endlicher Varianz  $\sigma^2$  und gibt eine allgemeine Schranke an, wie weit eine ZV um ihren Mittelwert streut. Diese Schranken können bei speziellen Annahmen, z. B. Normalverteilung, enger gefaßt werden:

	Tschebyschow	$N(\mu, \sigma^2)$
t	$P( X - \mu  < t\sigma) \geq 1 - 1/t^2$	$P( X - \mu  < t\sigma) = 2\Phi(t) - 1$
1	0	0.6826
2	0.7500	0.9546
3	0.8889	0.9974
4	0.9375	$1 - 6 \cdot 10^{-5}$
5	0.9600	$1 - 7 \cdot 10^{-7}$

Für den Namen Čebyšev des russischen Mathematikers sind viele Transliterationen gebräuchlich. Bei Schreibweisen wie *Tschebyscheff* sollte man daran denken, dass im russischen Original das letzte 'e' wie ein 'o' gesprochen wird.

## 4.2 Gesetze der großen Zahlen

Häufig wendet man die Ungleichung von Tschebyschow mit  $X = \bar{X}$  an, wobei  $\bar{X}$  das arithmetische Mittel von  $X_1, \dots, X_n$  ist.

### Satz 4.5 (Satz von Tschebyschow)

$X_1, \dots, X_n$  seien paarweise unkorreliert mit  $E(X_i) = \mu_i$  und beschränkten Varianzen  $\text{Var}(X_i) \leq \sigma^2$ ,  $i = 1, \dots, n$ ;  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  bezeichne das arithmetische Mittel. Dann gilt für beliebiges  $\epsilon > 0$  und  $0 < \eta < 1$

$$P\left(|\bar{X}_n - \frac{1}{n} \sum_{i=1}^n \mu_i| < \epsilon\right) \geq 1 - \eta, \quad \text{falls } n \geq \frac{\sigma^2}{\eta \epsilon^2}.$$

### Def. 4.6 (Stochastische Konvergenz)

Eine Folge  $Y_1, Y_2, \dots$  von ZV konvergiert **stochastisch** oder in **Wahrscheinlichkeit** gegen eine Zufallsvariable  $Y$  (i. Z.  $Y_n \xrightarrow{P} Y$ ) genau dann, wenn für alle  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(|Y_n - Y| < \epsilon) = 1.$$

### Satz 4.7 (Schwaches Gesetz der großen Zahlen)

- $X_1, X_2, \dots$  seien paarweise unkorreliert mit  $E(X_i) = \mu_i$  und beschränkten Varianzen  $\text{Var}(X_i) \leq \sigma^2$ , dann konvergieren die arithmetischen Mittel  $\bar{X}_n$  stochastisch gegen die arithmetischen Mittel ihrer Erwartungswerte, d. h.

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \frac{1}{n} \sum_{i=1}^n \mu_i.$$

- Die arithmetischen Mittel  $\bar{X}_n$  von iid-Variablen  $X_1, X_2, \dots$  mit existierenden zweiten Momenten konvergieren stochastisch gegen den Mittelwert  $\mu = E(X_i)$ ,  $i = 1, 2, \dots$ , d. h.

$$\bar{X}_n \xrightarrow{P} \mu \quad \text{oder}$$

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \epsilon) = 1 \quad \text{für alle } \epsilon > 0.$$

Wir haben bereits angesprochen, dass etwa beim Münzwurf die relativen Häufigkeiten "praktisch immer" gegen die gesuchten Wahrscheinlichkeiten konvergieren. Dies ist

das sog. **starke Gesetz der großen Zahlen**. Dabei können die Voraussetzungen des folgenden Satzes noch abgeschwächt werden.

**Satz 4.8 (Starkes Gesetz der großen Zahlen)**

Die arithmetischen Mittel  $\bar{X}_n$  von iid-Variablen  $X_1, X_2, \dots$  mit existierenden ersten Momenten konvergieren **fast sicher** (almost everywhere) oder **stark** oder **mit Wahrscheinlichkeit 1** gegen den Mittelwert  $\mu = E(X_i)$ ,  $i = 1, 2, \dots$ , d. h.

$$\bar{X}_n \xrightarrow{f.s.} \mu \quad \text{oder}$$

$$P\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) := P\left(\{\omega \mid \lim_{n \rightarrow \infty} \bar{X}_n(\omega) = \mu\}\right) = 1. \quad (4.6)$$

Ohne auf einen Beweis eingehen zu können, sei erwähnt, dass starke Konvergenz die stochastische impliziert, d. h.

$$Y_n \xrightarrow{f.s.} Y \quad \implies \quad Y_n \xrightarrow{P} Y.$$

Die Aussage von (4.6) hat also die Konvergenz von Satz 4.7 zur Folge und ist somit *stärker*. Warum Satz 4.8 mit geringeren Voraussetzungen auskommt, ist schnell gesagt: Das schwache Gesetz der großen Zahlen wurde in Satz 4.7 so formuliert, dass zum Beweis der Satz von Tschebyschow herangezogen werden kann. Die dortigen Voraussetzungen können aber noch gelockert werden.

Das eingangs beschriebene Phänomen der Konvergenz von relativen Häufigkeiten hat mit fast-sicherer Konvergenz zu tun. Es ist zwar ein Ereignis  $E_{fail}$  denkbar, so dass z. B. beim wiederholten (symmetrischen) Münzwurf eine Folge nicht gegen die Wahrscheinlichkeit  $\frac{1}{2}$  konvergiert. Für dieses Ereignis gilt aber  $P(E_{fail}) = 0$ .

## 4.3 Grenzwertsätze

Wir werden hier nur zwei der wichtigsten Grenzwertsätze erwähnen. Es gibt eine Vielzahl von Verallgemeinerungen, die der weiterführenden Literatur zu entnehmen sind.

### 4.3.1 Zentraler Grenzwertsatz

Nun wenden wir uns einem weiteren Konvergenzbegriff zu.

#### Def. 4.9 (Verteilungskonvergenz)

Eine Folge  $Y_1, Y_2, \dots$  von ZV konvergiert in Verteilung (in distribution) gegen eine Zufallsvariable  $Y$  (i. Z.  $Y_n \xrightarrow{d} Y$ ) genau dann, wenn die Verteilungsfunktionen  $F_n$  der  $Y_n$  gegen die Verteilungsfunktion  $F$  von  $Y$  konvergieren, und zwar in allen Stetigkeitspunkten von  $F$ .

Verteilungskonvergenz wird auch *schwache Konvergenz* genannt, denn sie ist die schwächste der drei in diesem Kapitel definierten Konvergenzarten. Es gilt:

$$Y_n \xrightarrow{f.s.} Y \quad \Longrightarrow \quad Y_n \xrightarrow{P} Y \quad \Longrightarrow \quad Y_n \xrightarrow{d} Y.$$

#### Satz 4.10 (Zentraler Grenzwertsatz für iid-Variable)

$X_1, X_2, \dots$  seien iid-ZV mit  $E(X_i) = \mu$ ,  $Var(X_i) = \sigma^2 < \infty$ ,  $i = 1, 2, \dots$ .

Dann gilt für  $S_n := \sum_{i=1}^n X_i$  und  $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i = S_n/n$  sowie

$$U_n := \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} :$$

$$\lim_{n \rightarrow \infty} U_n \sim N(0, 1), \quad d.h. \quad \lim_{n \rightarrow \infty} P(U_n \leq u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-t^2/2} dt,$$

oder  $U_n \xrightarrow{d} U$ , wobei  $U$  standardnormalverteilt ist.

#### Bemerkung 4.11 (zum Zentralen Grenzwertsatz, ZGWS)

1. Mit unseren Hilfsmitteln läßt sich der Beweis von Satz 4.10 gemäß [Stirzaker (1994)], pp. 294-295 führen. Besitzen zum Beispiel die iid-ZV  $X_i$  eine momenterzeugende Funktion  $M(s)$ , so sind etwa die behandelten Transformations- (vgl. 3.18) und Stetigkeitssätze (siehe 3.19) anzuwenden.
2. Summen  $S_n = \sum X_i$  von iid ZV mit  $E(X_i) = \mu$ ,  $Var(X_i) = \sigma^2$ ,  $i = 1, \dots, n$ , sind also für große  $n$  approximativ  $N(n\mu, n\sigma^2)$ -verteilt.



3. Der Grenzwertsatz von de Moivre-Laplace für binomialverteilte ZV ist ein Spezialfall des ZGWS 4.10. Der Beweis (in [Chung (1979)], pp. 210-216 findet man eine moderne Version) ist konstruktiv und beruht wesentlich auf der Stirlingschen Formel. Momenterzeugende Funktionen werden dort nicht benötigt.
4. Man sagt für  $n \geq 50$  ist  $U_n$  "praktisch" standardnormalverteilt. Für unabhängige  $X_i$  mit  $E(X_i) = \mu_i$  und  $Var(X_i) = \sigma_i^2 \leq \sigma^2$ ,  $i = 1, 2, \dots$  gilt die Aussage des Satzes für

$$U_n := \sum_{i=1}^n (X_i - \mu_i) / \sqrt{\sum_{i=1}^n \sigma_i^2}. \quad (4.7)$$

5. Der ZGWS kann unter weit allgemeineren Voraussetzungen bewiesen werden. Die bisherigen Formulierungen sollen aber für diese Einführungsveranstaltung genügen.
6. Wichtige Anwendungen sind etwa Approximationen der **Binomial-** und **Poisson-Verteilung**.

Eine binomialverteilte ZV  $X$  kann als Summe von  $n$  unabhängig mit dem Parameter  $p$  Bernoulli-verteilten ZV  $Y_1, \dots, Y_n$  angesehen werden (vgl. Bsp. 3.9, 1.). Für genügend großes  $n$ , etwa  $np \geq 4$  und  $n(1-p) \geq 4$ , gilt dann in guter Näherung

$$\begin{aligned} P(X = j) &= P(j - \frac{1}{2} < X \leq j + \frac{1}{2}) \\ &\approx \Phi\left(\frac{j + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{j - 0.5 - np}{\sqrt{np(1-p)}}\right), \end{aligned} \quad (4.8)$$

und  $\Phi$  bezeichnet wieder die Verteilungsfunktion der Standardnormalverteilung. Analog dazu kann für  $n \in \mathbb{N}$  und  $\lambda = n\mu$ ,  $\mu > 0$ , gemäß 2. von Bsp. 3.9 eine  $Poi(\lambda)$ -verteilte ZV  $X$  als unabhängige Summe von  $n$  mit Parameter  $\mu$  Poisson-verteilten  $Z_1, \dots, Z_n$  dargestellt werden. Für  $\lambda > 20$  (also  $n$  oder  $\mu$  genügend groß) gilt daher

$$\begin{aligned} P(X = j) &= P(j - \frac{1}{2} < X \leq j + \frac{1}{2}) \\ &\approx \Phi\left(\frac{j + 0.5 - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{j - 0.5 - \lambda}{\sqrt{\lambda}}\right). \end{aligned} \quad (4.9)$$

7. Der zentrale Grenzwertsatz gestattet es, in vielen Fällen die betrachteten Zufallsvariablen als normalverteilt anzusehen. Dabei geht man davon aus, dass die beobachtete Größe durch additive Überlagerung vieler nicht beobachteter Einflüsse entsteht.
8. Trotzdem ist eine Normalverteilungsannahme stets durch geeignete statistische Tests oder Methoden der explorativen Datenanalyse zu verifizieren; vgl. Kapitel 10. Dort spielt u.a. das Lemma 10.2 von Glivenko-Cantelli eine wichtige Rolle. Dieses besagt, dass die empirische Verteilungsfunktion  $F_n$  von iid ZV  $X_1, \dots, X_n$  für  $n \rightarrow \infty$  fast sicher gegen die Verteilungsfunktion  $F$  der  $X_i$ ,  $i = 1, \dots, n$ , konvergiert. Damit gehört auch diese Aussage zu den Gesetzen der großen Zahlen.

### 4.3.2 Approximation der Binomial- durch die Poisson-Verteilung

Die Approximation der Binomial- durch die Normalverteilung ist umso genauer, je näher  $p$  bei  $\frac{1}{2}$  liegt und je größer  $n$  ist. Für sehr kleine  $p$  und  $k$ , d.h.

$$p \ll \frac{1}{2} \quad \text{und} \quad k \ll n$$

ist die folgende Näherung weitaus besser geeignet. Sie spielt auch bei asymptotischen theoretischen Betrachtungen eine wichtige Rolle.

Man betrachtet zunächst für  $\lambda > 0$  die  $B(n; \frac{\lambda}{n})$ -verteilte ZV  $X_n$ ,  $n \in \mathbb{N}$ ,  $n > \lambda$ . Anders als bisher variiert jetzt der zweite Parameter  $p_n := \lambda/n$  mit  $n$ . Sei  $w_{k,n} := P(X_n = k)$ , dann gilt für  $k = 0, \dots, n$

$$\begin{aligned} w_{k,n} &= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n(n-1) \cdots (n-k+1)}{n^k \left(1 - \frac{\lambda}{n}\right)^k} \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \frac{1 \left(1 - \frac{\lambda}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)}{\left(1 - \frac{\lambda}{n}\right)^k}; \end{aligned} \tag{4.10}$$

und für festgehaltenes  $k$  wegen  $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$

$$\lim_{n \rightarrow \infty} w_{k,n} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Die Größen auf der rechten Seite entsprechen der Wahrscheinlichkeitsfunktion der Poisson-Verteilung; vgl. (1.17). Wir fassen zusammen

#### Satz 4.12 (Binomial- und Poisson-Verteilung)

Für  $\lambda > 0$ ,  $n \in \mathbb{N}$ ,  $p_n := \lambda/n$  und  $n \rightarrow \infty$  geht die Binomialverteilung  $B(n, p_n)$  mit der Wahrscheinlichkeitsfunktion

$$b_{x,n,p_n} := f_B(x) = \binom{n}{x} p_n^x (1 - p_n)^{n-x}, \quad x = 0, 1, \dots, n$$

über in die Poisson-Verteilung  $Poi(\lambda)$  mit der Wahrscheinlichkeitsfunktion

$$v_{x,\lambda} := f_P(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x \in \mathbb{N}_0.$$

Die Approximationsgüte wächst für großes  $n$ . Der Herleitung entnimmt man aber auch, dass die Approximation von Satz 4.12 mit  $\lambda = np$  ebenso für sehr kleines, festgehaltenes

$p$  und kleine  $k$  ausgezeichnet ist. Dies unterstreicht auch folgende Abschätzung, die wir ohne Beweis angeben. Für eine beliebige Teilmenge  $A \subset \mathbb{N}_0$  sowie jedes  $0 < p < 1$  und jedes  $n \in \mathbb{N}$  gilt die Abschätzung

$$\left| \sum_{k \in A} b_{k,n,p} - \sum_{k \in A} v_{k,np} \right| \leq p. \quad (4.11)$$

Die beiden Beispiele in den Tabellen von Anhang 11.4 geben einen Eindruck von der Approximationsgüte der Normal- bzw. Poissonverteilung.

Die numerische Approximation von Binomialwahrscheinlichkeiten durch den Poissonansatz hat im Computerzeitalter natürlich an Bedeutung verloren, da in den meisten Fällen eine exakte Berechnung der  $b_{k,n,p}$  durchgeführt werden kann, was früher mit Papier und Bleistift zumindest eine erhebliche Mühe bedeutet hätte.

Weitere interessante Eigenschaften der Poisson-Verteilung findet man zum Beispiel in [Chung (1979)], pp. 193, [Pfanzagl (1988)], S. 255-258 oder der Spezialliteratur über die Poissonverteilung.

# Kapitel 5

## Einfache Irrfahrt (Random Walk)

Bereits am Ende des Abschnitts 1 war im Beispiel vom Ruin des Spielers vom sog. *Random Walk* die Rede. Ausführliche Darstellungen hierzu findet man etwa in [Stirzaker (1994)], pp. 145 oder [Resnick (1994)], pp. 33. Dem Zweck dieser Materialien entsprechend, beschränken wir uns hier auf einführende Betrachtungen.

### 5.1 Definition, Rekurrenz

#### Def. 5.1 (Einfache Irrfahrt; Random Walk)

Seien  $X_1, X_2, \dots$  diskrete iid-Variablen, also gemäß Def. 2.29 unabhängig und identisch verteilt, mit

$$P(X_i = 1) = p, \quad P(X_i = -1) = q := 1 - p, \quad 0 < p < 1, \quad i = 1, 2, \dots,$$

und für gegebenes  $S_0 \in \mathbb{Z}$

$$S_n := S_0 + \sum_{i=1}^n X_i, \quad n = 1, 2, \dots \quad (5.1)$$

Dann nennt man  $(S_n; n \in \mathbb{N}_0)$  **einfache Irrfahrt** oder **simple Random Walk** (gelegentlich auch **Bernoulli Walk**).

Im Falle  $p = q = \frac{1}{2}$  spricht man von einem **symmetrischen Random Walk**.

$(S_n; n \in \mathbb{N}_0)$  ist ein **stochastischer Prozeß** mit **Parameterraum**  $\mathbb{N}_0$  und **Zustandsraum**  $\mathbb{Z}$ . Man schreibt auch  $(S_n; n \geq 0)$ .

Die Charakterisierung **einfach** bzw. **simple** bezieht sich u.a. darauf, dass der Zustandsraum der  $S_n$  eindimensional und ganzzahlig ist. Man kann ähnliche Modelle auch in allgemeineren Zustandsräumen, u.a. in höheren Dimensionen betrachten. Wir lassen ab jetzt die Zusätze *einfach* bzw. *simple* weg und sprechen in diesem Abschnitt nur noch vom *Random Walk*, meinen aber den Prozeß von Def. 5.1.

1. Gelegentlich wird der Parameterraum mit  $T$  bezeichnet. Dann interpretiert man  $S_t \in \mathbb{Z}$  als den Ort, an dem sich z. B. ein Partikel zum diskreten Zeitpunkt  $t \in T = \mathbb{N}_0$  gerade befindet.

2. Wegen der Unabhängigkeit der  $X_i$  gilt die sog. Markov-Eigenschaft:

$$P(S_{n+1} = j | S_n = i_n, S_{n-1} = i_{n-1}, \dots, S_1 = i_1, S_0 = i_0) = P(S_{n+1} = j | S_n = i_n),$$

d. h. die Kenntnis des Zustands  $S_n$  liefert genauso viel Information für die Vorhersage des Zustands  $S_{n+1}$  wie die Kenntnis der gesamten Vorgeschichte des Prozesses ( $S_j; j = 0, 1, \dots, n$ ).

3. Auf den Zustand  $S_n$  kann entweder der Zustand  $S_{n+1} = S_n + 1$  oder der Zustand  $S_{n+1} = S_n - 1$  folgen. Mit  $0 < p < 1$  gilt

$$P(S_{n+1} = j + 1 | S_n = j) = p \quad \text{und} \quad P(S_{n+1} = j - 1 | S_n = j) = q = 1 - p.$$

4. Der Ausgangszustand  $S_0 = k \in \mathbb{Z}$  ist vorgegeben. Bei vielen Überlegungen kann man sich mit der Transformation  $\tilde{S}_n := S_n - k$  auf den einfacheren Fall  $\tilde{S}_0 = 0$  beschränken und so den formalen Aufwand für Beweise verringern.

Wir wollen weitere Details der Vorlesung *Stochastische Prozesse* überlassen, bemerken aber an dieser Stelle:

1. Die  $S_n$ ,  $n = 1, 2, \dots$  sind Zufallsvariable über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{F}, P)$ , der sich als unendliches Produkt einfacherer Wahrscheinlichkeitsräume  $(\Omega_i, \mathfrak{F}_i, P_i)$ ,  $i \in \mathbb{N}_0$ , ergibt (vgl. iid-Variable  $X_i$ ).
2. Bisher wurde die Folge von ZV  $S_n : \Omega \rightarrow \mathbb{Z}$  für  $n \in \mathbb{N}_0$  betrachtet. Eine andere Sichtweise ist  $S_n : \Omega \times \mathbb{N}_0 \rightarrow \mathbb{Z}$ . Hält man bei diesem Ansatz ein  $\omega \in \Omega$  fest, so liefert  $(S_n(\omega); n \in \mathbb{N}_0)$  einen **Pfad** des stochastischen Prozesses  $S_n$ . Im Bild 10 ist  $(n, S_n(\omega))$  für einen solchen Pfad geplottet.

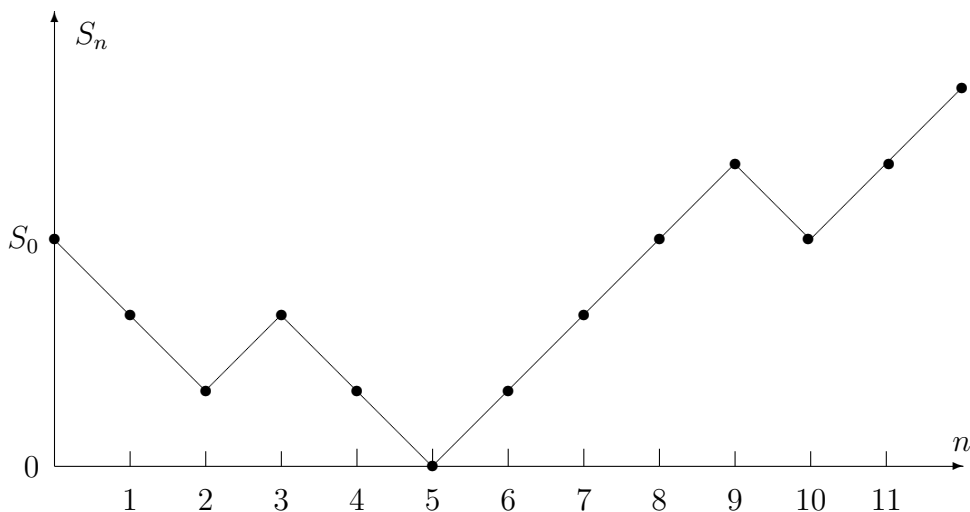


Bild 10: Random Walk

In fast jedem Lehrbuch findet man eine andere phantasievolle Illustration des Random Walk, z. B. der Betrunkene, der von einer Seite einer engen Gasse zur anderen torkelt etc.; siehe etwa [Chung (1979)], pp. 240. Wir lehnen uns bei Interpretationen in der Regel an das bereits eingeführte Beispiel 1.30 vom Ruin des Spielers an. Es gelten folgende Zusammenhänge:

1. Startkapital :  $k = S_0 \geq 0$
2.  $p = P(\text{Gewinn}) = P(S_{n+1} = j + 1 | S_n = j)$
3.  $S_n = 0 \iff$  Ruin des Spielers; in Bild 10 nach  $n = 5$  Spielen.
4.  $S_n = K$  und  $S_j \notin \{0, K\}$ ,  $j < n \iff$  Ende nach  $n$  Spielen mit gewonnenem Zielkapital  $K$ .

### Bemerkung 5.2 (Fragestellungen zum Random Walk)

In Abschnitt 1 wurden einige Fragen angeschnitten. Eine davon war:

Trifft ein Pfad des Random Walk einen beliebig vorgegebenen Wert  $z \in Z$ ; und wie oft? Die Antwort hierzu lautet (ohne Begründung):

Ein Pfad des symmetrischen Random Walk ( $p = q = \frac{1}{2}$ ) trifft, unabhängig vom Start  $S_0$ , jedes  $z \in \mathbb{Z}$  mit Wahrscheinlichkeit (Wkt) 1, und zwar beliebig oft. Diese Aussage gilt für  $p \neq q$  nicht mehr.

Weitere Fragen waren:

1. Verläßt ein beliebiger Pfad des Random Walk ein vorgegebenes endliches Intervall  $[a, b]$ ,  $a < b$ ; womöglich sogar mit Wkt 1 ?
2. Was geschieht im Fall  $K \rightarrow \infty$ , wenn die Bank unbeschränktes Kapital besitzt?
3.  $D$  sei die Spieldauer, also die Zeit, bei der das Kapital des Spielers (ausgehend von  $S_0 = k$ ) zum ersten Mal einen der Werte  $S_D = 0$  oder  $S_D = K$  erreicht. Es ist u.a. zu klären, ob  $D$  endlich ist.

Zur Beantwortung der Fragen ziehen wir zunächst die bereits vorliegenden Ergebnisse von Beispiel 1.30 heran. Dort erhielt man in (1.11) und (1.12) mit  $r := q/p$  für Startkapital  $S_0 = k \geq 0$  und Zielkapital  $K \geq k$ , (das Spiel wird beendet, wenn zum ersten Mal  $S_n = 0$  oder  $S_n = K$ )

$$\begin{aligned}
 p_k &= P(\exists n \geq 0 \text{ mit } S_n = 0 \text{ und } S_j < K, j < n \mid S_0 = k) \\
 &= \begin{cases} \frac{K - k}{K}, & \text{falls } p = \frac{1}{2} \\ \frac{r^k - r^K}{1 - r^K}, & \text{falls } p \neq \frac{1}{2}. \end{cases} \tag{5.2}
 \end{aligned}$$

und

$$\begin{aligned}
 q_k &= P(\exists n \geq 0 \text{ mit } S_n = K \text{ und } S_j > 0, 0 < j \leq n \mid S_0 = k) \\
 &= \begin{cases} \frac{k}{K} & , \text{ falls } p = \frac{1}{2} \\ \frac{1 - r^k}{1 - r^K} & , \text{ falls } p \neq \frac{1}{2}. \end{cases} \tag{5.3}
 \end{aligned}$$

$p_k$  gibt die Wahrscheinlichkeit an, dass  $S_n = 0$  vor dem Ereignis  $S_n = K$  eintritt; Analoges trifft für  $q_k$  zu. Gemäß Satz 1.31 gilt

$$p_k + q_k = 1.$$

Brechen wir den Pfad nicht bei Spielende ab, sondern setzen diesen unbeschränkt fort, so folgt daraus (mit geeigneten Translationen) die Beantwortung von **Frage 1** und **Frage 3** in Bem. 5.2:

Mit Wkt 1 verläßt ein Pfad des Random Walk jedes beliebige beschränkte Intervall  $[a, b]$ ,  $a < b$ . Damit ist die Spieldauer  $D$  mit Wkt 1 endlich.

Die Antwort auf **Frage 2** lautet mit (5.2): Wenn die Bank unbeschränktes Kapital besitzt, so ist die Wkt  $p_R$  für Eintritt des Ruins

$$p_R = \begin{cases} 1 & , \text{ falls } p \leq \frac{1}{2} \\ r^k = \left(\frac{1-p}{p}\right)^k & , \text{ falls } p > \frac{1}{2}. \end{cases} \tag{5.4}$$

Für  $p > \frac{1}{2}$  und genügend großes Startkapital  $k$  besteht also durchaus eine reelle Chance, Gewinn zu machen.

Wir wollen Gleichung (5.2) für  $K \rightarrow \infty$  näher untersuchen und stellen eine weitere Frage, nämlich nach der sog. **Rückkehr zur 0** (recurrence). Die Wahl  $S_0 = 0$  erfolgt nur aus formalen Gründen. Die Überlegungen ändern sich bei beliebigem  $S_0 \in \mathbb{Z}$  nicht. Wir folgen in der Darstellung [Isaac (1995)], pp. 192. Zunächst seien  $p = q = \frac{1}{2}$ , also  $r = 1$ . Aus dem Satz von der totalen Wkt ergibt sich wegen der Unabhängigkeit der  $X_i$

$$P(S_n = 0 \text{ für ein } n > 0 \mid S_0 = 0) = \tag{5.5}$$

$$\frac{1}{2} P(S_n = 0 \text{ für ein } n > 1 \mid S_1 = -1) + \frac{1}{2} P(S_n = 0 \text{ für ein } n > 1 \mid S_1 = 1).$$

Nun betrachtet man

$$P(S_n = 0 \text{ für ein } n > 1 \mid S_1 = 1) = P(S_n = 0 \text{ für ein } n > 0 \mid S_0 = 1), \tag{5.6}$$

also die Wkt je 0 zu erreichen, wenn man in 1 startet. Nun nützen wir das Resultat (5.2) vom Ruin des Spielers zunächst für endliches  $K > 0$  und erhalten in diesem symmetrischen Fall mit Startkapital  $k = 1$  und  $p_1 = 1 - 1/K$

$$\begin{aligned} P(S_n = 0 \text{ vor Eintritt von } S_n = K \mid S_0 = 1) &= \\ P(\text{Ruin des Spielers vor Eintritt von } S_n = K \mid S_0 = 1) &= 1 - \frac{1}{K}. \end{aligned} \quad (5.7)$$

Für  $K \rightarrow \infty$  strebt die rechte Seite  $1 - 1/K$  von (5.7) gegen 1 und die linke Seite konvergiert gegen die gesuchte Wkt von (5.6), die somit gleich 1 sein muss. Aus Symmetriegründen folgt dieselbe Aussage, wenn  $S_0 = 1$  durch  $S_0 = -1$  ersetzt wird. Gemäß (5.5) kehrt damit der symmetrische Random Walk ausgehend von  $S_0 = 0$  mit Wkt 1 zu 0 zurück.

Bei der Untersuchung des allgemeinen Falls  $p \neq q$  erhalten wir analog zu (5.5)

$$\begin{aligned} P(S_n = 0 \text{ für ein } n > 0 \mid S_0 = 0) &= \\ q P(S_n = 0 \text{ für ein } n > 1 \mid S_1 = -1) + p P(S_n = 0 \text{ für ein } n > 1 \mid S_1 = 1). \end{aligned} \quad (5.8)$$

Sei  $A_K$  das Ereignis  $S_n = 0$  vor Eintritt von  $S_n = K$ . Für  $p \neq q$  gilt entsprechend (5.6) gemäß (5.3) mit  $r = q/p \neq 1$  und  $k = 1$

$$\begin{aligned} P(S_n = 0 \text{ für ein } n > 0 \mid S_0 = 1) &= \\ \lim_{K \rightarrow \infty} P(A_K \mid S_0 = 1) &= \lim_{K \rightarrow \infty} p_1 = \lim_{K \rightarrow \infty} \frac{r - r^K}{1 - r^K} \\ &= \begin{cases} \frac{q}{p} & , \text{ falls } p > q \\ 1 & , \text{ falls } p < q. \end{cases} \end{aligned} \quad (5.9)$$

Bei einer **Drift** des Random Walk **nach rechts** ( $p > q$ ) gibt es also eine positive Wkt  $1 - q/p$ , nie wieder zur 0 zurückzukehren.

Für  $p < q$  liegt eine **Links-Drift** des Random Walk vor. Da die Rückkehrwahrscheinlichkeit (von  $S_0 = 1$  aus) schon im symmetrischen Fall  $p = q = \frac{1}{2}$  gleich 1 war, muss dies natürlich erst recht für  $p < q$  gelten. Für die Rückkehrwahrscheinlichkeit ausgehend von  $S_0 = -1$  ist die Formel (5.3) heranzuziehen. Weiterhin muss noch die iid-Eigenschaft der  $X_i$  berücksichtigt werden. Sei  $B_K$  das Ereignis  $S_n = K$  vor Eintritt von  $S_n = 0$ .

$$\begin{aligned} P(S_n = 0 \text{ für ein } n > 0 \mid S_0 = -1) &= \\ \lim_{K \rightarrow \infty} P(B_K \mid S_0 = K - 1) &= \lim_{K \rightarrow \infty} q_{K-1} = \lim_{K \rightarrow \infty} \frac{1 - r^{K-1}}{1 - r^K} \\ &= \begin{cases} \frac{1}{r} = \frac{p}{q} & , \text{ falls } p < q \\ 1 & , \text{ falls } p > q. \end{cases} \end{aligned} \quad (5.10)$$

Nun sind alle Formeln wieder zusammensetzen und man erhält mit (5.8)



**Satz 5.3 (Rückkehr zur 0, recurrence)**

Bei der einfachen Irrfahrt (Random Walk) von Def. 5.1 gilt mit  $q = 1 - p$

$$P(S_n = 0 \text{ für ein } n > 0 \mid S_0 = 0) = \begin{cases} 1 & , \text{ falls } p = q = \frac{1}{2} \\ 2p & , \text{ falls } p < q \\ 2q & , \text{ falls } p > q. \end{cases} \quad (5.11)$$

Im Fall  $p \neq q$  gibt es also eine positive Wkt  $1 - 2 \min\{p, q\}$ , nicht mehr zur 0 zurückzukehren.

**5.2 Stoppzeiten**

Zum Random Walk gibt es eine Reihe weiterer, interessanter Fragestellungen, die elementar, wie wir das bisher getan haben, oder mit modernen stochastischen Methoden angegangen werden können. Wir wollen hier an drei Beispielen nur die Problemstellung skizzieren und für eine genaue Untersuchung auf die Vorlesung Stochastische Prozesse oder entsprechende Lehrbücher hinweisen.

**1. Spieldauer; duration of the game**

Man betrachtet den Random Walk mit dem Startkapital  $S_0 = k \in \mathbb{N}_0$  und definiert für ein festgelegtes Zielkapital  $K \geq k$

$$D := \min\{n \in \mathbb{N}_0 : S_n = 0 \text{ oder } S_n = K\}. \quad (5.12)$$

Man spricht beim Random Walk auch von der ersten Passierzeit (*passage time*) durch  $(0, K)$ .

**2. Rekurrenzzeit, Übergangszeit**

In ähnlicher Weise definiert man die Rekurrenz- oder Rückkehrzeit (*recurrence time*)

$$T_{00} := \min\{n \geq 2 : S_n = 0, \text{ für } S_0 = 0\} \quad (5.13)$$

oder für beliebiges  $z \in \mathbb{Z}$  die Übergangszeit (*hitting time*)

$$T_{0z} := \min\{n \geq 0 : S_n = z, \text{ für } S_0 = 0\}. \quad (5.14)$$

Die Spieldauer  $D$  beim Ruin des Spielers ist gemäß Satz 1.31 mit Wkt 1 endlich. Damit ist  $D$  eine ZV im bisher verwendeten Sinn.

$T_{00}$  ist nur im symmetrischen Fall  $p = q = \frac{1}{2}$  mit Wkt 1 endlich und bei  $T_{0z}$  hängt die Endlichkeit vom Vorzeichen von  $z$  sowie dem Drift des Random Walk ab. Eine Herleitung der Wahrscheinlichkeitsverteilungen von  $T_{00}$  bzw.  $T_{0z}$  findet man z. B. in [Stirzaker (1994)], pp. 146 mit elementaren Methoden oder in [Resnick (1994)], pp. 33

über erzeugende Funktionen. Weiterhin ist ein Zugang über Markovketten oder Martingale möglich, die in weiterführenden Stochastikvorlesungen ausführlich besprochen werden.

Wir wollen wenigstens ein wichtiges Resultat zu Übergangszeiten referieren; vgl. z. B. [Stirzaker (1994)], pp. 146:

Wegen der iid-Eigenschaft der  $X_i$  gilt

$$T_{02} = T_{01} + T_{12},$$

wobei  $T_{01}$  und  $T_{12}$  dieselbe Verteilung haben und unabhängig sind ( $T_{12}$  ist analog zu  $T_{0z}$  definiert). Damit kann man sich auf die Analyse des Spezialfalls  $T_{01}$  beschränken. Hier gilt u.a.

$$E(T_{01}) = \begin{cases} \frac{1}{p-q} & , \text{ falls } p > q \\ \infty & , \text{ sonst.} \end{cases} \quad (5.15)$$

Dabei ist interessant, dass für  $p = q = \frac{1}{2}$  gilt:  $P(T_{01} < \infty) = 1$  aber  $E(T_{01}) = \infty$ .

$D$ ,  $T_{00}$  und  $T_{0z}$  sind spezielle **Stoppzeiten** (auch Stopzeiten). Bei diesem wichtigen Typ von ZV (hier werden also auch Werte wie  $\infty$  als Ergebnis zugelassen) kann zu jeder Zeit (hier  $n \in \mathbb{N}_0$ ) des stochastischen Prozesses (hier  $S_n$ ) festgestellt werden, ob das Stoppkriterium schon eingetreten ist oder nicht. Die praktische Relevanz dieser wichtigen Eigenschaft sei abschließend an einem Beispiel demonstriert:

*Ein Devisenspekulant hat zu einem gewissen Zeitpunkt  $t = 0$  den Betrag von  $A$  US \$ aufgekauft und möchte diese zur Zeit  $T_{opt}$  bei einem maximalen Kurs abstoßen.*

*$T_{opt}$  ist keine Stoppzeit, da am 1.12.1997 nicht entschieden werden kann, ob das Stoppkriterium schon eingetreten ist oder erst in der Zukunft eintreten wird.*

*Dagegen ist das Eintreten des Zeitpunkts  $t > 0$ , zu dem der US \$ zum ersten Mal einen Kurs von 2 DM erreicht, feststellbar. Das entsprechend definierte  $T_{2,00}$  ist eine Stoppzeit, die auch den Wert  $\infty$  annehmen kann, wenn nämlich der Kurs des US \$ nie mehr über 2 DM klettert.*

# Kapitel 6

## Eine Auswahl wichtiger Verteilungen

Dieses Kapitel dient dem schnellen Auffinden einiger Charakteristika wichtiger Verteilungen. Teilweise findet man die an dieser Stelle zusammengefaßten Ergebnisse verstreut über die vorangegangenen Abschnitte, andere werden hier erstmals aufgeführt. Dies trifft etwa zu auf die Multinomial- oder die Weibull-Verteilung. Die Stichprobenverteilungen zur statistischen Analyse von normalverteilten iid-Zufallsvariablen (ZV) wie die  $\chi^2$ -, Student- und Fisher-Verteilung werden in Kapitel 9 behandelt.

$f$  bezeichnet jeweils die Wahrscheinlichkeitsfunktion bzw. Dichte der betrachteten ZV  $X$ ,  $F$  deren Verteilungsfunktion sowie  $\bar{F}$  die *tail probability*, die oft als Überlebenswahrscheinlichkeit interpretiert werden kann. Neben Erwartungswert und Varianz sind ggf. auch Momente, Kumulanten, (moment-) erzeugende Funktion und Besonderheiten der zugrundeliegenden Verteilung angegeben.

Übersichten von Verteilungen mit weiteren Details findet man etwa im Lexikon der Stochastik von [Müller (1975)] oder in den drei Büchern von Johnson und Kotz über diskrete sowie univariate und multivariate stetige Verteilungen; zum Beispiel [Johnson & Kotz (1970)].

## 6.1 Einige diskrete Verteilungen

**Binomialverteilung:**  $B(n, p)$

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n, \quad 0 < p < 1 \quad (6.1)$$

$$E(X) = np, \quad \text{Var}(X) = np(1-p) \quad (6.2)$$

$$G(s) = E(s^X) = (1-p + ps)^n \quad (\text{erzeugende Funktion}) \quad (6.3)$$

Besonderheiten:

1. Approximation durch die Normalverteilung siehe Abschnitt 4.3.1
2. Approximation durch die Poisson-Verteilung siehe Abschnitt 4.3.2

**Poisson-Verteilung:**  $Poi(\lambda)$

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \dots, \quad \lambda > 0 \quad (6.4)$$

$$E(X) = \lambda, \quad \text{Var}(X) = \lambda \quad (6.5)$$

$$G(s) = E(s^X) = e^{\lambda(s-1)} \quad (\text{erzeugende Funktion}) \quad (6.6)$$

$$M(s) = E(e^{sX}) = \exp[\lambda(e^s - 1)] \quad (\text{momenterzeugende Funktion}) \quad (6.7)$$

$$K(s) = \lambda(e^s - 1) \quad (\text{Kumulantenfunktion}) \quad (6.8)$$

$$\kappa_k = \lambda, \quad k = 1, 2, 3, \dots \quad (\text{Kumulanten}) \quad (6.9)$$

Besonderheiten:

1. Approximation durch die Normalverteilung siehe Abschnitt 4.3.1
2. Grenzverteilung der Binomialverteilung siehe Abschnitt 4.3.2

**Geometrische Verteilung**

$$f(x) = p(1-p)^x, \quad x = 0, 1, 2, \dots \quad (6.10)$$

$$E(X) = \frac{1-p}{p}, \quad \text{Var}(X) = \frac{1-p}{p^2} \quad (6.11)$$

$$G(s) = E(s^X) = \frac{p}{1-(1-p)s} \quad (\text{erzeugende Funktion}) \quad (6.12)$$

**Negative Binomialverteilung  $NB(v, p)$** 

$$f(x) = \binom{-v}{x} (-p)^x (1-p)^v, \quad x = 0, 1, \dots \quad 0 < p < 1 \quad v > 0 \quad (6.13)$$

$$E(X) = \frac{pv}{1-p}, \quad \text{Var}(X) = \frac{pv}{(1-p)^2} \quad (6.14)$$

Besonderheiten:

1. Für  $v \in \mathbb{N}_0$  ergibt eine negativ binomialverteilte ZV  $X$  die Anzahl der Fehlversuche vor dem  $v$ -ten Erfolg bei unabhängigen Bernoulli-Versuchen.
2. Auch hier ist die Poisson-Verteilung eine Grenzverteilung; siehe etwa [Müller (1975)].

**Hypergeometrische Verteilung:  $H(n, N, M)$** 

Diese spielt eine wichtige Rolle in der Stichprobentheorie. In einer Urne seien  $N$  Kugeln, davon  $0 < M < N$  rot gefärbt und der Rest weiß.  $X$  sei die Anzahl der roten Kugeln beim  $n$ -maligen (unabhängigen) Ziehen einer Kugel ohne Zurücklegen. Gemäß Satz 1.22 gilt

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}, \quad 0 \leq x \leq \min\{M, n\}. \quad (6.15)$$

$$E(X) = n \frac{M}{N}, \quad \text{Var}(X) = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1} \quad (6.16)$$

Besonderheiten:

1. Wichtige Stichprobenverteilung.
2. Für  $M \rightarrow \infty$  und  $N \rightarrow \infty$ , so dass  $\lim \frac{M}{N} = p$ , ergibt sich in der Grenze die Binomialverteilung.

**Multinomialverteilung**  $M(n, \boldsymbol{\pi})$ 

Seien  $k \in \mathbb{N}$ ,  $A_1, \dots, A_k$  disjunkte Ereignisse und

$$A_{k+1} := \Omega \setminus \left( \bigcup_{j=1}^k A_j \right) \quad (6.17)$$

mit

$$P(A_j) = \pi_j > 0, \quad j = 1, \dots, k+1.$$

Wir setzen  $\boldsymbol{\pi} := (\pi_1, \dots, \pi_k)^T \in \mathbb{R}^k$  und erhalten wegen (6.17)

$$\pi_{k+1} = 1 - \sum_{j=1}^k \pi_j. \quad (6.18)$$

Bei  $n$  unabhängigen Versuchsdurchführungen trete  $X_1$  mal das Ereignis  $A_1$ ,  $X_2$  mal das Ereignis  $A_2$ , ... und  $X_{k+1}$  mal das Ereignis  $A_{k+1}$  auf. Dann gilt wieder wegen (6.17)

$$\sum_{j=1}^{k+1} X_j = n \quad \text{oder} \quad X_{k+1} = n - \sum_{j=1}^k X_j. \quad (6.19)$$

Für  $x_j \in \mathbb{N}_0$ ,  $j = 1, \dots, k+1$  und  $\sum_{j=1}^{k+1} x_j = n$  lautet die Wahrscheinlichkeitsfunktion der  $k$ -dimensionalen ZV  $\mathbf{X} := (X_1, \dots, X_k)^T$  mit  $\mathbf{x} := (x_1, \dots, x_k)^T$

$$\begin{aligned} f_M(\mathbf{x}) &= P(X_1 = x_1, \dots, X_k = x_k) \\ &= \frac{n!}{x_1! \cdot \dots \cdot x_k! \cdot x_{k+1}!} \pi_1^{x_1} \cdot \dots \cdot \pi_k^{x_k} \cdot \pi_{k+1}^{x_{k+1}}. \end{aligned} \quad (6.20)$$

Wegen der Nebenbedingungen (6.18), (6.19) und  $x_{k+1} = n - \sum_{j=1}^k x_j$  genügt es,  $\mathbf{X}$  als eine  $k$ -dimensionale ZV zu betrachten. Wir haben übrigens für  $k = 1$  diese Vorgehensweise bei der Binomialverteilung ganz selbstverständlich angewendet. Dort galt für  $0 < p < 1$  mit

$$\begin{aligned} \boldsymbol{\pi} &= (\pi_1)^T = \pi_1 := p, \quad \pi_2 := 1 - p, \\ \mathbf{x} &= (x_1)^T = x_1 = x, \quad x_2 := n - x_1, \quad 0 \leq x_1, x_2 \leq n \end{aligned}$$

für die Wahrscheinlichkeitsfunktion

$$f(x_1, x_2) = \frac{n!}{x_1! \cdot x_2!} \pi_1^{x_1} \cdot \pi_2^{x_2} = \binom{n}{x} p^x (1-p)^{n-x} = f_B(x).$$

Die Multinomialverteilung ist eine  $k$ -dimensionale diskrete Verteilung mit

$$E(X_j) = n \pi_j \quad (6.21)$$

und

$$\text{Cov}(X_i, X_j) = \begin{cases} n \pi_i (1 - \pi_i) & , \text{ falls } i = j \\ -n \pi_i \pi_j & , \text{ sonst,} \end{cases} \quad 1 \leq i, j \leq k. \quad (6.22)$$

Zur Herleitung siehe etwa [Chung (1979)], section 6.4.

Die Multinomialverteilung bildet wie die Poisson-, Normal- und Gammaverteilung eine sog. Exponentialfamilie. Diese Eigenschaft spielt in der Mathematischen Statistik bei der Konstruktion von optimalen Schätzfunktionen und Tests eine wesentliche Rolle; vgl. etwa [Witting (1985)].

## 6.2 Beispiele stetiger Verteilungen

**Normal (Gauß)-Verteilung**  $N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \mu \in \mathbb{R}, \quad \sigma^2 > 0 \quad (6.23)$$

$N(0, 1)$  heißt **Standardnormalverteilung** und hat die spezielle Dichte

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

Die Verteilungsfunktion

$$\Phi(u) = \int_{-\infty}^u \varphi(t) dt$$

der Standardnormalverteilung  $N(0, 1)$  ist analytisch. Eine geschlossene Darstellung mit bekannten Standardfunktionen ist jedoch für  $\Phi$  nicht möglich. Allerdings gibt es ausgezeichnete numerische Approximationen. Zum Ablesen von Quantilen (vgl. Abschnitt 1.5.4) verwendet man Tabellen von  $\Phi$ , die in den meisten Statistik-Einführungsbüchern zu finden sind.

Nun sei  $X$  wieder  $N(\mu, \sigma^2)$

$$E(X) = \mu, \quad \text{Var}(X) = \sigma^2 \quad (6.24)$$

$$M(s) = E(e^{sX}) = \exp\left(\mu s + \frac{1}{2}\sigma^2 s^2\right) \quad (\text{momenterzeugende Funktion}) \quad (6.25)$$

Besonderheiten:

1. Zentrale Momente:

$$E[(X - \mu)^k] = \begin{cases} 0 & , \text{ falls } k \text{ ungerade} \\ 1 \cdot 3 \cdot 5 \cdot \dots \cdot (k-1) \sigma^k & , \text{ sonst} \end{cases} \quad (6.26)$$

2.  $\gamma_3 := E[(X - \mu)^3] / \sigma^3 = 0$  (Schiefe, skewness)

3.  $\delta_4 := E[(X - \mu)^4] / \sigma^4 - 3 = 0$  (Exzess, Wölbung, kurtosis).

4. Alle Kumulanten ab der Ordnung 3 verschwinden.

5. Summen normalverteilter ZV sind wieder normalverteilt. Speziell gilt für unabhängige  $X \sim N(\mu_x, \sigma_x^2)$  und  $Y \sim N(\mu_y, \sigma_y^2)$

$$X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2).$$

**Gleichverteilung (uniform distribution)  $U(a, b)$  in  $(a, b)$ ,  $a < b$** 

Man spricht auch von der Rechteckverteilung. Es spielt keine Rolle, ob das offene oder das abgeschlossene Intervall zwischen  $a$  und  $b$  betrachtet wird, weil hier eine stetige ZV  $X$  vorliegt, für die  $P(X = a) = P(X = b) = 0$  gilt.

$$f(x) = \frac{1}{b-a} 1_{(a,b)}(x) = \begin{cases} \frac{1}{b-a} & , \text{ falls } a < x < b \\ 0 & , \text{ sonst,} \end{cases} \quad (6.27)$$

$$E(X) = (a+b)/2, \quad Var(X) = (b-a)^2/12 \quad (6.28)$$

Die Gleichverteilung hat folgende zentralen Momente

$$E[(X - \mu)^k] = \begin{cases} 0 & , \text{ falls } k \text{ ungerade} \\ \frac{1}{k+1} \left(\frac{b-a}{2}\right)^k & , \text{ sonst.} \end{cases} \quad (6.29)$$

Besonderheiten für den Spezialfall  $X \sim U(0, c)$ ,  $c > 0$ :

1. Die momenterzeugende Funktion lautet

$$M(s) = \frac{e^{cs} - 1}{cs} \quad (6.30)$$

2. Für den Minimum-Varianz-Schätzer  $\hat{c}$  von  $c$  siehe Kapitel 8.

**Exponentialverteilung  $ED(\lambda)$** 

$$f(x) = \lambda e^{-\lambda x} 1_{(0,\infty)}(x) \quad (6.31)$$

$$F(x) = (1 - e^{-\lambda x}) 1_{(0,\infty)}(x) \quad (6.32)$$

$$\bar{F}(x) = e^{-\lambda x} \quad \text{Überlebenswahrscheinlichkeit für } x > 0 \quad (6.33)$$

$$E(X) = \frac{1}{\lambda}, \quad Var(X) = \frac{1}{\lambda^2} \quad (6.34)$$

$$M(s) = E(e^{sX}) = \frac{\lambda}{\lambda - s} \quad (\text{momenterzeugende Funktion}) \quad (6.35)$$

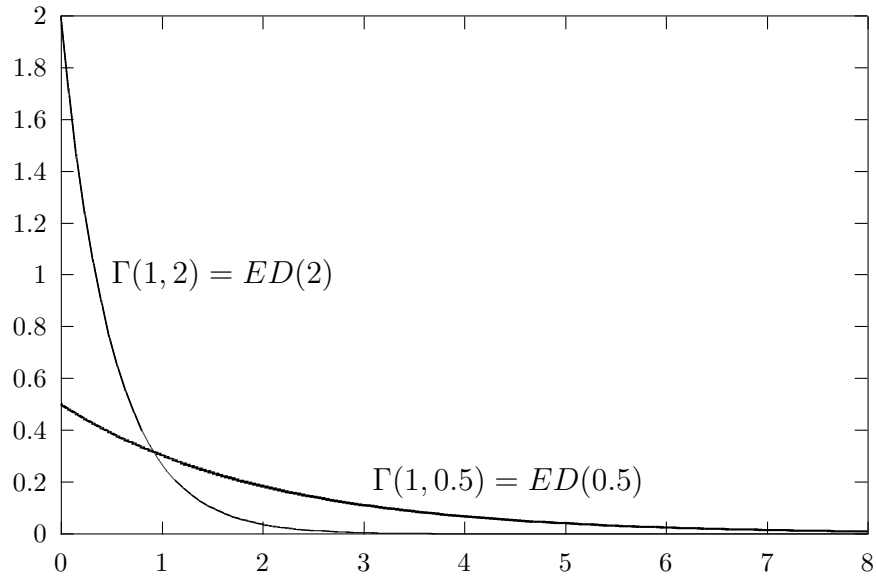
Besonderheiten:

1. Lebensdauerverteilung
2. Gedächtnislosigkeit; vgl. Bsp. 1.45:

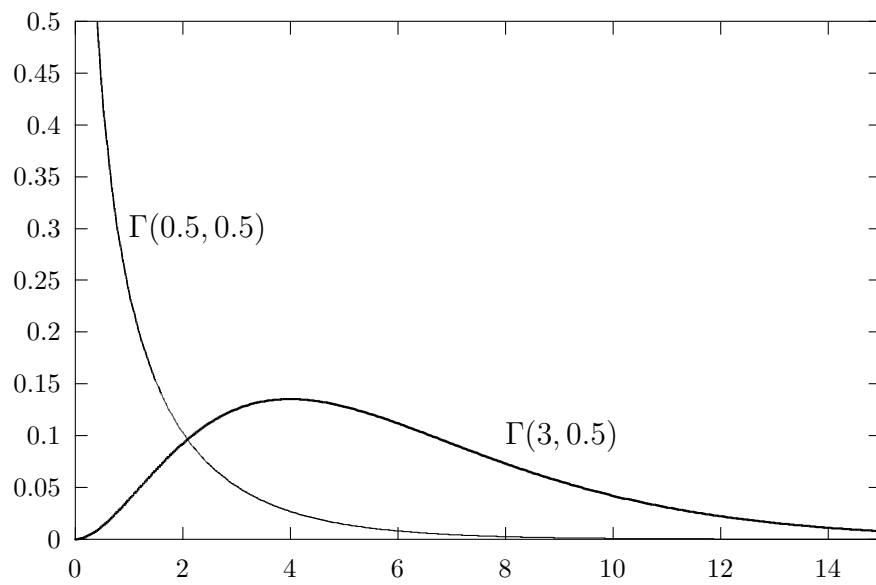
$$P(X > x + t_0 | X > t_0) = P(X > x).$$

3. Das Minimum von  $n$  iid  $ED(\lambda)$ -verteilten ZV ist wieder exponentialverteilt und zwar mit Parameter  $n\lambda$ ; d. h. der Erwartungswert des Minimums ist ein  $n$ -tel des ursprünglichen Erwartungswerts einer der iid-Variablen.





**Bild 11:** Dichten von Exponentialverteilungen



**Bild 12:** Dichten von Gammaverteilungen mit  $\lambda = 0.5$

**Gammaverteilung**  $\Gamma(q, \lambda)$ ,  $q > 0$  und  $\lambda > 0$

Für  $z > 0$  bezeichne

$$\Gamma(z) := \int_0^{\infty} t^{z-1} e^{-t} dt \quad (6.36)$$

die **Gammafunktion** mit  $\Gamma(z+1) = z\Gamma(z)$ ;  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ ;  $\Gamma(n) = (n-1)!$ ,  $n \in \mathbb{N}$ . Die Dichte der Gammaverteilung lautet

$$f(x) = \frac{\lambda^q}{\Gamma(q)} x^{q-1} e^{-\lambda x} 1_{(0, \infty)}(x). \quad (6.37)$$

Damit erhält man

$$E(X) = \frac{q}{\lambda}, \quad Var(X) = \frac{q}{\lambda^2} \quad (6.38)$$

$$M(s) = E(e^{sX}) = \left( \frac{\lambda}{\lambda - s} \right)^q \quad (\text{momenterzeugende Funktion}) \quad (6.39)$$

Besonderheiten:

1. Momente:

$$E[X^k] = \frac{q(q+1) \cdots (q+k-1)}{\lambda^k}, \quad k = 1, 2, \dots \quad (6.40)$$

2. Spezialfälle dieser Verteilungsfamilie sind u.a.

- Exponentialverteilung  $ED(\lambda) = \Gamma(1, \lambda)$
- $\chi_n^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$  Summe von  $n$  Quadraten von unabhängigen  $N(0, 1)$ -verteilten ZV; siehe Kapitel 9.

3. Die Summe  $X + Y$  unabhängiger  $\Gamma(q, \lambda)$ - bzw.  $\Gamma(r, \lambda)$ -verteilter ZV  $X$  bzw.  $Y$  ist  $\Gamma(q+r, \lambda)$ -verteilt, wobei  $q, r, \lambda > 0$ .

4. Die Summe  $V$  von  $n$  unabhängigen  $ED(\lambda)$ -verteilten ZV ist  $Erlang(n, \lambda) = \Gamma(n, \lambda)$ -verteilt; siehe Bsp. 3.14. Die Dichte der Erlangverteilung lautet

$$f(v) = \frac{\lambda^n}{(n-1)!} v^{n-1} e^{-\lambda v} 1_{(0, \infty)}(v). \quad (6.41)$$

**Cauchy-Verteilung**

$$f(x) = \frac{a}{\pi(a^2 + x^2)}, \quad a > 0 \quad (6.42)$$

Besonderheiten:

1.  $E(X)$  und  $Var(X)$  existieren nicht.
2. Der Quotient zweier unabhängig  $N(0, 1)$ verteilter ZV ist Cauchy-verteilt mit Parameter  $a = 1$ . Damit ist die Cauchy-Verteilung ( $a = 1$ ) eine Studentverteilung (siehe Kapitel 9) mit Freiheitsgrad 1.

**Pareto-Verteilung**

Seien  $c > 0$  ein gewisser Schwellenwert, den die entsprechende ZV  $X$  nicht erreichen oder unterschreiten kann und  $\alpha > 0$

$$f(x) = \frac{\alpha}{c} \left(\frac{c}{x}\right)^{\alpha+1} 1_{(c, \infty)}(x) \quad (6.43)$$

$$F(x) = 1 - \left(\frac{c}{x}\right)^\alpha 1_{(c, \infty)}(x). \quad (6.44)$$

$$E(X) = \frac{\alpha c}{\alpha - 1}, \quad \text{für } \alpha > 1 \quad (6.45)$$

$$Var(X) = \frac{\alpha c^2}{(\alpha - 1)^2(\alpha - 2)}, \quad \text{für } \alpha > 2 \quad (6.46)$$

**Logistische Verteilung**

Für  $\mu \in \mathbb{R}$ ,  $\sigma > 0$  und  $y(x) := (x - \mu)/\sigma$  lauten Verteilungsfunktion, Erwartungswert und Varianz

$$F(x) = \frac{1}{1 + e^{-\pi y(x)/\sqrt{3}}}, \quad E(X) = \mu, \quad Var(X) = \sigma^2. \quad (6.47)$$

**Lognormalverteilung**  $LogN(\mu, \sigma^2)$ 

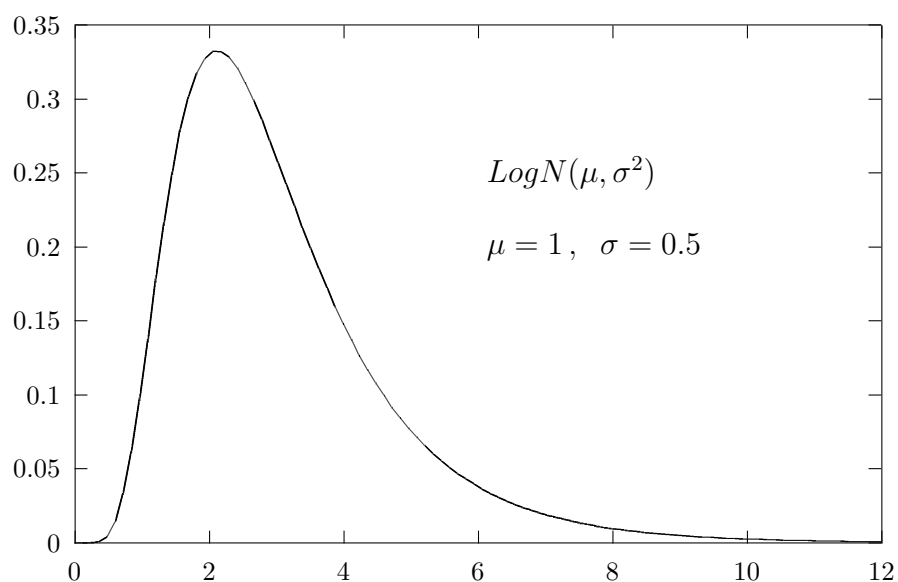
Eine positive ZV  $X$  heißt logarithmisch normalverteilt, wenn  $\ln(X)$  normalverteilt ist mit Mittelwert  $\mu$  und Varianz  $\sigma^2$ .

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma x} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) 1_{(0,\infty)}(x) \quad (6.48)$$

$$E(X) = e^{\mu+\sigma^2/2}, \quad Var(X) = e^{2\mu+2\sigma^2} (e^{\sigma^2} - 1) \quad (6.49)$$

Besonderheiten:

1. Median:  $x_{0.5} = e^\mu$
2. Modus:  $x_M = e^{\mu-\sigma^2}$
3. Die Lognormalverteilung ist eine einseitige unsymmetrische Verteilung und wird u.a. zur Modellierung von Lebensdauern herangezogen.



**Bild 13:** Dichte der Lognormalverteilung  $LogN(\mu, \sigma^2)$

**Weibull-Verteilung**  $W(r, \lambda)$ 

$$f(x) = \lambda r x^{r-1} \exp(-\lambda x^r) 1_{(0, \infty)}(x), \quad r > 0, \lambda > 0 \quad (6.50)$$

$$F(x) = 1 - e^{-\lambda x^r}, \quad \bar{F}(x) = e^{-\lambda x^r}, \quad x > 0 \quad (6.51)$$

$$E(X) = \lambda^{-1/r} \Gamma\left(\frac{1}{r} + 1\right) \quad \text{Var}(X) = \lambda^{-2/r} \left[ \Gamma\left(\frac{2}{r} + 1\right) - \Gamma^2\left(\frac{1}{r} + 1\right) \right] \quad (6.52)$$

Besonderheiten:

1. Median:  $x_{0.5} = \left(\frac{\ln 2}{\lambda}\right)^{1/r}$
2. Modus:  $x_M = \left(\frac{r-1}{r\lambda}\right)^{1/r}$  für  $r \geq 1$
3. Lebensdauerverteilung
4. Eine der drei Grenzverteilungen für Extremwerte (bei geeigneter Parametrisierung)
5. Die Exponentialverteilung  $ED(\lambda) = W(1, \lambda)$  ist eine spezielle Weibull-Verteilung.

**n-dimensionale Normalverteilung  $N(\boldsymbol{\mu}, \mathbf{C})$** 

Für  $\mathbf{X} = (X_1, \dots, X_n)^T$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T \in \mathbb{R}^n$ , eine positiv definite, und damit symmetrische Matrix  $\mathbf{C} = (c_{ij}) \in \mathbb{R}^{n,n}$  lautet die Dichte der n-dimensionalen Normalverteilung

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C})}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right], \quad \mathbf{x} \in \mathbb{R}^n. \quad (6.53)$$

Die Parameter sind der Erwartungswert und die Kovarianzmatrix des Vektors  $\mathbf{X}$

$$E(\mathbf{X}) = \boldsymbol{\mu}, \quad \mathbf{Cov}(\mathbf{X}) = \mathbf{C}. \quad (6.54)$$

Im wichtigen Spezialfall  $n = 2$  erhält man die bivariate Normalverteilung, vgl. auch die Kapitel 1 bis 2.

Sei  $(Z, Y)^T$  ein gemeinsam normalverteilter zweidimensionaler Zufallsvektor mit  $E(Z) = \mu_Z$ ,  $E(Y) = \mu_Y$ ,  $Var(Z) = \sigma^2 > 0$ ,  $Var(Y) = \tau^2 > 0$  und  $Cov(Z, Y) = \rho \sigma \tau$  mit dem Korrelationskoeffizienten  $-1 < \rho < 1$ . In diesem Fall kann die Dichte (6.53) geschrieben werden als

$$f_{Z,Y}(z, y) = \frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}} \exp \left[ -\frac{1}{2(1-\rho^2)} \left( \frac{(z-\mu_Z)^2}{\sigma^2} - \frac{2\rho(z-\mu_Z)(y-\mu_Y)}{\sigma\tau} + \frac{(y-\mu_Y)^2}{\tau^2} \right) \right]. \quad (6.55)$$

# Kapitel 7

## Deskriptive Statistik

### Was ist Statistik?

Im Rahmen der Mathematik versteht man unter dem Begriff **Statistik** vor allem zweierlei, nämlich:

1. Eine wissenschaftliche Disziplin, die sich u.a. mit dem Studium gesetzmäßiger **Massenerscheinungen** befaßt (keine Aussagen für den Einzelfall; vgl. etwa Sterbetafeln bei Lebensversicherungen).
2. Die Bezeichnung von sog. **Schätzfunktionen**, z. B.  $\bar{X}$  für die übliche Durchschnittsbildung beim arithmetischen Mittel.

Für eine weitere Präzisierung siehe die Einleitung zu Kapitel 8.

**Statistiken** (im landläufigen Sinn) nennt man anschauliche, meist komprimierte Darstellungen von quantifizierbaren Vorgängen aller Art (z. B. in Tabellen oder Schaubildern), die sich über einen gewissen Zeitraum erstrecken oder die sonst wegen einer Fülle unübersichtlichen Datenmaterials einer direkten Interpretation nicht zugänglich sind; z. B.

Bundesligatabelle, Arbeitslosenstatistik, Statistisches Jahrbuch, Volkszählung, Bevölkerungsstatistik, Inventur, Mietspiegel, Umsatz- und Wirtschaftsstatistik, Eingänge von Forderungen an Versicherungen, Soziologische Befragungen und Auswertungen etc.

Eine nicht ganz feinfühlige, aber treffende Charakterisierung von *Statistik* wird Gérard Calot zugeschrieben (vgl. [Becker (1993)], S. 40)

”Wenn ein Mensch stirbt, ist’s ein Malheur,  
bei 100 Toten eine Katastrophe,  
bei 1000 Toten eine Statistik.”

Zu einer **statistischen Auswertung** gehören u.a.:

Planung der gesamten Untersuchung, Festlegung eines Stichprobenauswahlverfahrens, Datenerfassung, -codierung und -verarbeitung, mathematische Analyse mit einem geeignet gewählten Modell, Datenpräsentation und Interpretation der Ergebnisse.

In angewandten Gebieten wie den Natur-, Ingenieur- oder Wirtschaftswissenschaften muss man sich gelegentlich auf eine anschauliche Darstellung von Versuchs- und Untersuchungsergebnissen beschränken, weil den Adressaten der Präsentation möglicherweise wahrscheinlichkeitstheoretische Grundlagen fehlen. Um die interdisziplinäre Zusammenarbeit zu fördern, beginnen auch wir damit, Möglichkeiten der grafischen Darstellung von Daten und deren Charakterisierung mit geeigneten Maßzahlen aufzuzeigen.

Man spricht von **deskriptiver** oder auch **beschreibender** bzw. **empirischer** Statistik. Auf Elemente der sog. **induktiven** oder **schließenden Statistik** wird am Ende dieser Vorlesung eingegangen. Dort kommen u.a. Eigenschaften von Schätzfunktionen und die statistische Analyse normalverteilter Daten kurz zur Sprache.

Die deskriptive Statistik ist das Bindeglied von einer eher anschaulich orientierten Datenanalyse hin zur abstrakten Wahrscheinlichkeitstheorie. Die Zusammenhänge zwischen Grundgesamtheit und Ergebnismenge, Merkmal und Zufallsvariable, Summenhäufigkeitsfunktion und Verteilungsfunktion sowie vielen statistischen Maßzahlen und entsprechenden Kenngrößen von Verteilungen sind offensichtlich.

## 7.1 Grundgesamtheit, Merkmale

### Def. 7.1 (Grundgesamtheit, Population $\Omega$ )

*Eine statistische Untersuchung bezieht sich stets auf eine klar festgelegte **Grundgesamtheit** oder **Population**  $\Omega$ , die Menge aller denkbaren Beobachtungseinheiten. In der mathematischen Terminologie ist  $\Omega$  also eine nichtleere **Menge**. Die **Elemente**  $\omega \in \Omega$  nennt man auch **Merkmalssträger**.*

### Beispiel 7.2 (Grundgesamtheiten)

1.  $\Omega_1 := \{ \omega \mid \omega \text{ ist Student an der TU München im Wintersemester 1997/98} \}$ .  
 $\Omega_1$  kann mit der Datei aller Matrikelnummern identifiziert werden.
2.  $\Omega_2 := \{ \omega \mid \omega \text{ ist Klient der Versicherung V am 1.11.1997} \}$ .  
 $\Omega_2$  ist die Datei der Kundennummern.
3.  $\Omega_3 := \{ \omega \mid \omega \text{ ist Mietwohnung in der Stadt M am 1.1.1998} \}$ .
4.  $\Omega_4 := \{ \omega \mid \omega \text{ ist Bürger der BRD am 1.12.1997} \}$ .

Eine Schwierigkeit bei statistischen Untersuchungen bzgl.  $\Omega_3$  und  $\Omega_4$  ist, dass es wohl keine Datei gibt, die **alle** zu untersuchenden Objekte enthält.



**Def. 7.3 (Merkmal)**

Gegenstand statistischer Erhebungen ist in der Regel nicht die Grundgesamtheit  $\Omega$  selbst, sondern Eigenschaften ihrer Elemente. Unter einem **Merkmal** versteht man eine Abbildung

$$X : \Omega \rightarrow \mathbb{R}, \quad (7.1)$$

die jedem Merkmalsträger  $\omega \in \Omega$  eine Zahl  $x = X(\omega)$  zuordnet.

$$X(\Omega) := \{ x \in \mathbb{R} \mid x = X(\omega) \} \quad (7.2)$$

nennt man die Menge der **Merkmalsausprägungen**.

Zufallsvariable sind nach dieser Definition spezielle Merkmale. In vielen Büchern werden die Begriffe *Merkmal* und *Zufallsvariable* synonym verwendet.

### 7.1.1 Klassifizierung von Merkmalen, Skalentypen

**Def. 7.4 (diskret, stetig, qualitativ, quantitativ)**

Falls

$$X(\Omega) = \{a_1, \dots, a_j, \dots\}, \quad (7.3)$$

nennen wir  $X$  **diskret**. Merkmale, die alle Werte eines Intervalls  $I \subseteq \mathbb{R}$  annehmen können, d. h.

$$X(\Omega) = [a, b], [0, \infty), \mathbb{R} \text{ usw.}, \quad (7.4)$$

heißen **stetig**.

Weiterhin wird zwischen **qualitativen** und **quantitativen** Merkmalen unterschieden. Alle stetigen Merkmale sind quantitativ.

Beispiele:

Diskrete qualitative Merkmale sind etwa Geschlecht, Religionszugehörigkeit und Prädikat beim Diplomabschluß (mit Auszeichnung, sehr gut, gut, befriedigend, bestanden). Zählvariable wie die Anzahl der Kunden vor einem Bankschalter sind diskrete quantitative Merkmale.

Stetige Merkmale sind z. B. Körpergröße, Auftragsvolumen, Nettomiete/qm, Einkommen etc.

Vor allen in den Wirtschafts- und Sozialwissenschaften ist folgende weitere Klassifizierung üblich:

**Def. 7.5 (Skalentypen: nominal, ordinal, metrisch (kardinal))**

Mit **metrisch** oder **kardinal** werden solche Merkmale bezeichnet, die auf einer **Intervallskala**  $Int \subseteq \mathbb{R}$  gemessen werden. Zu einer solchen Skala gehört u.a. eine Normierung der Abstände durch eine festgelegte Einheit. Stetige Merkmale und diskrete Zählvariable sind metrisch.

Die nächst einfachere Stufe sind **ordinale** Merkmale, bei denen zwar noch eine Rangfolge der Merkmalsausprägungen vorliegt, die Abstände dazwischen aber nicht mehr sinnvoll quantifiziert werden können, z. B. Noten oder Bewertungen wie (einfach, normal, gehoben, sehr gut), etwa für die Ausstattung von Wohnungen.

Überhaupt keine Struktur liegt bei sog. **nominalen** Merkmalen wie Geschlecht, Religionszugehörigkeit, Wahlverhalten etc. vor.

Ordinale und nominale Merkmale sind diskret.

Die Berechnung vieler statistischer Kenngrößen ist nur für metrische (quantitative) Merkmale sinnvoll.

Bei ordinalen Merkmalen sind immerhin noch Rangstatistiken von Interesse, während im nominalen Fall nur die Angabe relativer Häufigkeiten für die verschiedenen Merkmalsausprägungen erfolgen kann.

**Beispiel 7.6 (Merkmale)**

1.  $X_1 : \Omega_1 \rightarrow \{0 = \text{männlich}, 1 = \text{weiblich}\}$ , Geschlecht (nominal).
2.  $X_2 : \Omega_1 \rightarrow \{ \text{mit Auszeichnung} = 0, \text{sehr gut} = 1, \text{gut} = 2, \text{befriedigend} = 3, \text{bestanden} = 4 \}$ , Prädikat im Diplom-Zeugnis an der TUM (ordinal).
3.  $X_3 : \Omega_2 \rightarrow \mathbb{N}_0$ , Kinderzahl (diskrete Zählvariable, metrisch)
4.  $X_4 : \Omega_2 \rightarrow [0, \infty)$ , Auftragsvolumen des Kunden im Abrechnungsmonat Oktober 1997 (stetig)
5.  $X_5 : \Omega_3 \rightarrow (0, \infty)$ , Nettomiete/qm am 1.1.1997 (stetig)

Im ersten Stochastikkurs sei es erlaubt, dem Hörer den Unterschied zwischen der Abbildung  $X$  (Merkmal, Zufallsvariable) und deren Realisierung  $x = X(\omega)$  an einem weiteren konstruierten Beispiel aufzuzeigen. Vielleicht trägt die Tatsache, dass das zu beschreibende Vorgehen vom Gesetzgeber ausdrücklich verboten wurde, dazu bei, dass sich der darzustellende Sachverhalt besser einprägt.

### Beispiel 7.7 (Datenbank)

Gäbe es also eine bundesdeutsche Datenbank, in der für jeden Bürger über eine Personenkennzahl (wir nennen sie  $\omega$ ) alle Informationen abrufbar wären, z. B. auch Geschlecht, Schulabschluß, Beruf, Kinderzahl, Wahlverhalten, Einkünfte, Nebentätigkeiten, usw., so ließen sich interessante statistische Untersuchungen über die Bundesbürger folgendermaßen vereinfachen:

1. Lege das zu untersuchende Merkmal  $X$  fest; etwa Religionszugehörigkeit als eine gewisse Spalte im jeweiligen Personen-Datensatz.
2. Wähle aus der Datenbank (das ist  $\Omega$ ) eine Person  $\omega$  aus. Durch Auswerten der Position für Religionszugehörigkeit erhält man mit  $x := X(\omega)$  das Merkmal  $X$  im jeweiligen Datensatz  $\omega$ . Die Realisierung  $x \in \mathbb{R}$  ist eine Zahl im Gegensatz zu der Abbildung  $X : \Omega \rightarrow \mathbb{R}$ .
3. Aus zwei vorhandenen Merkmalen wie  $X = \text{Einkünfte}$  und  $Y = \text{Nebentätigkeiten}$  läßt sich ein neues Merkmal  $Z := X + Y$  Gesamteinkünfte konstruieren, indem jeder Datensatz um eine Spalte erweitert wird. Dort ist dann jeweils  $Z(\omega) := X(\omega) + Y(\omega)$  für alle  $\omega \in \Omega$  neu einzutragen, wobei  $X(\omega)$  und  $Y(\omega)$  bereits vorhanden sind.
4. Sei  $g : \mathbb{R} \rightarrow \mathbb{R}$  eine geeignet gewählte Funktion. Dann ist durch  $W := g(U)$  ein neues Merkmal definiert; z. B.  $U$  Einkommen in DM und  $W$  Einkommen umgerechnet in US \$.
5. Sind also Häufigkeiten, Mittelwerte, Maxima etc. für  $X, Y, U$  in den genannten Beispielen bekannt, so können die entsprechenden Werte auch für die transformierten Größen  $Z$  und  $W$  leicht angegeben werden.

### 7.1.2 Messreihen, Stichproben

Bei der Gewinnung statistischen Datenmaterials sind generell sog. **Vollerhebungen** wünschenswert. Dagegen sprechen meist Datenschutzvorschriften oder Kostengründe. Beispielsweise läßt sich eine Volkszählung nicht jedes Jahr durchführen. Auf der anderen Seite verbieten auch praktische Überlegungen eine vollständige Ausschöpfung der Grundgesamtheit: z. B. beim Prüfen der Lebensdauer von Glühbirnen. Deswegen sind in den meisten statistischen Untersuchungen sog. **Teilerhebungen** oder Stichproben notwendig. Bei der Datenerfassung hat man hier eine Reihe von Regeln zu beachten.

**Def. 7.8 (Empirische Stichprobe, Messreihe)**

Gegeben sei eine Population (Grundgesamtheit)  $\Omega$ , ein Merkmal  $X : \Omega \rightarrow \mathbb{R}$  und eine Teilmenge  $\{\omega_1, \omega_2, \dots, \omega_n\} \subset \Omega$ .

Dann heißt

$$x_1, x_2, \dots, x_n := X(\omega_1), X(\omega_2), \dots, X(\omega_n) \quad (7.5)$$

**empirische Stichprobe oder Messreihe oder Urliste oder nur 'die Daten'**. Die  $x_i$  sind in der Regel ungeordnet. Im Gegensatz zu Mengen können die Stichprobeneintragungen  $x_i$  und  $x_j$ ,  $i \neq j$  durchaus gleich sein.

Eine Stichprobe wird genannt:

1. **zufällig**, wenn jedes Element aus der Grundgesamtheit die gleiche Chance hat, in die Auswertung der Stichprobe bzgl.  $X$  zu gelangen
2. **repräsentativ**, wenn die der Stichprobe zugrundeliegenden ausgewählten Elemente alle Aspekte der Grundgesamtheit bzgl. des Merkmals  $X$  repräsentieren (z. B. das ganze Parteienspektrum bei einer Wahlprognose).

Natürlich können die Daten  $x_1, x_2, \dots, x_n$  auch Realisierungen von iid-Variablen  $X_1, X_2, \dots, X_n$  sein (iid = independent identically distributed); vgl. Def. 2.29. Die folgenden Überlegungen setzen dies jedoch nicht voraus.

## 7.2 Empirische Verteilung eindim. diskreter Merkmale

Hier sollen für verschiedene Merkmalstypen (relative) Häufigkeiten tabellarisch und grafisch dargestellt werden. Es liege eine empirische Stichprobe  $x_1, x_2, \dots, x_n$  bzgl. des Merkmals  $X$  in  $\Omega$  vor. Für diskrete Merkmale aller Skalentypen gibt es eine Reihe grafischer Darstellungsmöglichkeiten, die auch in vielen kommerziellen Programmen angewählt werden können. Die relative Summenhäufigkeitsfunktion ist bereits auf ordinale und metrische Merkmale eingeschränkt. Statistische Maßzahlen wie Mittelwert und Varianz sind nur für metrische Merkmale erklärt.

### 7.2.1 Tabellierung und grafische Darstellung

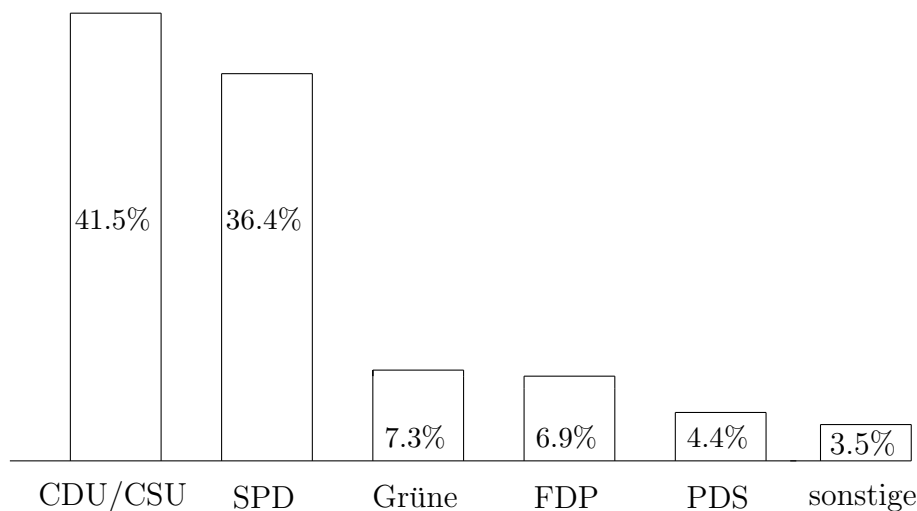
Zunächst soll ein nominales Merkmal untersucht werden.

**Beispiel 7.9 (Bundestagswahl 1994)**

Wir legen  $n = 50$  Mio. abgegebene gültige Stimmen zugrunde. Dann entspricht etwa das Säulendiagramm von Bild 14 folgender Häufigkeitstabelle

<b>Tabelle 4:</b> Absolute und relative Häufigkeiten						
Partei	CDU/CSU	SPD	Grüne	FDP	PDS	sonstige
absolute Hfkt. in Mio.	20.75	18.20	3.65	3.45	2.20	1.75
relative Häufigkeit	41.5%	36.4%	7.3%	6.9%	4.4%	3.5%

Für nominale Merkmale werden in der Regel nur absolute und relative Häufigkeiten angegeben. Eine grafische Darstellung erfolgt über verschiedenartige Diagramme.



**Bild 14: Säulendiagramm:** Ergebnisse der Bundestagswahl 1994

Daneben gibt es auch noch räumliche Säulendiagramme, Stabdiagramme, Kreissektordiagramme, 'Torten'-Diagramme, Balkendiagramme (waagrechte Säulen) und sog. Pictogramme, z. B. mit kleinen Autos für die Anzahl von Pkw-Zulassungen usw.

### 7.2.2 Relative Summenhäufigkeit

Wir wenden uns jetzt ordinalen und metrischen Merkmalen zu mit den Merkmalsausprägungen

$$a_1 < a_2 < \dots < a_j < \dots < a_m .$$

**Beispiel 7.10 (Kunden am Schalter)**

Über mehrere Wochen standen an verschiedenen Tagen jeweils bei der Kfz-Zulassungsstelle München um 12 Uhr (Ende der ffnungszeit) an geöffneten Schaltern folgende Anzahlen von Personen in der Schlange:

Anzahl Personen $a_j = j$	0	1	2	3	4	5	6	7	8	9	10
Häufigkeit $n_j$	0	0	0	5	10	15	20	50	100	25	25
rel. Häufigkeit $r_j = n_j/n$	0	0	0	$\frac{5}{250}$	$\frac{10}{250}$	$\frac{15}{250}$	$\frac{20}{250}$	$\frac{50}{250}$	$\frac{100}{250}$	$\frac{25}{250}$	$\frac{25}{250}$
kumulierte rel. Hfkt. $H_j := \sum_{k=1}^j r_k$	0	0	0	$\frac{5}{250}$	$\frac{15}{250}$	$\frac{30}{250}$	$\frac{50}{250}$	$\frac{100}{250}$	$\frac{200}{250}$	$\frac{225}{250}$	$\frac{250}{250}$

Einer übersichtlichen Darstellung wegen unterdrücken wir die Ausprägung  $a_0$ . Das ändert nichts an den folgenden Berechnungen, da  $a_0 = 0$  nicht als Stichprobenwert auftritt. Sodann setzen wir (wie schon in der Tabelle angegeben)

$$a_1 := 1, a_2 := 2, \dots, a_9 := 9, a_{10} = a_m := 10.$$

**Satz 7.11 (Regeln für absolute, relative und kumulierte Häufigkeiten)**

Mit den absoluten Häufigkeiten  $0 \leq n_j \leq n$ ,  $\sum n_j = n$  gilt für die relativen Häufigkeiten

$$0 \leq r_j = \frac{n_j}{n} \leq 1, \quad \sum_{j=1}^m r_j = 1. \quad (7.6)$$

Daraus folgt für die kumulierten (relativen) Häufigkeiten

$$0 \leq H_1 = r_1 \leq H_2 = r_1 + r_2 \leq \dots \leq H_m = 1. \quad (7.7)$$

$H_6 = \frac{50}{250} = 0.2$  im Beispiel oben besagt, dass in 20% der Fälle sechs oder weniger Kunden um 12 Uhr anstanden.

Dies motiviert folgende Festlegung

**Def. 7.12 (Empirische Verteilungsfunktion)**

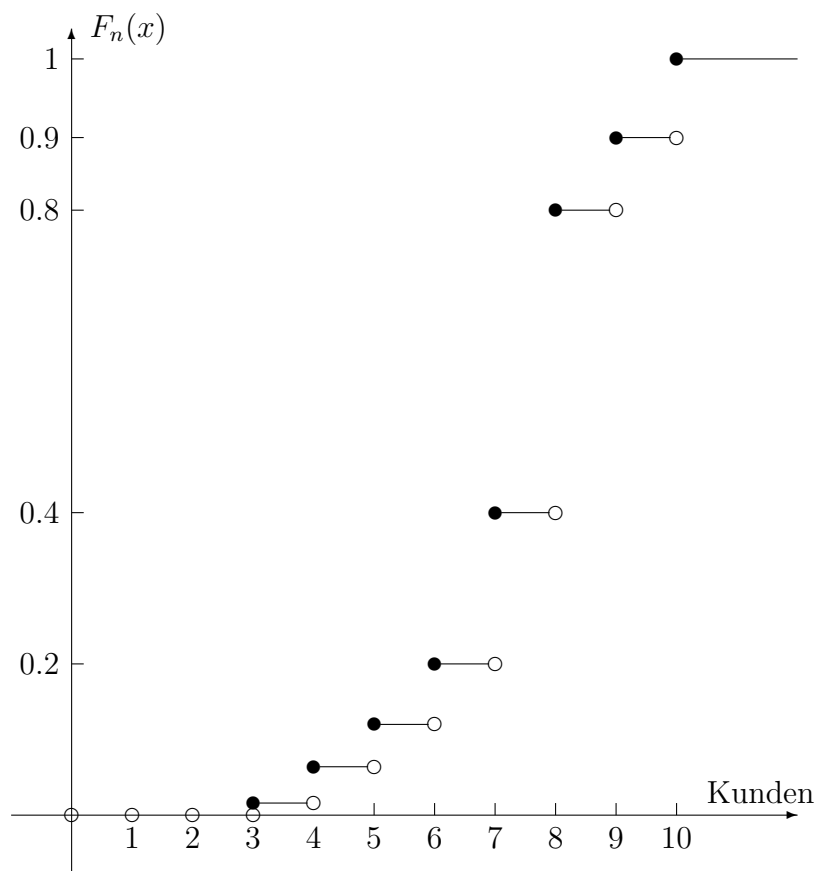
Die Funktion  $F_n : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$F_n(x) := \frac{|\{x_i, i = 1, \dots, n : x_i \leq x\}|}{n} = \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, x]}(x_i), \quad (7.8)$$

heißt **empirische Verteilungsfunktion** oder auch **relative Summenhäufigkeitsfunktion**.  $1_A$  bezeichnet die Indikatorfunktion der Menge  $A$ , also  $1_A(x) = 1$ , falls  $x \in A$  und  $1_A(x) = 0$ , sonst (siehe auch Def. 1.44).

Es gilt

$$F_n(x) = \begin{cases} 0 & , \text{ falls } x < a_1 \\ H_j = r_1 + \dots + r_j & , \text{ falls } a_j \leq x < a_{j+1} \\ 1 & , \text{ falls } x \geq a_m. \end{cases} \quad (7.9)$$



**Bild 15: Empirische Verteilungsfunktion: Kunden**

**Satz 7.13 (Eigenschaften)**

Die empirische Verteilungsfunktion ist für ordinale und metrische Merkmale definiert. Für sie gilt

1.  $F_n$  ist monoton nicht fallend
2.  $\lim_{x \rightarrow -\infty} F_n(x) = 0$ ,  $\lim_{x \rightarrow \infty} F_n(x) = 1$
3.  $F_n$  ist eine rechtsseitig stetige Treppenfunktion.

## 7.3 Empirische Verteilung eindim. stetiger Merkmale

Ein Merkmal  $X : \Omega \rightarrow \mathbb{R}$  heißt **stetig**, falls  $X(\omega)$  alle Werte aus einem Intervall, z. B.

$$[a, b] = \{x \mid a \leq x \leq b\}, \quad a < b$$

- oder wie in Gleichung (7.4) beschrieben - annehmen kann.  $x_1, \dots, x_n$  sei wieder eine empirische Stichprobe (Urliste).

### Beispiel 7.14 (Körpergewichte von Schülern einer 8. Klasse)

Tabelle 6: Urliste Körpergewichte						
57.8	61.3	60.1	60.0	63.7	60.5	64.8
58.1	65.9	61.1	63.2	56.2	64.4	61.9
59.7	60.2	61.3	65.6	63.2	58.7	62.8

Ziel: Die gesammelte Information möglichst kurz und übersichtlich darstellen. Die Rolle der diskreten Merkmalsausprägungen übernehmen nun disjunkte Klassen, die alle Stichprobenwerte überdecken.

### Regeln zur Klassenbildung

1.  $x_{\min} = \min \{x_i\}$ ,  $x_{\max} = \max \{x_i\}$  :

bestimme Zahlen  $a_0 (< x_{\min}) < a_1 < \dots < a_m (\geq x_{\max})$

und setze  $I_j = (a_{j-1}, a_j] = \{x \mid a_{j-1} < x \leq a_j\}$ ,  $j = 1, \dots, m$ .

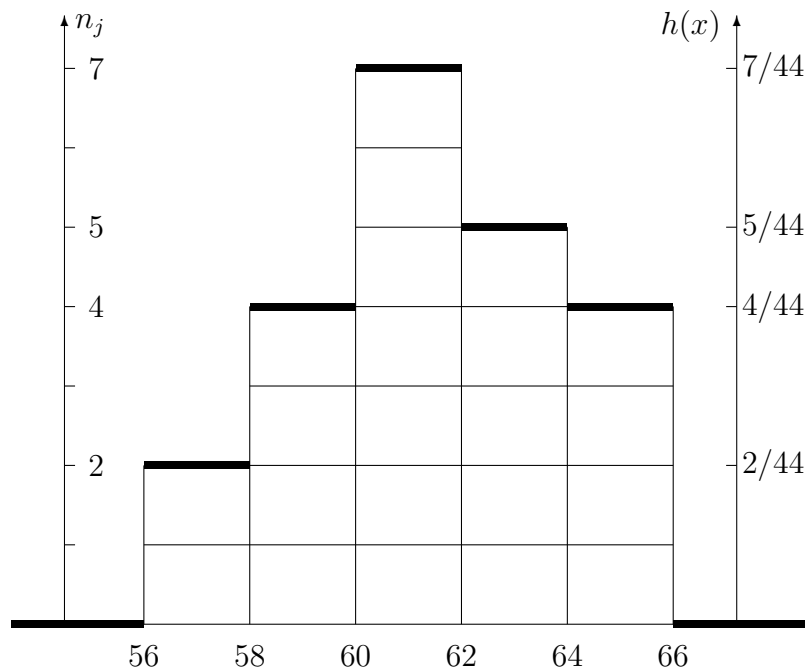
$a_0$  heißt **Reduktionslage**,  $v = x_{\max} - x_{\min}$  bezeichnet die **Variationsbreite**.



2.  $a_j$  (falls möglich) durch praktische Überlegungen im Zusammenhang mit der Problemstellung wählen.
3.  $m \approx \sqrt{n}$  (im Zweifel ungerade),  $5 \leq m \leq 25$  (Faustregel);
4. i. a. **Klassenbreiten**  $w_j = a_j - a_{j-1} = w = \text{const.}$   
 $a_j$  möglichst "einfache Zahlen".
5. Manchmal bietet sich ein **logarithmischer Maßstab** oder eine andere geeignete Transformation an.
6. Die Klasseneinteilung ist nicht eindeutig vorgeschrieben.
7. Alle Größen dieses Abschnitts hängen von der gewählten Klasseneinteilung  $I := I_1, \dots, I_m$  ab.

Im Beispiel gilt  $x_{\min} = 56.2$  und  $x_{\max} = 65.9$ .

### Histogramm und empirische Verteilungsfunktion



**Bild 16:** Histogramm für die oben gewählte Klasseneinteilung

Ein **Histogramm** veranschaulicht die Dichte der Daten. Bei geeigneter Skalierung (im Bild die rechte Ordinate) ist das Histogramm der Graph der **empirischen Dichtefunktion**

$$h(x) = h_I(x) := \sum_{j=1}^n \frac{n_j}{nw_j} 1_{I_j}(x) = \begin{cases} \frac{n_j}{nw_j} & , x \in I_j \\ 0 & , \text{sonst} \end{cases} \quad (7.10)$$

$n_j/n$  Rel. Häufigkeit der Klasse  $I_j$   
 $w_j$  Klassenbreite

**Satz 7.15 (Empirische Dichtefunktion, Histogramm)**

Es gilt

1.  $h \geq 0$
2.  $\int_{-\infty}^{\infty} h(x)dx = 1.$

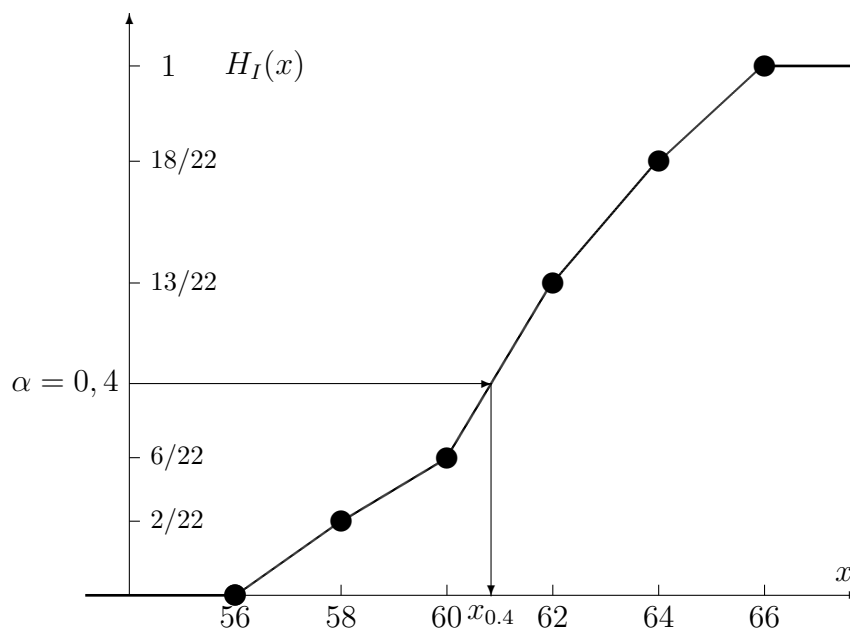
$(a_{j-1}, a_j]$	$(a_0, a_1]$	$(a_1, a_2]$	$(a_2, a_3]$	$(a_3, a_4]$	$(a_4, a_5]$
Klasse	(56,58]	(58,60]	(60,62]	(62,64]	(64,66]
$n_j$	2	4	7	5	4
$n_j/n$	2/22	4/22	7/22	5/22	4/22
kumul.rel.Hfkt.	2/22	6/22	13/22	18/22	22/22

Bei stetigen Merkmalen wird oft die integrale **empirische Verteilungsfunktion**  $H_I(x)$  bzgl. der gewählten Klasseneinteilung herangezogen. Man definiert:

$$H_I(x) := \int_{-\infty}^x h(t)dt. \quad (7.11)$$

Daraus ergeben sich **Nichtnegativität** bzw. **Monotonie**

$$0 \leq H_I(x) \leq 1, \quad H_I(x) \leq H_I(y), \quad \text{falls } x \leq y. \quad (7.12)$$



**Bild 17:** Stetige empirische Verteilungsfunktion bzgl. Klasseneinteilung

An den Klassenenden  $a_j$  gilt  $H_I(a_j) = (n_1 + n_2 + \dots + n_j)/n$ .

Damit stimmt  $H_I$  dort mit der unabhängig von der Klasseneinteilung definierten, stückweise konstanten **relativen Summenhäufigkeitsfunktion**

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, x]}(x_i), \quad x \in \mathbb{R},$$

überein; vgl. Abschnitt 7.2.2. Weiterhin gilt

$$\lim_{x \rightarrow -\infty} F_n(x) = \lim_{x \rightarrow -\infty} H_I(x) = 0, \quad \lim_{x \rightarrow \infty} F_n(x) = \lim_{x \rightarrow \infty} H_I(x) = 1, \quad (7.13)$$

und beide Funktionen sind monoton wachsend. Bei diskreten Merkmalen verwendet man nur  $H(x)$ , das höchstens an den Merkmalsausprägungen  $a_j$  Sprünge aufweist.

Näherungen für empirische  $\alpha$ -Quantile (Prozentpunkte), vgl. Abschnitt 7.4.2, können grafisch aus Bild 17 bestimmt werden, indem man die Gerade  $y = \alpha$  mit dem Graphen von  $H_I$  schneidet.  $x_\alpha$  ist die  $x$ -Koordinate des Schnittpunkts.

Stichproben werden häufig durch Parameter, d. h. statistische Maßzahlen, charakterisiert. Wir erläutern die wichtigsten **Lage-** und **Streuungsparameter**.

Fast alle der zu behandelnden Größen haben ihre Entsprechung als Kenngrößen von Verteilungen. Zur Unterscheidung verwenden wir in den folgenden Abschnitten jeweils bei der Definition einer Maßzahl Bezeichnungen wie *empirisches* Quantil oder *empirische* Varianz etc. Da in diesem Kapitel aber keine Verwechslungen mit den theoretischen Kenngrößen der Verteilungen, etwa von Def. 1.50, 1.59, 1.63 usw. zu befürchten sind, wird der Zusatz *empirisch* in der Folge meist wieder weggelassen.

## 7.4 Lageparameter, Lokationsmaße

Diese beschreiben die ‐Lage‐ der Stichprobe auf der Zahlengeraden.

### 7.4.1 Modus (Modalwert)

#### Def. 7.16 (Modus, Modalwert)

Der **Modus** oder **Modalwert**  $x_{\text{Modus}}$  ist die am häufigsten auftretende Merkmalsausprägung. Er wird hauptsächlich für nominale Merkmale verwendet, ist aber auch für alle anderen (diskreten) Merkmalstypen sinnvoll. Bei stetigen Merkmalen hängt die Angabe des Modalwerts von der Form der Dichteschätzung ab (vgl. Klasseneinteilung, Kerndichteschätzer).

**Beispiel 7.17 (Modalwerte)**

1. Beim Wahlverhalten von Beispiel 7.9 wäre der Modus (Modalwert)  $x_{Modus} = a_1$ , dem "CDU/CSU" zugeordneten Wert.
2. Bei stetigen Merkmalen nimmt man als Modalwert die Mitte der Klasse mit den größten Häufigkeiten. Also in Beispiel 7.14

$$x_{Modus} = 61 = \frac{a_2 + a_3}{2}.$$

Ab jetzt betrachten wir nur mehr ordinale oder metrische Merkmale.  $x_{1:n}, \dots, x_{n:n}$  bezeichne die geordnete Messreihe; also  $x_{1:1} \leq x_{1:2} \leq \dots \leq x_{n:n}$ . Gelegentlich schreibt man für die geordnete Stichprobe auch  $x_{(1)}, \dots, x_{(n)}$ .

Weiterhin sei  $[y]$  die größte ganze Zahl kleiner oder gleich  $y \in \mathbb{R}$ .

**7.4.2 Empirische Quantile, Quartile, Median****Def. 7.18 (Quantile)**

Seien  $0 < \alpha < 1$ ,  $1 \leq q \leq n - 1$  und  $0 \leq r < 1$ , so dass

$$n \cdot \alpha = q + r, \quad q = [n \cdot \alpha].$$

Für ordinale und metrische Merkmale ist das sog.  $\alpha$ -Quantil  $x_\alpha$  definiert als eine Zahl, für die mindestens  $\alpha \cdot 100\%$  der Daten kleiner oder gleich  $x_\alpha$  und mindestens  $(1 - \alpha) \cdot 100\%$  der Daten größer oder gleich  $x_\alpha$  sind, wobei  $x_\alpha$  selbst gegebenenfalls mitgezählt wird. Eine stets eindeutige Festlegung erreicht man durch

$$x_\alpha := \begin{cases} x_{(q+1)} & , \text{ falls } r > 0 \\ \frac{x_{(q)} + x_{(q+1)}}{2} & , \text{ falls } r = 0. \end{cases} \quad (7.14)$$

Der Vollständigkeit halber wird noch  $x_{0.0} := x_{(1)} = x_{min}$  und  $x_{1.0} := x_{(n)} = x_{max}$  festgelegt.

Beispielsweise erhält man für  $x_1, \dots, x_n = 1, 2, 3, \dots, 19, 20$  mit  $n = 20$

$$x_{0.0} = 1, x_{0.25} = 5.5, x_{0.5} = 10.5, x_{0.75} = 15.5, x_{1.0} = 20,$$

oder für  $x_1, \dots, x_n = 1, 2, 3, \dots, 12, 13$  mit  $n = 13$

$$Q_0 := x_{0.0} = 1, Q_1 := x_{0.25} = 4, Q_2 := x_{0.5} = 7, Q_3 := x_{0.75} = 10, Q_4 := x_{1.0} = 13.$$

Die speziellen Quantile  $Q_0, Q_1, Q_2, Q_3, Q_4$  heißen empirische **Quartile**. Unter diesen nennt man  $Q_1$  **erstes** bzw.  $Q_3$  **drittes** Quartil. Von besonderem Interesse ist der Median  $Q_2$ .

## Median (Zentralwert)

### Def. 7.19 (Median)

Für ordinale und metrische Merkmale ist der empirische **Median** oder **Zentralwert** definiert als

$$\text{med}(x_1, \dots, x_n) := x_{0.5} := \begin{cases} x_{(\frac{n+1}{2})} & , \text{ falls } n \text{ ungerade} \\ \frac{x_{(n/2)} + x_{(n/2+1)}}{2} & , \text{ falls } n \text{ gerade.} \end{cases} \quad (7.15)$$

Mindestens 50% der Stichprobenwerte liegen also links und mindestens 50% rechts des Medians (den Median selbst ggf. mit eingerechnet).

### Beispiel 7.20 (Median)

Bei den Körpergewichten von Beispiel 7.14 ergibt sich

$$x_{0.5} = (x_{(11)} + x_{(12)})/2 = (61.1 + 61.3)/2 = 61.2.$$

$\text{med}(x_1, \dots, x_n)$  ist ein sehr robustes Lokationsmaß. Robuste statistische Kenngrößen sind wenig anfällig gegen Datenausreißer. Man muss die Hälfte der Daten gegen  $\infty$  oder  $-\infty$  verschieben, um den Median selbst gegen  $\pm\infty$  wandern zu lassen. Zur quantitativen Untersuchung dieses Phänomens definiert man den sog.

### Def. 7.21 (Breakdownpoint $\epsilon$ )

Es gilt

$$\epsilon(\text{med}(x_1, \dots, x_n)) = \begin{cases} \frac{k+1}{n} = \frac{1}{2} + \frac{1}{2n} & , \quad n = 2k+1 \\ \frac{k}{n} = \frac{1}{2} & , \quad n = 2k \end{cases} \xrightarrow{n \rightarrow \infty} \frac{1}{2}. \quad (7.16)$$

### 7.4.3 Arithmetisches Mittel

**Def. 7.22 (Arithmetisches Mittel)**

Falls nichts anderes gesagt ist, bezeichnet man mit **Mittelwert** das **arithmetische Mittel** (*mean*) oder den **Durchschnitt**

$$\bar{x} = \bar{x}_n := \frac{1}{n} \sum_{i=1}^n x_i. \quad (7.17)$$

$\bar{x}$  ist sinnvoll für beliebige metrische Merkmale.

$\bar{x}$  ist im Gegensatz zum empirischen Median sehr anfällig gegen sog. Datenausreißer. Mit dem Breakdownpoint  $\epsilon(\bar{x}) = 1/n$  genügt das Abdriften eines einzigen Datenpunktes, um  $\bar{x}$  gegen  $\pm\infty$  streben zu lassen.

## 7.5 Streuungsparameter

Diese geben Auskunft, ob die Stichprobenwerte eng um den Mittelwert konzentriert sind oder weit “streuen”. In der Regel werden Streuungsparameter nur für metrische Merkmale angegeben. Formal sind sie zwar auch für ordinale Merkmale definiert, hängen aber von der dort nicht eindeutig festgelegten Kodierung der Merkmalsausprägungen ab.

### 7.5.1 Spannweite, Variationsbreite

**Def. 7.23 (Spannweite, Variationsbreite)**

Für ordinale und metrische Merkmale nennt man die **Ausdehnung der Stichprobenwerte**

$$v = x_{(n)} - x_{(1)} = x_{max} - x_{min} \quad (7.18)$$

**Spannweite** oder **Variationsbreite** (*range*).

**Beispiel 7.24 (Variationsbreite)**

Körpergewichte von Beispiel 7.14:

$$v = x_{max} - x_{min} = 65.9 - 56.2 = 9.7.$$

## 7.5.2 Mittlere quadratische Abweichung, Standardabweichung

### Def. 7.25 (Empirische Varianz, Standardabweichung)

Das am meisten verwendete Streuungsmaß ist die mittlere quadratische Abweichung

$$s^2 = s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (7.19)$$

$s^2$  heißt auch empirische **Varianz** (variance).  $s_x^2$  ist zunächst eine abstrakte Größe. Anschaulicher ist die empirische **Standardabweichung** (standard deviation)

$$s = s_x = \sqrt{s_x^2} \geq 0. \quad (7.20)$$

$s$  hat im Gegensatz zu  $s^2$  dieselbe Benennung wie  $\bar{x}$ ; also [m] oder [sec] etc.

Wie bei  $\bar{x}$  sind die Breakdownpoints  $\epsilon(s^2) = \epsilon(s) = 1/n$ .

Die Darstellung  $\sum (x_i - \bar{x})^2 / (n - 1)$  eignet sich besser zur numerisch stabilen Berechnung von  $s^2$ . Zu Übungszwecken mit einfachen Daten ist immer noch die sog. **Verschiebungsregel**

$$s^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n \bar{x}^2 \right) \quad (7.21)$$

in Gebrauch. Die Gleichheit gilt natürlich bei exakter Rechnung. Durch Auslöschung entstehen aber bei endlicher Arithmetik durch Anwendung von (7.21) unakzeptabel große Rundungsfehler. Es lassen sich leicht Beispiele konstruieren, bei denen die Verschiebungsregel mit dreistelliger Rechnung negative Varianzen liefert.

### Beispiel 7.26

Sei  $x_1, \dots, x_n = 6, 4, 7, 3, 7$ , also  $n = 5$  und  $\bar{x} = 5.4$ .

$$\begin{aligned} s^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 3.3 \\ s &= 1.817 \quad (\text{Standardabweichung}) \end{aligned} \quad (7.22)$$

Aus numerischen Gründen sollte stets mit der im Beispiel verwendeten Formel gerechnet werden.

### Bemerkung 7.27 (Bedeutung der Standardabweichung)

Bei normalverteilten Daten liegen ca. 95% der Daten im Intervall  $[\bar{x} - 2s, \bar{x} + 2s]$ .

$\bar{x}$  und  $s^2$  hängen eng zusammen. Es gilt nämlich

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \leq \frac{1}{n-1} \sum_{i=1}^n (x_i - \lambda)^2 \quad \forall \lambda \in \mathbb{R}. \quad (7.23)$$

Gelegentlich arbeitet man auch mit

$$\tilde{s}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2. \quad (7.24)$$

Die Wahl von  $s$  ist jedoch besser, da für iid-Variable  $X, X_1, \dots, X_n$  mit  $\text{Var}(X) = \sigma^2$  gemäß Beispiel 2.34 mit  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  gilt:

$$E \left( \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \right) = \sigma^2. \quad (7.25)$$

### 7.5.3 Mittlere lineare Streuung

Ein ähnlicher Zusammenhang wie für  $\bar{x}$  und  $s^2$  besteht zwischen dem Median und folgendem Streuungsmaß.

**Def. 7.28 (Mittlere lineare Streuung)**

$x_{0.5}$  bezeichne den Median. Dann nennt man für ordinale und metrische Merkmale

$$s_L = \frac{1}{n} \sum_{i=1}^n |x_i - x_{0.5}| \quad (7.26)$$

die **mittlere lineare Streuung**.

Die lineare Streuung ist nicht so anfällig gegen extreme Datenausreißer wie  $s^2$  oder die Standardabweichung  $s$ .

**Bemerkung 7.29 (Optimalität des Medians)**

$$s_L = \frac{1}{n} \sum_{i=1}^n |x_i - x_{0.5}| \leq \frac{1}{n} \sum_{i=1}^n |x_i - \lambda| \quad \forall \lambda \in \mathbb{R}. \quad (7.27)$$

### 7.5.4 Interquartilsabstand, inter quartile range (IQR)

Noch robuster als die mittlere lineare Streuung ist ein Streuungsmaß, das durch den Abstand zwischen dem ersten Quartil  $Q_1$  und dem dritten  $Q_3$  (vgl. Def. 7.18) erklärt ist.

**Def. 7.30 (Empirischer Interquartilsabstand, IQR)**

Ein weiteres Maß für die Streuung der Daten ist

$$IQR = Q_3 - Q_1. \quad (7.28)$$



**Beispiel 7.31 (Ausreißer)**

Wir betrachten  $x_1, \dots, x_n = 1, 7, 9, 11, 13, 20, 79$  mit  $n = 7$ . Es gilt

$Q_1 = 7$ ,  $Q_2 = x_{0.5} = 11$ ,  $Q_3 = 20$ ,  $IQR = 13$ ,  $s_L = 13.57$  (mittlere lineare Streuung)

$$\bar{x} = 20, \quad s^2 = 608.86, \quad s = 24.68$$

$$v = x_{max} - x_{min} = 78.$$

Läßt man die jeweils kleinste und größte Beobachtung als mögliche Ausreißer weg (**gestutztes** oder **getrimmtes** Mittel), so ergibt sich

$$\bar{x}_{2..6} = 12,$$

$$s_{2..6}^2 = \frac{1}{5} \sum_{i=2}^6 (x_i - \bar{x}_{2..6})^2 = 20, \quad s_{2..6} = 4.47,$$

$$s_L^{2..6} = \frac{1}{5} \sum_{i=2}^6 |x_i - x_{0.5}| = 3.4,$$

$$IQR^{2..6} = Q_3^{2..6} - Q_1^{2..6} = 13 - 9 = 4.$$

Arithmetisches Mittel und  $s^2$  sind also sehr anfällig gegen extreme Beobachtungen (im Beispiel  $x_7 = 79$ ). Hätten wir diesen Wert **zufällig nicht erhoben**, so ergäbe sich ein Mittelwert von  $\frac{1}{6} \sum_{i=1}^6 x_i = 10.17$ .

Eine annähernd gleich realistische Beschreibung der mittleren Lage liefern aber sowohl Median als auch das gestutzte Mittel (ohne kleinsten und größten Stichprobenwert) bereits aus der ursprünglichen Stichprobe mit der Beobachtung  $x_7 = 79$ .

Man sollte also zunächst  $\bar{x}$ , Median, gestutzte Mittel sowie die Standardabweichung  $s$  und die lineare Streuung  $s_L$  berechnen. Stimmen diese Werte einigermaßen überein, so kann man bedenkenlos mit  $\bar{x}$  und  $s$  bzw.  $s^2$  weiterarbeiten. Andernfalls ist eine evtl. Eingabedatei auf Tippfehler bzw. die Stichprobe auf sog. **Ausreißer** zu überprüfen.

**7.5.5 Variationskoeffizient**

Die folgende Größe dient dem Vergleich der Streuung zweier verschiedener Grundgesamtheiten.

**Def. 7.32 (Variationskoeffizient, coefficient of variation, CV)**

*Der empirische Variationskoeffizient*

$$CV := \frac{s_x}{\bar{x}}, \quad \bar{x} \neq 0, \quad (7.29)$$

*ist eine dimensionslose Größe.*

Beispielsweise seien für die monatlichen Einkommen im Jahre 1990 von Studenten einer bestimmten Bevölkerungsschicht bekannt:

BRD	$\bar{x} = 720 \text{ DM}$	$s_x = 180 \text{ DM}$	$CV_x = \frac{180}{720} = 0.25$
USA	$\bar{y} = 480 \text{ \$}$	$s_y = 160 \text{ \$}$	$CV_y = \frac{160}{480} = 0.33$

Die Studenteneinkommen streuen also in den USA mehr als in der BRD.

## 7.6 Schiefe und Exzess

Hierbei handelt es sich um sog. höhere Momente.  $s$  bezeichne die oben definierte Standardabweichung.

**Def. 7.33 (Schiefe, skewness)**

$$b_3 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3} \quad (7.30)$$

heißt empirische **Schiefe**. Im Falle  $b_3 < 0$  spricht man von **linksschiefen** und für  $b_3 > 0$  von **rechtsschiefen** Daten.

**Bemerkung 7.34 (Interpretation)**

1. Die Schiefe ist ein standardisiertes Maß für die Symmetrie der Stichprobenwerte zum Mittelwert.
2. Für normalverteilte Daten ist  $b_3$  ungefähr gleich Null.

**Def. 7.35 (Exzess, kurtosis)**

$$b_4 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{s^4} - 3 \quad (7.31)$$

heißt empirischer **Exzess**.

**Bemerkung 7.36 (Interpretation)**

1. Für normalverteilte Daten ist  $b_4$  ungefähr gleich Null.
2. Der Exzess kann u.a. als ein standardisiertes Maß für die Abweichung der Stichprobenwerte von normalverteilten Daten betrachtet werden.
3.  $b_4$  ist eine äußerst sensitive Größe und darf niemals allein zur Bewertung der Normalität von Daten herangezogen werden.

## 7.7 Mittelwerte, Streuungen für gruppierte Daten

Dieser Abschnitt enthält eher triviale Umformungen. Die Resultate sind trotzdem bemerkenswert, weil hier die Motivation für die Festlegung von Erwartungswerten bei diskreten Zufallsvariablen zu finden ist.

Nehmen wir an im Beispiel 7.10 der Kunden am Schalter der Kfz-Zulassungsstelle wären nicht die Rohdaten, sondern nur die relativen Häufigkeiten bekannt. Also

Tabelle 8: Schalterkunden											
Anzahl Personen $a_j = j$	0	1	2	3	4	5	6	7	8	9	10
rel. Häufigkeit $r_j = n_j/n$	0	0	0	$\frac{5}{250}$	$\frac{10}{250}$	$\frac{15}{250}$	$\frac{20}{250}$	$\frac{50}{250}$	$\frac{100}{250}$	$\frac{25}{250}$	$\frac{25}{250}$

Wie viele Personen standen durchschnittlich in der Schlange?

Wird wieder die Ausprägung  $a_0$  unterdrückt und

$$a_1 := 1, a_2 := 2, \dots, a_9 := 9, a_{10} = a_m := 10$$

gesetzt, dann kann das arithmetische Mittel direkt aus den relativen Häufigkeiten  $r_j = \frac{n_j}{n}$  berechnet werden:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \left( \underbrace{a_1 + \dots + a_1}_{n_1 \text{ Mal}} + \dots + \underbrace{a_m + \dots + a_m}_{n_m \text{ Mal}} \right) = \sum_{j=1}^m \frac{n_j}{n} \cdot a_j = \sum_{j=1}^m r_j \cdot a_j = 7.5.$$

### Satz 7.37 (Mittelwerte und Streuungen für gruppierte Daten)

Das Merkmal  $X$  sei metrisch. Für eine empirische Stichprobe  $x_1, \dots, x_n$  mit den Merkmalsausprägungen  $a_1 < \dots < a_m$  und den absoluten Häufigkeiten  $n_1, \dots, n_m$  (gruppierte Daten) gilt

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \sum_{j=1}^m \frac{n_j}{n} \cdot a_j = \sum_{j=1}^m r_j \cdot a_j. \quad (7.32)$$

Analog erhält man für die (leicht modifizierte) empirische Varianz

$$\tilde{s}^2 := \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{j=1}^m \frac{n_j}{n} (a_j - \bar{x})^2 = \sum_{j=1}^m r_j \cdot (a_j - \bar{x})^2. \quad (7.33)$$

Für die Schalterkunden ergibt sich  $\tilde{s}^2 = 2.45$  mit der Standardabweichung  $\tilde{s} = 1.57$ .

Ersetzt man die relativen Häufigkeiten durch Wahrscheinlichkeiten, so kommt man zu den bekannten Erwartungswertformeln für diskrete Zufallsvariable, wenn  $\tilde{s}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$  anstelle von  $s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$  verwendet wird.

## 7.8 Affine Transformationen

Es bleibt noch die Auswirkung linearer Transformationen der  $x_i$  auf die angegebenen statistischen Kenngrößen zu untersuchen.

### Satz 7.38 (Transformationen)

Für die empirische Stichprobe  $x_1, \dots, x_n$  bezeichne  $x_{0.5}$  den Median,  $\bar{x}$  das arithmetische Mittel,  $s_x$  die Standardabweichung und  $s_L^{(x)}$  die lineare Streuung. Ferner seien  $d \neq 0$ ,  $c$  beliebig. Durch eine affine Transformation

$$y_i := c + d \cdot x_i \quad (7.34)$$

entsteht eine neue empirische Stichprobe  $y_1, \dots, y_n$ . Für diese lauten die entsprechend mit 'y' bezeichneten Größen

$$\bar{y} = c + d \cdot \bar{x} \quad (7.35)$$

$$y_{0.5} = c + d \cdot x_{0.5} \quad (7.36)$$

$$s_y^2 = d^2 s_x^2 \quad (7.37)$$

$$s_y = |d| s_x \quad (7.38)$$

$$s_L^{(y)} = |d| s_L^{(x)}. \quad (7.39)$$

### Beispiel 7.39 (Transformation: Celsius nach Fahrenheit)

Die mittleren Temperaturen einer Tropenregion seien in der Stichprobe  $x_1, \dots, x_n$  in  $^{\circ}C$  (Celsius) erhoben mit  $\bar{x} = 27^{\circ}$  und  $s_x = 3^{\circ}C$ . Die entsprechenden Größen sollen in  $^{\circ}F$  (Fahrenheit) umgerechnet werden. Mit der Festlegung

$$0^{\circ}C = 33\frac{1}{3}^{\circ}F \quad \text{und} \quad 36^{\circ}C = 100^{\circ}F \quad (7.40)$$

lautet die Transformation

$$y_i := 33\frac{1}{3} + \frac{100}{54} \cdot x_i. \quad (7.41)$$

Also

$$\bar{y} = 33\frac{1}{3} + \frac{100}{54} \bar{x} = 33\frac{1}{3} + \frac{100}{54} \cdot 27 = 83\frac{1}{3}^{\circ}F$$

$$s_y = |d| s_x = \frac{100}{54} \cdot 3 = \frac{100}{18} = 5.56^{\circ}F.$$

## 7.9 Empirische Regression

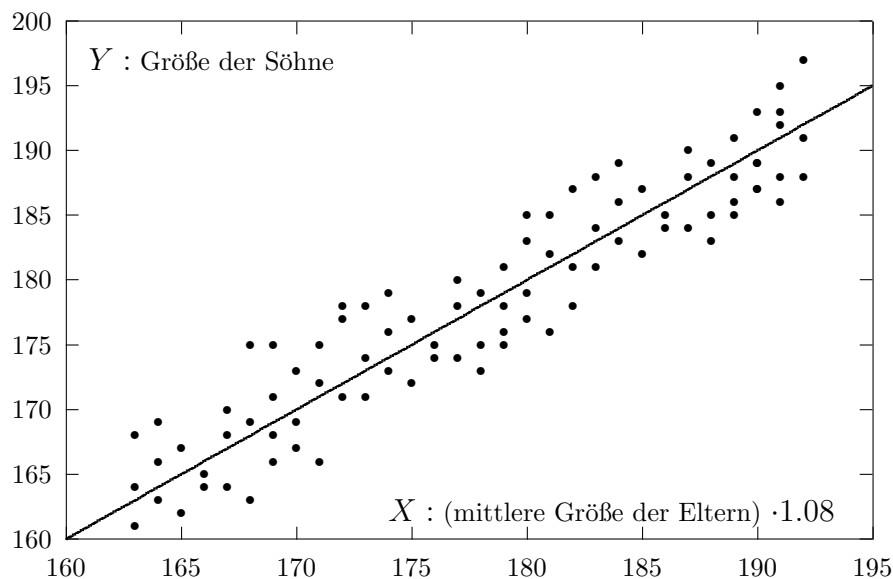
Man untersucht den Zusammenhang mehrerer Merkmale; z. B.

$Y$	$X_1$	$X_2$	...	$X_p$
Umsatz	Werbung	Vertreterbesuche	...	Preis
Nettomiete	Wohnfläche	Baualter	...	Ausstattung

Oft möchte man das  $Y$ -Merkmal aus den (leichter zu ermittelnden oder besser kontrollierbaren)  $X$ -Merkmalen vorhersagen. Wir beschränken uns hier auf den Fall  $p = 1$ , die **einfache lineare Regression**.

### 7.9.1 Statistische Maßzahlen

$X$  und  $Y$  seien Merkmale auf der Grundgesamtheit  $\Omega$  und  $(x_1, y_1), \dots, (x_n, y_n)$  eine Stichprobe. Zunächst veranschaulicht man sich den Zusammenhang grafisch in einer **Punktewolke (Streudiagramm, Scatterplot)**. Eine der ersten Regressionsanalysen wurde von F. Galton (1886) durchgeführt; vgl. Bild 18.



**Bild 18: Galton (1886):** Größe der Söhne und mittlere Größe der Eltern

Als statistische Maßzahlen gehen in die folgenden Berechnungen ein:

- die **Mittelwerte**  $\bar{x}$ ,  $\bar{y}$
- die **emp. Kovarianz**

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left( \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right) \quad (7.42)$$

- die **emp. Varianzen (Streuungen)**

$$s_x^2 = s_{xx} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n \bar{x}^2 \right). \quad (7.43)$$

Eine analoge Formel gilt für  $s_y^2 = s_{yy}$ .

Ein Maß für den **linearen** Zusammenhang der Merkmale  $X$  mit Daten  $x_i$  und  $Y$  mit Daten  $y_i$  ist der **empirische Korrelationskoeffizient**

$$\begin{aligned}
 r_{xy} &= \frac{s_{xy}}{s_x s_y} & (7.44) \\
 &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right) \left(\sum_{i=1}^n (y_i - \bar{y})^2\right)}} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n \bar{x}^2\right) \left(\sum_{i=1}^n y_i^2 - n \bar{y}^2\right)}}.
 \end{aligned}$$

Für alle Datensätze mit nichtkonstanten  $x_i$  bzw.  $y_i$  folgt

$$-1 \leq r_{xy} \leq 1. \quad (7.45)$$

Es stellt sich heraus, dass  $r_{xy} = \pm 1$  genau dann gilt, wenn alle Punkte exakt auf einer Geraden liegen. Bis hierher waren die den Daten zugrundeliegenden Merkmale  $X$  und  $Y$  völlig gleichberechtigt.

### 7.9.2 KQ-Gerade, Bestimmtheitsmaß

In der Praxis ist eines der Merkmale (hier  $X$ ) oft leichter zu beobachten oder zu kontrollieren als das andere. Beispielsweise variiert der Dollarkurs  $Y$  abhängig vom zeitlichen Verlauf  $X$ . Numeriert man die Tage eines Jahres etwa von 1 bis 365 durch und ist der Kurs  $Y$  am Tag  $X = 267$  bekannt, so kann der "Wert" der Zeit am darauffolgenden Tag zweifelsfrei mit  $X = 268$  festgelegt werden.

Wir besprechen nun eine häufig angewendete Prognosemethode für das schwerer zu vorhersagende Merkmal  $Y$ ; im Beispiel Dollarkurs.

Diese sog. "**Regressionsmethode**" beruht auf einem linearen Ansatz der Form

$$Y = \alpha + \beta X + E,$$

wobei  $E$  ein nichtbeobachtbarer Fehler ist. Die unbekannt Parameter  $\alpha$  und  $\beta$  schätzt man nach der **Methode der kleinsten Quadrate (KQ-Methode)**. Die Schätzungen  $a$  bzw.  $b$  der Parameter  $\alpha$  bzw.  $\beta$  gewinnen wir, indem eine Gerade

$$y = a + b x$$

in gewissem Sinne **optimal** durch die Punktwolke der  $(x_i, y_i)_{i=1, \dots, n}$  gelegt wird.

Wählt man einen Achsenabschnitt  $a$  und eine Steigung  $b$  für die Gerade, dann liegt mit  $\hat{y}_i := a + b x_i$  der Punkt  $(x_i, \hat{y}_i)$  genau auf der Geraden, und zwar vertikal über  $x_i$ ; vgl. nachfolgendes Bild 19.

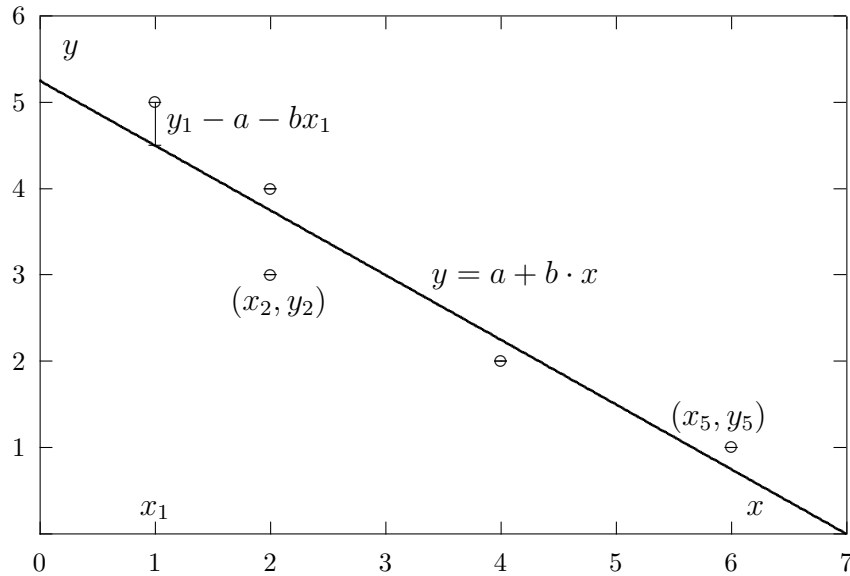
$$y_i - \hat{y}_i = y_i - (a + b x_i) = y_i - a - b x_i, \quad i = 1, \dots, n \quad (7.46)$$

sind also die **parallel zur  $y$ -Achse gemessenen Abstände** der Datenpunkte  $(x_i, y_i)$  zur Geraden  $y = a + b x$  über den Werten  $x_i$ .

Mit den gelegentlich verwendeten Bezeichnungen  $Y = \beta_0 + \beta_1 X + E$  für den linearen Ansatz und  $y = b_0 + b_1 x$  für die KQ-Gerade will man andeuten, dass die einfache lineare Regression als Spezialfall der sog. **Multiplen Regression**

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + E$$

mit mehreren erklärenden Variablen  $X_1, \dots, X_p$  angesehen werden kann.

Bild 19:  $(x, y)$ -Punkte und Gerade  $y = a + b x$ 

Als Maß für eine Anpassung der Geraden an die Datenpunkte wird die Summe der oben definierten **Abstandsquadrate** herangezogen. Diese Quadratsumme ist für eine optimale Anpassung zu minimieren, d. h. *Achsenabschnitt*  $a$  und *Steigung*  $b$  sind so zu wählen, dass die Fehlerquadratsumme

$$Q(a, b) = \sum_{i=1}^n [y_i - (a + b x_i)]^2 = \sum_{i=1}^n [y_i - \hat{y}_i]^2$$

möglichst klein wird.

#### Satz 7.40 (KQ-Schätzungen)

Die optimalen **KQ-Schätzungen** berechnen sich zu

$$\hat{\beta} = b = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \quad (7.47)$$

und

$$\hat{\alpha} = a = \bar{y} - b \bar{x}. \quad (7.48)$$

$a$  und  $b$  sind eindeutig bestimmt, wenn wenigstens zwei verschiedene Werte  $x_i$  vorliegen.

$$y = a + b x \quad (7.49)$$

heißt **empirische Regressionsgerade** oder **KQ-Gerade**.



Beweis von Satz 7.40:

Sei  $e_i(a, b) := y_i - a - b x_i$ . An einem Minimum der Residuensumme

$$Q = Q(a, b) := \sum_{i=1}^n [y_i - (a + b x_i)]^2 = \sum_{i=1}^n [e_i(a, b)]^2$$

muss der Gradient verschwinden, d. h.

$$\frac{\partial Q}{\partial a} = \sum_i 2e_i \frac{\partial e_i}{\partial a} = \sum_i 2e_i (-1) = 0 \implies \frac{1}{n} \sum e_i = 0 \quad (7.50)$$

$$\frac{\partial Q}{\partial b} = \sum_i 2e_i \frac{\partial e_i}{\partial b} = \sum_i 2e_i (-x_i) = 0 \implies \frac{1}{n} \sum x_i e_i = 0. \quad (7.51)$$

(7.50) impliziert

$$\frac{1}{n} \sum_i (y_i - a - b x_i) = 0 \iff \bar{y} - a - b \bar{x} = 0 \iff a = \bar{y} - b \bar{x}.$$

Weiter ergeben (7.51) und (7.42) mit  $s_x^2 > 0$  (da die  $x_i$  nichtkonstant)

$$\begin{aligned} \sum_i (y_i - a - b x_i) x_i = 0 &\iff \sum_i x_i y_i - a \sum_i x_i - b \sum_i x_i^2 = 0 \\ &\iff \sum_i x_i y_i - n \bar{x} a - b \sum_i x_i^2 = 0 \\ &\iff \sum_i x_i y_i - n \bar{x} (\bar{y} - b \bar{x}) - b \sum_i x_i^2 = 0 \\ &\iff \sum_i x_i y_i - n \bar{x} \bar{y} - b (\sum_i x_i^2 - n \bar{x}^2) = \\ &= (n-1) (s_{xy} - b s_x^2) = 0 \\ &\iff b = \frac{s_{xy}}{s_x^2} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2}. \end{aligned}$$

Die Hessematrix von  $Q$  ist konstant (bzgl.  $a, b$ ) und lautet

$$\nabla^2 Q = \begin{pmatrix} \frac{\partial^2 Q}{\partial a^2} & \frac{\partial^2 Q}{\partial a \partial b} \\ \frac{\partial^2 Q}{\partial b \partial a} & \frac{\partial^2 Q}{\partial b^2} \end{pmatrix} = 2 \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}. \quad (7.52)$$

Diese Matrix ist unter den gegebenen Voraussetzungen

$$x_i \text{ nichtkonstant} \iff s_x^2 > 0 \iff n \sum x_i^2 - (\sum x_i)^2 > 0$$

positiv definit (warum?). Damit sind  $a$  und  $b$  globale Minimierer von  $Q$ .  $\square$

Als weitere Bezeichnung sei noch die **empirische Streuung des Fehlers** eingeführt:

$$s_e^2 := \frac{Q(a, b)}{n-1} = \frac{1}{n-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - (a + b x_i))^2. \quad (7.53)$$

**Satz 7.41 (Bestimmtheitsmaß)**

Das sog. **Bestimmtheitsmaß** lautet im linearen Fall

$$R^2 = r_{xy}^2 = \frac{s_{xy}^2}{s_x^2 s_y^2} = 1 - \frac{s_e^2}{s_y^2}, \quad (7.54)$$

wobei jetzt  $0 \leq R^2 \leq 1$ .

**Lemma 7.42**

$$s_e^2 = \frac{1}{n-1} Q(a, b) = s_y^2 - b^2 s_x^2 = s_y^2 (1 - r_{xy}^2) = \frac{1}{n-1} \sum_{i=1}^n e_i^2. \quad (7.55)$$

Beweis des Lemmas:

Mit  $a = \bar{y} - b\bar{x}$  gilt

$$\begin{aligned} Q(a, b) &= \sum_{i=1}^n (y_i - (a + b x_i))^2 \\ &= \sum_{i=1}^n (y_i - (\bar{y} - b\bar{x}) - b x_i)^2 \\ &= \sum_{i=1}^n (y_i - \bar{y} - b(x_i - \bar{x}))^2 \\ &= \sum_{i=1}^n \left[ (y_i - \bar{y})^2 - 2b(y_i - \bar{y})(x_i - \bar{x}) + b^2(x_i - \bar{x})^2 \right] \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 - 2b \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) + b^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= (n-1) \left( s_y^2 - 2b s_{xy} + b^2 s_x^2 \right) \\ &= (n-1) \left( s_y^2 - 2 \frac{s_{xy}}{s_x} + \frac{s_{xy}^2}{s_x^2 s_x} s_x^2 \right) \\ &= (n-1) \left( s_y^2 - \frac{s_{xy}^2}{s_x^2 s_x} s_x^2 \right) = (n-1) \left( s_y^2 - b^2 s_x^2 \right) \\ &= (n-1) s_y^2 \left( 1 - \frac{s_{xy}^2}{s_x^2 s_y^2} \right) = (n-1) s_y^2 (1 - r_{xy}^2). \end{aligned}$$

□

Beweis von Satz 7.41:

Da  $(a, b)$  globale Minimierer von  $Q$  sind, folgt

$$0 \leq \frac{1}{n-1} Q(a, b) = s_e^2 \leq \frac{1}{n-1} Q(\bar{y}, 0) = s_y^2.$$

1. Fall:

$$y_i = \text{const} = \bar{y} \implies s_y^2 = 0 = Q(\bar{y}, 0).$$

$$a = \bar{y}, b = 0, r_{xy} \text{ nicht definiert.}$$

2. Fall:

$$y_i \neq \text{const} \iff s_y^2 > 0$$

$$\implies 0 \leq \frac{s_e^2}{s_y^2} \leq \frac{s_y^2}{s_y^2} = 1$$

$$\iff 0 \leq \frac{s_y^2}{s_y^2} (1 - r_{xy}^2) \leq 1$$

$$\iff 0 \leq 1 - r_{xy}^2 \leq 1$$

$$\iff 0 \leq r_{xy}^2 \leq 1$$

$$\iff -1 \leq r_{xy} \leq 1. \tag{7.56}$$

Zur Interpretation dieser Größen zunächst ein Beispiel:

Beschreiben die  $x_i$  diskrete Zeitpunkte wie Jahre, Halbjahre, Quartale oder Monate, so spricht man von **Zeitreihen**. Das  $Y$ -Merkmal kann dann für Aktienkurse, Umsatzentwicklungen, Arbeitslosenzahlen etc. stehen. Für Zeitreihen wird die KQ-Gerade auch **Trendgerade** genannt. In den folgenden Daten schlägt sich eine Halbjahres-Saison (zum Beispiel Winter/Sommer) nieder:

<b>Tabelle 9:</b>						
Beispiel Zeitreihe						
i	$x_i$	$y_i$	$x_i y_i$	$x_i^2$	$y_i^2$	Saison und Jahr
1	1	2	2	1	4	Winter 1993/94
2	2	4	8	4	16	Sommer 1994
3	3	5	15	9	25	Winter 1994/95
4	4	7	28	16	49	Sommer 1995
5	5	8	40	25	64	Winter 1995/96
6	6	10	60	36	100	Sommer 1996
$\Sigma$	21	36	153	91	258	

Man erhält

$$\bar{x} = \frac{7}{2} = 3.5, \quad \bar{y} = 6,$$

$$\sum_{i=1}^6 x_i y_i - 6 \bar{x} \bar{y} = 153 - 6 \frac{7}{2} 6 = 153 - 126 = 27,$$

$$\sum_{i=1}^6 x_i^2 - 6\bar{x}^2 = 91 - 6 \left(\frac{7}{2}\right)^2 = 91 - 6 \frac{49}{4} = \frac{182}{2} - \frac{147}{2} = \frac{35}{2} = 17.5,$$

$$\sum_{i=1}^6 y_i^2 - 6\bar{y}^2 = 258 - 6 \cdot 6^2 = 258 - 216 = 42.$$

und die Regressionsgerade  $y = a + bx$  mit

$$b = \frac{\sum_{i=1}^6 x_i y_i - 6\bar{x}\bar{y}}{\sum_{i=1}^6 x_i^2 - 6\bar{x}^2} = \frac{27}{35/2} = \frac{54}{35} = 1.543 \quad (7.57)$$

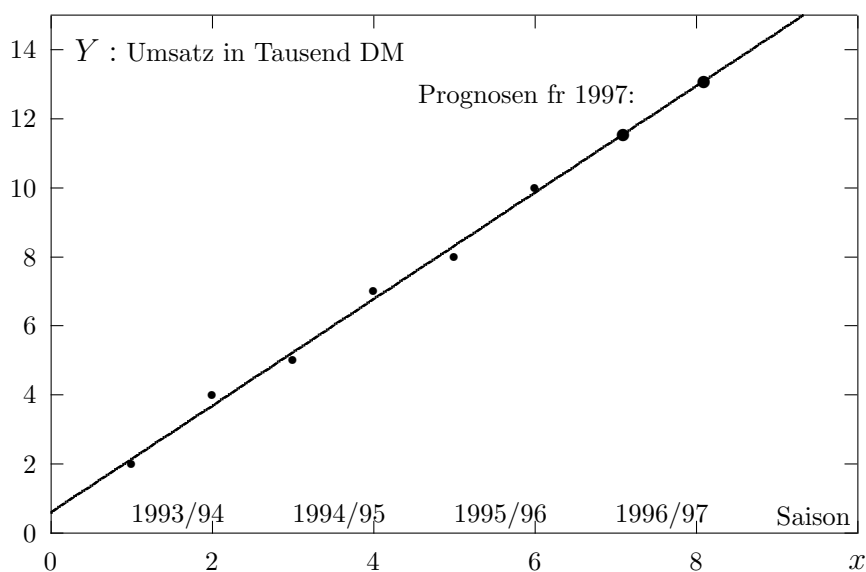
$$a = \bar{y} - b\bar{x} = 6 - \frac{54}{35} \cdot \frac{7}{2} = 6 - \frac{27}{5} = \frac{30 - 27}{5} = \frac{3}{5} = 0.6. \quad (7.58)$$

Als Bestimmtheitsmaß  $R^2$  ergibt sich

$$R^2 = \frac{\left(\sum_{i=1}^6 x_i y_i - 6\bar{x}\bar{y}\right)^2}{\left(\sum_{i=1}^6 x_i^2 - 6\bar{x}^2\right) \left(\sum_{i=1}^6 y_i^2 - 6\bar{y}^2\right)} = \frac{27^2}{\frac{35}{2} \cdot 42} = \frac{729 \cdot 2}{35 \cdot 42} = \frac{1458}{1470} = 0.9918 \quad (7.59)$$

$$r_{xy}^2 = R^2 = 0.9918 \quad \implies \quad r_{xy} = 0.9959.$$

Obwohl nicht alle Datenpunkte exakt auf einer Geraden liegen, wird doch ein ausgesprochen hoher Korrelationskoeffizient sehr nahe bei 1 erreicht. In anderen Anwendungen ist man schon mit  $r_{xy} \simeq 0.9$  zufrieden.



**Bild 20:** Zeitreihe aus Winter- und Sommersaison mit Prognosen

Die Prognosen sind berechnet nach der Formel  $\hat{y}(x) = a + bx$ . Dem Winter 1996/97 wird dabei der Wert  $x = 7$  und dem Sommer 1997  $x = 8$  zugeordnet:

Saison	$x$ -Wert	Prognose
Winter 1996/97	7	$\hat{y}(7) = 0.6 + 1.543 \cdot 7 = 11.401$
Sommer 1997	8	$\hat{y}(8) = 0.6 + 1.543 \cdot 8 = 12.944$

Auf nachträgliche Saison-Korrekturen der Prognosen (im Winter nach unten und im Sommer nach oben) können wir in diesem Rahmen nicht näher eingehen.

In Computerprogrammen sollten zunächst  $\bar{x}$  und  $\bar{y}$  und danach aus numerischen Gründen die Steigung  $b$  unbedingt nach der Formel

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (7.60)$$

erzeugt werden. Für manuelle Rechnungen in Übungs- und Klausuraufgaben wird dagegen gerne das oben dargestellte Schema mit **Verschiebungsregeln** benützt. Dieses ist für die Handrechnung insbesondere von Vorteil, wenn  $\bar{x}$  oder  $\bar{y}$  keine ganzen Zahlen sind. Bei größeren Datenmengen wird man sowieso Rechner benützen.

Über die Regressionsgerade sind also **Prognosen** der Art

$$\hat{y}(x) = a + bx \quad (7.61)$$

möglich. Bei Zeitreihen wird  $x$  in der Regel ein zukünftiger Zeitpunkt sein.  $\hat{y}(x)$  liegt immer genau auf der Trendgeraden.

Bestimmtheitsmaß und empirischer Korrelationskoeffizient messen die **lineare Abhängigkeit** der Merkmale  $X$  und  $Y$  und sind folgendermaßen zu interpretieren

$$\begin{aligned} R^2 = r_{xy}^2 = 1 & : \text{ alle Datenpunkte liegen exakt auf der Regressionsgeraden} \\ r_{xy} > 0 & : \text{ positive Steigung der emp. Regressionsgeraden, d. h. } b > 0 \\ r_{xy} < 0 & : \text{ negative Steigung der emp. Regressionsgeraden, d. h. } b < 0 \\ r_{xy} = 0 & : \text{ Regressionsgerade horizontal, d. h. } b = 0 ; \\ & : \text{ kein linearer Zusammenhang zwischen } X \text{ und } Y ! \end{aligned}$$

Von einem ausreichend "guten" linearen Zusammenhang kann man allenfalls bei

$$R^2 = r_{xy}^2 > \frac{1}{2} \quad \text{bzw.} \quad |r_{xy}| > 0.7$$

ausgehen, denn dann wird die ursprüngliche Varianz  $s_y^2$  durch den Regressionsansatz in  $s_e^2$  zumindest auf die Hälfte reduziert. **Nichtlineare Abhängigkeiten**, z. B.  $Y = X^2$ , werden von der linearen Regressionsanalyse bzw. vom empirischen Korrelationskoeffizienten nicht erfaßt.

**Vorsicht mit Korrelationen!**

Tabelle 10: Ein Beispiel mit $r_{xy} = 0.957$		
Jahr	Autozulassungen x (in Tausend)	Hühner y (in Tausend)
1942	1.847	53.502
1943	1.544	46.371
1944	1.599	50.242
1945	2.599	56.666
1946	3.113	61.723
1947	3.521	64.880
1948	3.734	79.219
1949	4.113	89.152
1950	4.414	90.798
1951	4.625	90.067

Bei der Interpretation von Korrelationen darf nie der sachlogische Hintergrund unberücksichtigt bleiben. Bei obigem Beispiel führen die hohen Korrelationen

Kombination	$r_{xy}$
Jahr/Autozulassungen:	$r_{xy} = 0.971$
Jahr/Hühner:	$r_{xy} = 0.950$

zu der hohen positiven "Scheinkorrelation" von Autozulassungen/Hühner von  $r_{xy} = 0.957$ , was beim Betrachter ein Schmunzeln hervorruft:

Man stellt sich eher auf eine fallende Anzahl von Hühnern bei steigendem Autoverkehr ein, da zu erwarten ist, dass das eine oder andere (freilaufende) Tier versehentlich überfahren wird.

**Nichtlineare Zusammenhänge und Transformationen**

Die unbekannt Parameter, etwa  $a$ ,  $b$ ,  $c$ , in allgemeinen nichtlinearen Beziehungen der Form

$$Y = a + b e^{cX} + E$$

müssen mit iterativen Methoden geschätzt werden; vgl. u.a. Seber & Wild (1989). Eine Reihe wichtiger Ansätze läßt sich aber durch Transformationen auf den bereits besprochenen Fall zurückführen. Wichtig ist dabei, dass am Ende die zu schätzenden Parameter nur **linear** in die Beziehung eingehen. Wir geben einige Beispiele an, wobei die ebenfalls zu transformierende Fehlervariable weggelassen wird:

$$Y = a + b X^2 \quad \rightarrow \quad Y = a + b \tilde{X}, \quad \tilde{X} = X^2$$

$$Y = a e^{bX} \quad \rightarrow \quad \tilde{Y} = \tilde{a} + b X, \quad \tilde{Y} = \ln(Y), \quad \tilde{a} = \ln(a).$$

Eine weitere Transformation soll am folgenden Beispiel erläutert werden. Gemäß [Kredler & Ritter (1995)], Abschnitt 1.4.5 erhöht der Einsatz von  $X$  Mengeneinheiten eines biologischen Schädlingsbekämpfungsmittels den Ertrag einer Beersorte um  $Y$  Mengeneinheiten. Da sich der Ertrag nicht beliebig erhöhen läßt, sondern wohl eine Sättigung zu erwarten ist, scheidet ein linearer Ansatz aus. Eine einfache Modellierung könnte lauten

$$Y = a + b \frac{1}{1 + X},$$

was sich mit  $\tilde{X} = 1/(1 + X)$  in der bekannten Form  $Y = a + b\tilde{X}$  schreiben läßt. Die Versuchsreihen auf vergleichbaren Feldern ergaben

Tabelle 11: Nichtlineare Transformation						
	i	1	2	3	4	5
Schädlingsbekämpfungsmittel	$x_i$	0	1	3	6	11
Transformation	$\tilde{x}_i = 1/(1 + x_i)$	1	0.5	0.25	0.1428	0.0833
Ertrag	$y_i$	1	3	4	5	6

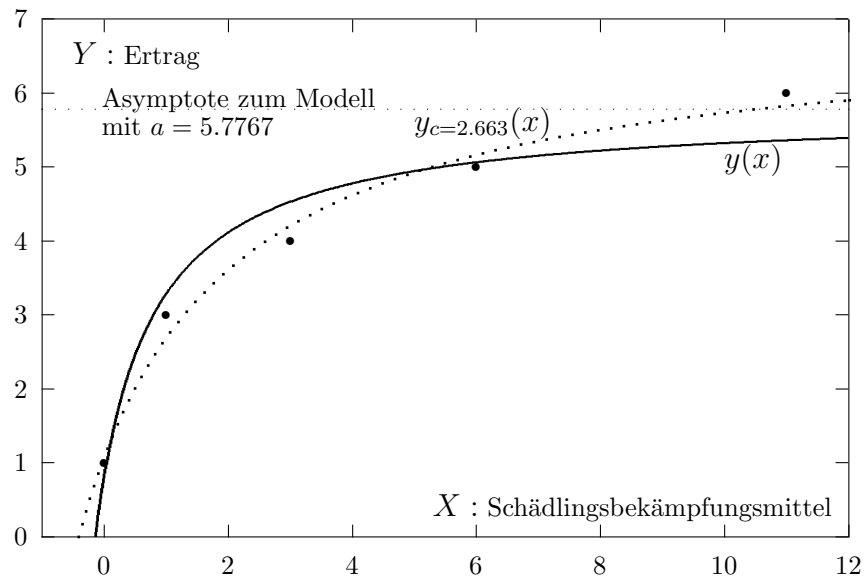
Man erhält bei einem Bestimmtheitsmaß (für das lineare Modell bzgl.  $Y$  und  $\tilde{X}$ ) von  $R^2 = 0.9448$  die Koeffizienten  $a = 5.7767$  und  $b = -5.0020$  sowie die Regressionsfunktion

$$y = a + b\tilde{x} = 5.7767 - \frac{5.0020}{1 + x} = y(x),$$

wobei der mittlere Ertrag ohne Schädlingsbekämpfungsmittel auf  $y(0) = 0.7747$  und bei beliebig hohem Einsatz auf  $y(\infty) = 5.7767$  geschätzt wird. Die Fehlerquadratsumme beträgt 0.817. Diese läßt sich auf 0.2 drücken, wenn man einen zusätzlichen Parameter  $c$  einführt:

$$Y = a + b \frac{1}{c + X}.$$

In diesem für  $c$  nichtlinearen Modell lauten die optimalen Parameter  $a = 6.966$ ,  $b = -15.615$ ,  $c = 2.663$ . Wie auch die Graphen im folgenden Bild zeigen, war die Fixierung der 1 im Nenner nicht unbedingt günstig. Insbesondere liegt die Sättigungsgrenze im zweiten Modell erst bei 6.966, was angesichts der Daten realistischer erscheint.

**Bild 21: Transformation auf lineares Modell**



## 7.10 Beispiele

Ein wichtiges Ziel der zugrundeliegenden Vorlesung ist es, die Hörer zum Umgang mit dem Rechner anzuleiten und Daten am Rechner zu analysieren. Hier folgen einige Grafiken und Ausdrücke von kommerziellen Programmen und Hinweise auf weitere Beispieldaten.

### 7.10.1 Grafiken und Ausdrücke von Statistikprogrammen

Eine statistische Datenanalyse kann mit speziellen Programmsysteme wie **SAS**, **SPSS** oder **S-Plus**, um nur einige zu nennen, wesentlich effektiver und oft auch einfacher durchgeführt werden als etwa mit Excel. Um dem Leser aber eine schnelle Einstiegsmöglichkeit für statistische Berechnungen am Computer zu bieten, wird hier auch auf das verbreitete Excel zurückgegriffen. Sicherlich gibt es noch viele weitere Programme, die einem ähnlichen Zweck dienen können.

#### Grafiken in Excel

Einen Excel-Ausdruck für die Daten von Tabelle 9 erhält man etwa mit den Kommandos wie *Diagramm* (Icon in Funktionsleiste), *Punkt (XY)*, *Diagramm* (in Menüzeile), *Trendgerade hinzufügen*, *Formel in Diagramm darstellen* etc.

Zur Berechnung wichtiger Populationskenngrößen wie Mittelwert, Varianz, Quartilen etc. benütze man die vorgesehenen **Excel Statistik-Funktionen** mit gleichen oder ähnlichen Namen.

Für häufig vorkommende Fragestellungen gibt es außerdem Sammlungen von Excel-Macros. Dazu wählt man **Extras** und dort den Untermenüpunkt **Analyse-Funktionen**, der beim ersten Bedienen mit dem sog. **Add-In-Manager...** aktiviert werden muss.

Als Statistik-Macros sind für unsere Zwecke etwa **Populationskenngrößen** (z.B. Mittelwert, Median, Quartile, *IQR*) und **Histogramm** interessant.

#### Stichprobe mit 100 normalverteilten Beobachtungen

Bei der SAS-Grafik (Box-Plot + Histogramm) und dem Ausdruck mit Populationskenngrößen handelt es sich um eine Simulation von 100 unabhängigen standardnormalverteilten Zufallszahlen. Der Abschnitt 10.3 befaßt sich genauer mit sog. Box-Plots. Im Unterschied zu einem dort diskutierten ähnlichen Beispiel liegt hier keine der Beobachtungen außerhalb der Nadeln. Man beachte:

1. Die empirische Schiefe = *Skewness* = 0.0668 stimmt schon fast mit dem Sollwert 0 überein. Histogramm und Box-Plot zeigen ebenfalls eine hinreichende Symmetrie der Daten an.
2. Der empirische Exzess = *Kurtosis* = -0.4146 mißt in etwa, wie nah die Daten um den Mittelwert konzentriert sind. Bei exakt normalverteilten Beobachtungen

ist der Idealwert 0. Das Beispiel zeigt, dass beim empirischen Exzess Abweichungen ungefähr im Bereich  $[-0.5; 0.5]$  akzeptiert werden müssen.

## Sonnendaten; garantiert nicht normalverteilt

Zum Abschluß sei noch ein Beispiel aus dem Unterverzeichnis SASHELP mit  $n = 126$  Quartalen aus den Jahren 1960 bis Mitte 1991 mit dem Exportüberschuss eines großen Landes in Mrd. \$ angegeben. Negative Werte zeigen an, dass mehr Güter importiert als ins Ausland verkauft wurden.

Zum Vergleich mit dem vorherigen Beispiel (Normalverteilung) sind auffallend:

- Der empirische Median liegt nicht in der Mitte der Box.
- Eine große Anzahl von Datenpunkten ist "links" der Box-Plot-Nadelspitze
- $Skewness = -1.4920 \ll 0$ ,  $Kurtosis = +1.2653 \gg 0$ .

Der Ausdruck wurde wieder mit SAS erstellt.

### 7.10.2 Einige Datenbeispiele

Die folgenden Daten sind im Netz verfügbar und werden neben anderen Datensätzen im Statistik-Praktikum zu dieser Vorlesung analysiert.

$n$  = Anzahl der Beobachtungen (Datenzeilen)

$p$  = Anzahl der Variablen (Datenspalten)

Datei.sd2	n	p	Kurzbeschreibung
<i>huhnauto</i>	10	3	Hühner = $10612444 + 5487 \cdot \text{Jahr}$ ; $\rho = 0.95$ Hühner = $23494 + 14.39 \cdot \text{Auto}$ ; $\rho = 0.96$ Auto = $-722108 + 372.6 \cdot \text{Jahr}$ ; $\rho = 0.97$
<i>kornstoy</i>	30	2	Korndurchmesser und Druckfestigkeit aus Stoyan (1993)
<i>gewicht</i>	22	1	Gewichte von Schülern: $\bar{x} = 61.55$ , $s_x = 2.53$ , $x_{0.5} = 61.3$
<i>indian</i>	27	2	Höchstgeschwindigkeit im Indianapolis-Rennen; $\rho = 0.99$ $v = 61.13 + 2.715 \cdot \text{Jahr}$
<i>groegew</i>	40	1	Gewicht = $-96.15 + 0.96 \cdot \text{Größe}$ ; $\rho = 0.735$
<i>outpkost</i>	10	2	Kosten = $1000 + 24 \cdot \text{Output}$ ; $\rho = 0.946$
<i>mendele</i>	7	2	Löslichkeit/100g $H_2O$ von Natriumnitrat bzgl. Temperatur in $^{\circ}C$ des Wassers nach Mendelejew Löslichkeit = $66.95 + 1.036 \cdot \text{Temperatur}$ ; $\rho = 0.99$
<i>normal</i>	100	1	standardnormalverteilte Daten (simuliert)
<i>umsatz</i>	23	1	$\bar{x} = 226.0$ , $s = 91.31$ , $Q_1 = 154.9$ , $Q_3 = 285.4$
<i>zeitreih</i>	12	2	Quartalsdaten: $Trend = 0.6288 + 0.4065 \cdot \text{Monat}$
<i>eufirm</i>	94	1	Auszug aus <i>gnp</i> : $\bar{x} = 115.95$ , $s = 289.20$
<i>gnp</i>	94	8	aus <i>SASHELP</i>

# Kapitel 8

## Schätzfunktionen, ML-Prinzip

Die Wahrscheinlichkeitstheorie setzt stets einen Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{F}, P)$  als gegeben voraus. Damit sind Verteilungen, Erwartungswerte etc. im Prinzip bekannt, auch wenn sich gewisse Herleitungen als schwierig oder gar nicht realisierbar erweisen. Im Gegensatz dazu betrachtet die mathematische Statistik das in der Praxis häufiger vorliegende Problem, Aussagen über die unbekannte Verteilungsstruktur bzw. gewisse Kenngrößen von Zufallsvariablen (ZV) aus der Beobachtung von Stichproben zu gewinnen.

Ohne Zusatzannahmen ist diese Aufgabe nur in trivialen Fällen zu lösen. Meist setzt man voraus, dass  $F$  einer bestimmten Funktionenfamilie angehört, die nur noch von wenigen (in unserem Fall meist ein bis zwei) Parametern abhängt. Während die optimale Schätzung dieser unbekannt Parameter für einen Großteil der wichtigen Fälle gelöst ist, muss die Wahl der Verteilungsfamilie (etwa Normal-, Exponential-, Gleich- oder sonstige Verteilung) vom Statistiker bei jeder Anwendung neu vorgenommen werden. Als Hilfsmittel können u.a. dienen

1. Verteilungstests oder grafische Methoden wie QQ-Plots; siehe Abschnitt 10.2
2. Ergebnisse gleichartiger Untersuchungen aus der Vergangenheit
3. Theoretische Überlegungen: Man kann etwa die ZV  $X =$  Körpergröße als linear-additive Überlagerung vieler (teilweise nicht beobachtbarer) ZV wie Größe des Vaters, mittlerer Sauerstoffgehalt der Luft während der Schwangerschaft, mittlerer Eiweißgehalt der Ernährung in den ersten zehn Lebensjahren usw. ansehen. Aus dem zentralen Grenzwertsatz folgt dann, dass  $X$  als approximativ normalverteilt angesehen werden kann.

Sei also eine geeignete parametrisierte Verteilungsfamilie festgelegt. Wir diskutieren nun kurz die Eigenschaften von **Schätzfunktionen** für die noch zu bestimmenden unbekannt Parameter. Die Information zur Schätzung gewinnt man durch unabhängige Wiederholung des Zufallsexperiments für  $X$ .

$\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$  seien unbekannt Parameter; z. B.  $\boldsymbol{\theta} = (\mu, \sigma^2)^T$  bei der Normalverteilung.  $X, X_1, \dots, X_n$  seien iid, d. h.  $X_1, \dots, X_n$  sind unabhängig und jedes  $X_i, i = 1, \dots, n$ , hat dieselbe Verteilungsfunktion (VF)  $F(x; \boldsymbol{\theta}) = F_X(x; \boldsymbol{\theta})$  wie

$X$ . Man nennt in diesem Fall  $X_1, \dots, X_n$  auch mathematische Stichprobe; vgl. Def. 2.29.

Im Gegensatz zu früheren Kapiteln wird jetzt die Abhängigkeit der VF (und später der Wahrscheinlichkeitsfunktion bzw. Dichte) von den unbekannt Parametern  $\theta \in \mathbb{R}^m$  durch die Schreibweise  $F(x; \theta)$  hervorgehoben.

Zur Schätzung  $\hat{\theta}$  für  $\theta \in \Theta \subset \mathbb{R}^m$  benötigt man eine Borel-messbare Funktion

$$T : \mathbb{R}^n \rightarrow \Theta, \quad T(X_1, \dots, X_n) = \hat{\theta} \in \Theta,$$

mit gewissen zusätzlichen Eigenschaften, z. B. Erwartungstreue und Konsistenz, die im Anschluss genauer erklärt sind. Ein weiteres Gütekriterium ist die minimale Varianz einer Schätzfunktion.  $T$  wird auch einfach **Statistik** oder **Schätzer** (estimator) genannt.  $T$  ist selbst eine ZV, deren Verteilung u.a. vom Parameter  $\theta$  bestimmt ist.

## 8.1 Erwartungstreue und Konsistenz

### Def. 8.1 (Erwartungstreu, unverzerrt; unbiased)

Eine Schätzfunktion  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  heißt **erwartungstreu** oder **unverzerrt** (unbiased), falls

$$E[T(X_1, \dots, X_n)] = \theta.$$

Die Abweichung  $E[T(X_1, \dots, X_n)] - \theta$  heißt **Bias** des Schätzers  $T$ .

### Beispiel 8.2 (Erwartungstreue Schätzer)

1. Seien  $X_1, \dots, X_n$  iid mit  $E(X_i) = \mu$ , dann ist

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

ein erwartungstreuer Schätzer für  $\mu$ .

2. Seien  $X_1, \dots, X_n$  iid mit  $E(X_i) = \mu$  und  $Var(X_i) = \sigma^2$ . Eine erwartungstreu Schätzfunktion für  $\sigma^2$  lautet gemäß Bsp. 2.34

$$T(X_1, \dots, X_n) = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Die Begriffe Schätzer, Statistik und Schätzfunktion bzw. erwartungstreu und unverzerrt werden abwechselnd verwendet, um alle Bezeichnungen einzuüben.

Zur Diskussion einer weiteren Eigenschaft von Schätzern schreiben wir im nächsten Abschnitt  $T_n$  für  $T$ , um die Abhängigkeit vom Stichprobenumfang  $n$  anzudeuten. Einer

einfachen Darstellung wegen beschränken wir uns hier außerdem auf eindimensionale Statistiken  $T_n$  und Parameter  $\theta$ . Der Übergang zum mehrdimensionalen Fall läßt sich leicht vollziehen.

**Def. 8.3 (Konsistenz)**

Ein Schätzer  $T_n : \mathbb{R}^n \rightarrow \mathbb{R}$  heißt **konsistent** (*consistent*) für  $\theta$ , falls

$$\lim_{n \rightarrow \infty} P(|T_n(X_1, \dots, X_n) - \theta| \leq \epsilon) = 1, \quad \forall \epsilon > 0.$$

Mit den Bezeichnungen von Def. 4.6 bedeutet Konsistenz, dass  $T_n$  stochastisch gegen  $\theta$  konvergiert, i.Z.

$$T_n(X_1, \dots, X_n) \xrightarrow{P} \theta.$$

**Beispiel 8.4 (Konsistenz und Erwartungstreue)**

1. Für iid  $X_1, \dots, X_n$  mit  $E(X_i) = \mu$  und  $Var(X_i) = \sigma^2$  ist

$$T_n(X_1, \dots, X_n) = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

nach dem Satz von Tschebyschow 4.5 ein konsistenter Schätzer für  $\mu$ .

2. Nun seien die  $X_1, \dots, X_n$  iid  $N(\mu, \sigma^2)$ . Damit gilt für die folgenden drei  $\sigma^2$ -Schätzer

$$\tilde{S}_n^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad \text{ist konsistent, aber nicht unverzerrt,}$$

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad \text{ist konsistent und unverzerrt,}$$

$$\hat{S}_n^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \quad \text{ist konsistent und unverzerrt.}$$

## 8.2 Schätzfunktionen minimaler Varianz

$T$  heißt **MV**(Minimum-Varianz)- oder **wirksamste Schätzfunktion**, wenn  $T$  unter allen erwartungstreuen Schätzfunktionen für  $\theta$  minimale Varianz besitzt (wir schreiben hier wieder  $T$  statt  $T_n$  und  $\bar{X}$  statt  $\bar{X}_n$ ).

**Beispiel 8.5 (MV-Schätzer)**

Für die Parameter der meisten hier behandelten Verteilungen können wirksamste Schätzer angegeben werden.

1. Binomialverteilung:  $\bar{X}$  ist MV-Schätzer für  $p$
2. Poissonverteilung:  $\bar{X}$  ist MV-Schätzer für  $\lambda$
3. Normalverteilung:  $\bar{X}$  ist MV-Schätzer für  $\mu$ ,  $S^2$  ist MV-Schätzer für  $\sigma^2$
4. Exponentialverteilung:  $\bar{X}$  ist MV-Schätzer für  $1/\lambda$
5. Gleichverteilung in  $[0, b]$ :  $\hat{b} = \frac{n+1}{n} \max_i \{X_i\}$  ist MV-Schätzer für  $b$ .

$\bar{X}$  ist linear in den  $X_i$ . Damit ist  $\bar{X}$  in den genannten Beispielen auch ein sog. **BLUE-Schätzer** (Best Linear Unbiased Estimator). Die Eigenschaft "best" bedeutet wieder, dass BLUE-Schätzer unter allen linearen, erwartungstreuen Schätzern minimale Varianz haben. BLUE-Schätzer spielen auch bei linearen Regressionsmodellen eine wichtige Rolle.

Zum Nachweis der BLUE-Eigenschaft von  $\bar{X}$  für iid  $X_1, \dots, X_n$  mit  $E(X_i) = \mu$ ,  $Var(X_i) = \sigma^2$ ,  $i = 1, \dots, n$ , macht man den Ansatz einer allgemeinen linearen Schätzfunktion

$$T(X_1, \dots, X_n) = \sum_{i=1}^n a_i X_i, \quad a_i \in \mathbb{R}, i = 1, \dots, n. \quad (8.1)$$

Die Erwartungstreue wird dann durch die Nebenbedingung

$$\sum_{i=1}^n a_i = 1 \quad (8.2)$$

gesichert. Die Minimierung von

$$Var(T) = Var\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 Var(X_i) = \sigma^2 \sum_{i=1}^n a_i^2 \quad (8.3)$$

unter der Nebenbedingung (8.2) liefert als optimale Lösung

$$a_i = \frac{1}{n}, \quad i = 1, \dots, n.$$

## 8.3 Konstruktion von Schätzern

Von vielen denkbaren Schätzmethoden seien drei wichtige kurz erwähnt.



### 8.3.1 Methode der Kleinsten Quadrate (KQ)

Dieses Prinzip haben wir bereits im Kapitel 7 über empirische Statistik kennengelernt. Zum Beispiel ist  $\bar{X}$  u.a. KQ-Schätzer, da

$$\sum_{i=1}^n [X_i - \bar{X}]^2 \leq \sum_{i=1}^n [X_i - c]^2, \quad \text{für alle } c \in \mathbb{R}. \quad (8.4)$$

In analoger Weise wurden auch die KQ-Schätzungen für Achsenabschnitt und Steigung der Regressionsgerade in Abschnitt 7.9 eingeführt.

### 8.3.2 Momentenmethode

Seien  $X, X_1, \dots, X_n$  iid mit existierendem absolutem  $k$ -ten Moment  $E[|X^k|]$ . Dann sind gemäß Satz 1.71 auch die Momente  $\mu_j = E[X^j]$ ,  $j = 1, \dots, k$ , erklärt. Als Schätzungen für die  $\mu_j$  bietet sich etwa

$$\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^n X_i^j, \quad j = 1, \dots, k \quad (8.5)$$

an.

### 8.3.3 Maximum-Likelihood (ML)-Methode

Die wohl am häufigsten angewendete Technik zur Konstruktion von Schätzfunktionen ist die sog. **ML-** oder **Maximum-Likelihood-Methode**. Bei normalverteilten  $X_i$  ist diese äquivalent mit der Methode der kleinsten Quadrate.

Die iid-Variablen  $X, X_1, \dots, X_n$  seien diskret oder stetig mit Wahrscheinlichkeitsfunktion bzw. Dichte  $f(x; \boldsymbol{\theta}) = f_X(x; \boldsymbol{\theta})$  (Schätzfunktion  $T$  und der unbekannte Parameter  $\boldsymbol{\theta}$  sind jetzt wieder  $m$ -dimensional). Wegen der Unabhängigkeit der  $X_i$  lautet die gemeinsame Dichte von  $(X_1, \dots, X_n)$

$$l(x_1, \dots, x_n; \boldsymbol{\theta}) = \prod_{i=1}^n f(x_i; \boldsymbol{\theta}). \quad (8.6)$$

Bei gegebenen Stichprobenrealisierungen  $x_1, \dots, x_n$  hängt  $l$  nur noch vom Parameter  $\boldsymbol{\theta} \in \mathbb{R}^m$  ab. Man nennt  $l$  in diesem Fall **Likelihoodfunktion**. Nach der ML-Methode wählt man nun die Schätzungen  $\hat{\boldsymbol{\theta}}_{ML}$  für den unbekannt Parameter  $\boldsymbol{\theta}$ , so dass das Eintreten der beobachteten Stichprobe maximale Wahrscheinlichkeit (im Englischen verwendet man hier den Begriff *Likelihood* im Unterschied zu *probability*) besitzt. Also

$$l(x_1, \dots, x_n; \hat{\boldsymbol{\theta}}_{ML}) \geq l(x_1, \dots, x_n; \tilde{\boldsymbol{\theta}}), \quad \text{für alle } \tilde{\boldsymbol{\theta}} \in \mathbb{R}^m. \quad (8.7)$$

Alle nachfolgenden Beispiele zeigen, dass der Übergang zur sog. **Log-Likelihoodfunktion**

$$L(x_1, \dots, x_n; \boldsymbol{\theta}) = \sum_{i=1}^n \ln[f(x_i; \boldsymbol{\theta})] \quad (8.8)$$

große Erleichterungen bei der Berechnung von  $\hat{\boldsymbol{\theta}}_{ML}$  bringt. Da  $\ln(\cdot)$  streng monoton ist, stimmen die Maximalstellen von  $l$  und  $L$  überein. Generell gilt:

**Def. 8.6 (ML-Schätzer)**

Besitzen die iid-Variablen  $X, X_1, \dots, X_n$  eine Wahrscheinlichkeitsfunktion bzw. Dichte  $f(x; \boldsymbol{\theta}) = f_X(x; \boldsymbol{\theta})$ , so heißt jede Maximalstelle

$$\hat{\boldsymbol{\theta}}_{ML} = \operatorname{argmax}\{L(X_1, \dots, X_n; \tilde{\boldsymbol{\theta}}); \tilde{\boldsymbol{\theta}} \in \mathbb{R}^m\} = \operatorname{argmax}\{l(X_1, \dots, X_n; \tilde{\boldsymbol{\theta}}); \tilde{\boldsymbol{\theta}} \in \mathbb{R}^m\}$$

**ML- oder Maximum-Likelihood-Schätzer** für  $\boldsymbol{\theta}$ .  $\hat{\boldsymbol{\theta}}_{ML}$  muss nicht immer existieren; außerdem hat  $L$  gelegentlich mehrere Maximalstellen.

Gemäß dieser Definition ist der ML-Schätzer eine Zufallsvariable. Bei der praktischen Berechnung bestimmt man zunächst eine Formel  $T_{ML}$  bzw. eine Berechnungsvorschrift durch Nullsetzen der partiellen Ableitungen nach  $\theta_i$  bzw. über einen numerischen Optimierungsalgorithmus, wobei die beobachteten Realisierungen  $x_1, \dots, x_n$  als feste Konstanten angesehen werden. Im Beispiel der Exponentialverteilung erhält man

$$l(x_1, \dots, x_n; \boldsymbol{\theta}) = \prod_{i=1}^n \lambda e^{-\lambda x_i} 1_{(0, \infty)}(x_i), \quad \boldsymbol{\theta} = \lambda > 0. \quad (8.9)$$

Wegen  $x_i > 0, i = 1, \dots, n$  können die Indikatorfunktionen weggelassen werden. Damit lautet die Loglikelihoodfunktion

$$L(x_1, \dots, x_n; \boldsymbol{\theta}) = \sum_{i=1}^n (\ln(\lambda) - \lambda x_i) = n \ln(\lambda) - \lambda \sum_{i=1}^n x_i. \quad (8.10)$$

Nullsetzen der Ableitung nach  $\lambda$  liefert  $\lambda = T_{ML}(x_1, \dots, x_n) = 1/\bar{x}$ . Mit dieser Vorschrift lautet der ML-Schätzer als Zufallsvariable

$$\hat{\lambda} = T_{ML}(X_1, \dots, X_n) = 1/\bar{X}. \quad (8.11)$$

**Beispiel 8.7 (ML-Schätzer)**

1. Binomialverteilung  $B(n, p)$ :  $\bar{X}$  ist ML-Schätzer für  $p$
2. Poissonverteilung  $Poi(\lambda)$ :  $\bar{X}$  ist ML-Schätzer für  $\lambda$
3. Gleichverteilung in  $(0, b)$ :  $\hat{b}_{ML} = \max_i \{X_i\}$  ist ML-Schätzer für  $b$ .
4. Normalverteilung  $N(\mu, \sigma^2)$ :  $\bar{X}$  ist ML-Schätzer für  $\mu$

$$S_{ML}^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ ist ML-Schätzer für } \sigma^2.$$

Vergleicht man die Beispiele 8.5 und 8.7, so sind die MV-Schätzer entweder die ML-Schätzer selbst oder man erhält sie, im Falle von  $S^2$  bzw.  $\hat{b}$ , durch Multiplikation der entsprechenden ML-Schätzer  $S_{ML}^2$  bzw.  $\hat{b}_{ML}$  mit einer Normierungskonstanten, so dass

der neue Schätzer erwartungstreu ist.

Gemäß Satz 2.34 gilt  $E[S_{ML}^2] = \frac{n-1}{n}\sigma^2$ . Erwartungstreue erhält man durch Wahl eines Schätzers

$$T = S^2 = \frac{n}{n-1} S_{ML}^2.$$

Analog geht man bei der Schätzung für  $b$  im Falle der Gleich- oder Rechteckverteilung vor. Nach Satz 2.43 über die Erwartungswerte der Ordnungsstatistiken gilt nämlich

$$E[\hat{b}_{ML}] = E[\max\{X_i\}] = E[X_{n:n}] = \frac{n}{n+1}.$$

Dieses Normierungsverfahren liefert in einigen Fällen MV-Schätzer. Der Beweis minimaler Varianz muss jedoch jeweils im Einzelfall geführt werden.

# Kapitel 9

## Statistik normalverteilter Daten

### 9.1 Stichprobenverteilungen

#### 9.1.1 Normalverteilung

Die Dichte der Normal- oder Gauß-Verteilung  $N(\mu, \sigma^2)$  lautet

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (9.1)$$

$N(0, 1)$  heißt **Standardnormalverteilung**.

#### Grundlegende Eigenschaften

1.  $U \sim N(0, 1) \implies X := \mu + \sigma U \sim N(\mu, \sigma^2)$ . Durch die Standardisierung

$$U := \frac{X - \mu}{\sigma}$$

kann eine  $N(\mu, \sigma^2)$ -verteilte ZV  $X$  in eine  $N(0, 1)$ -verteilte ZV  $U$  transformiert werden.

2. Für die Standardnormalverteilungsfunktion

$$\Phi(u) = P(U \leq u) = \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

gibt es ausgezeichnete numerische Approximationen (siehe etwa Abschnitt 11.2 im Anhang). Wegen der Symmetrie der Gaußschen Glockenkurve  $\varphi$  gilt

$$\Phi(-u) = 1 - \Phi(u) \quad \text{und} \quad u_{1-p} = -u_p,$$

so dass  $\Phi$  nur für  $u \geq 0$  tabelliert werden muss.

3. Zum Ablesen der (einseitigen) Quantile  $u_p$  mit

$$\Phi(u_p) = p, \quad 0 < p < 1$$

im Lehr- und Übungsbetrieb verwendet man Tabellen (vgl. Abschnitt 11.3 im Anhang) von  $\Phi$ , die auch in den meisten Statistikbüchern zu finden sind.

4.  $E(X) = \mu, \text{Var}(X) = \sigma^2$
5. Alle Kumulanten (Semiinvarianten) ab der Ordnung 3 verschwinden, also u.a. Schiefe und Exzess.
6. Linearkombinationen gemeinsam normalverteilter ZV sind wieder normalverteilt. Dies gilt auch, falls die ZV korreliert sind. Speziell für unabhängige  $X \sim N(\mu_x, \sigma_x^2)$  und  $Y \sim N(\mu_y, \sigma_y^2)$  folgt

$$a + bX + cY \sim N(a + b\mu_x + c\mu_y, b^2\sigma_x^2 + c^2\sigma_y^2).$$

7. Schätzung der unbekannt Parameter  $\mu$  und  $\sigma^2$   
Seien  $x_1, \dots, x_n$  Realisierungen einer Stichprobe unabhängiger  $N(\mu, \sigma^2)$ -verteilter Zufallsvariablen, dann sind

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{und} \quad (9.2)$$

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (9.3)$$

optimale (im Sinne minimaler Varianz) Schätzungen für die unbekannt Parameter  $\mu$  und  $\sigma^2$ .

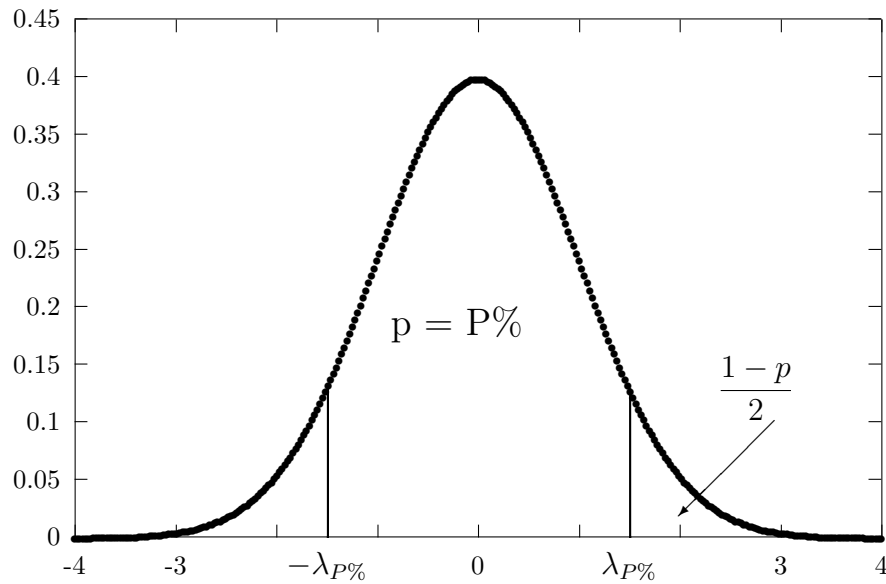
### Quantile der Standardnormalverteilung

Mit  $u_p, \Phi(u_p) = p, 0 < p < 1$ , bezeichnet man die (einseitigen) Quantile. Daneben benötigt man symmetrische Quantile der Form

$$\lambda_p = u_{(1+p)/2}, \quad \text{mit} \quad P(-\lambda_p \leq U \leq \lambda_p) = p. \quad (9.4)$$

Gebäuchlich sind auch die Bezeichnungen

$$u_p = u_{1-\alpha}, \quad \lambda_p = \lambda_{1-\alpha} = u_{1-\alpha/2}, \quad \text{mit} \quad \alpha = 1 - p.$$

Bild 22: Symmetrische Quantile  $\lambda_p$ 

Es folgen oft benützte Quantile der Normalverteilung:

Tabelle 12: $N(\mu, \sigma^2)$ , symmetrische Quantile $\lambda_{P\%}$		
Fläche $P\%$	symmetrisches Intervall $[\mu - \sigma \cdot \lambda_{P\%}; \mu + \sigma \cdot \lambda_{P\%}]$	symmetrisches Quantil $\lambda_{P\%}$
50%	$[\mu - 0.6745 \sigma; \mu + 0.6745 \sigma]$	0.6745
68.26%	$[\mu - \sigma; \mu + \sigma]$	1
95%	$[\mu - 1.96 \sigma; \mu + 1.96 \sigma]$	1.96
95.46%	$[\mu - 2 \sigma; \mu + 2 \sigma]$	2
99%	$[\mu - 2.576 \sigma; \mu + 2.576 \sigma]$	2.576
99.73%	$[\mu - 3 \sigma; \mu + 3 \sigma]$	3

Gemäß obiger Tabelle müssen ca. 95% normalverteilter Daten im sog.  $2\sigma$ -Intervall und über 99% im  $3\sigma$ -Intervall liegen.

### 9.1.2 $\chi^2$ -Verteilung

Die Verteilung der Summe der Quadrate von  $\nu$  unabhängigen  $N(0,1)$ -verteilten Zufallsvariablen  $U_1, \dots, U_\nu$  nennt man  $\chi^2$ -verteilt mit  $\nu$  Freiheitsgraden, d.h.

$$\chi_\nu^2 = \sum_{i=1}^{\nu} U_i^2.$$

Die  $\chi^2$ -Verteilungen sind spezielle Gammaverteilungen. Es gilt

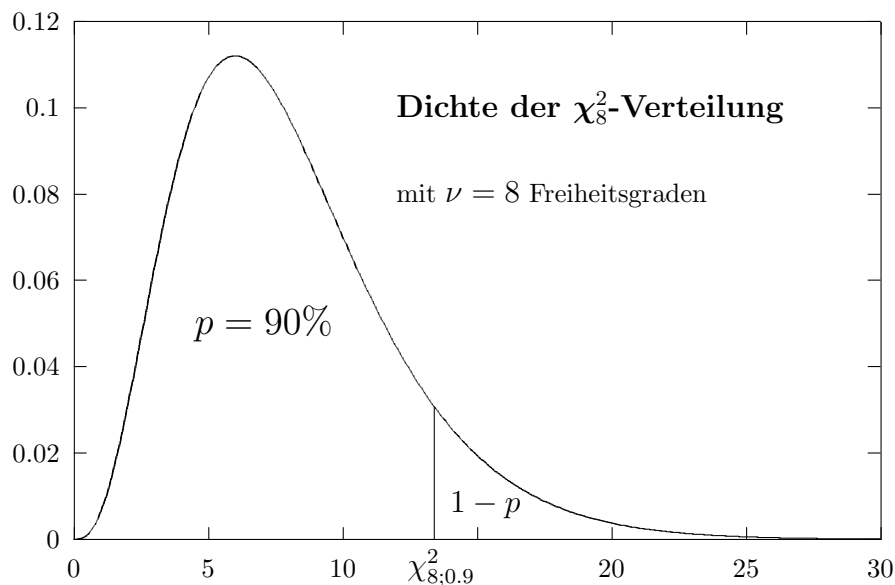
$$\chi_\nu^2 = \Gamma\left(\frac{\nu}{2}, \frac{1}{2}\right).$$

Damit lautet die Dichte der  $\chi^2$ -Verteilung mit  $\nu$  Freiheitsgraden

$$f_{\chi_\nu^2}(z) = \frac{1}{2^{\nu/2} \Gamma(\frac{\nu}{2})} z^{\nu/2-1} e^{-z/2} \mathbf{1}_{(0,\infty)}(z). \quad (9.5)$$

Gemäß Beispiel 1.58, Nr. 7, ist  $U_1^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$ . Mit der momenterzeugenden Funktion der Gammaverteilung aus den Beispielen 3.13 und 3.14 folgt die Behauptung für unabhängige  $U_1, \dots, U_\nu$ .  $\chi^2$ -verteilte Zufallsvariable können keine negativen Werte annehmen. Die Masse der  $\chi^2$ -Verteilung konzentriert sich um deren Erwartungswert  $E(\chi_\nu^2) = \nu$ , so dass für die häufig verwendeten 95%-Quantile  $\nu < \chi_{\nu,0.95}^2$  gilt (und zwar für alle  $\nu \in \mathbb{N}$ ) sowie weiterhin

$$\begin{aligned} E(\chi_\nu^2) &= \nu \\ \text{Var}(\chi_\nu^2) &= 2\nu \\ \text{Modus}(\chi_\nu^2) &= \nu - 2, \quad \nu \geq 2 \\ \text{Modus}(\chi_1^2) &\text{ existiert nicht.} \end{aligned}$$



**Bild 23:**  $\chi^2$ -Verteilung

### 9.1.3 Student- oder t-Verteilung

Seien  $U \sim N(0, 1)$  und  $\chi_\nu^2$  unabhängig. Dann heißt

$$T_\nu := \frac{U}{\sqrt{\chi_\nu^2/\nu}}$$

$t$ - oder Student-verteilt mit  $\nu$  Freiheitsgraden. Die zugehörige Dichte lautet

$$f_{T_\nu}(t) = \frac{1}{\sqrt{\nu} \pi} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\left(1 + \frac{t^2}{\nu}\right)^{(\nu+1)/2}}; \quad t \in \mathbb{R}. \quad (9.6)$$

Zur Herleitung der Dichte (vgl. etwa [Heinhold & Gaede (1979)], S. 235) benützt man die Transformationsregel für Quotienten aus (2.30) und (9.5). Für  $\nu = 1$  sind Zähler und Nenner unabhängig standardnormalverteilt, so dass man in diesem Fall die *Cauchy-Verteilung* mit der Dichte

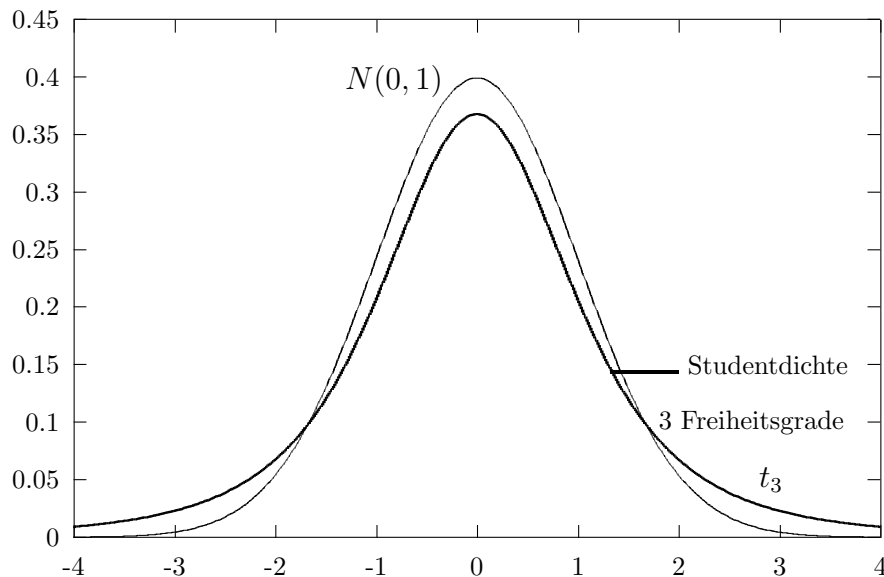
$$f_{T_1}(t) = \frac{1}{\pi} \frac{1}{1+t^2}$$

erhält. Analog zur Normalverteilung gilt für die Quantile der Student-Verteilung

$$P(T_\nu \leq t_{\nu;p}) = p \quad \text{und} \quad P(|T_\nu| \leq \gamma_{\nu;p}) = p,$$

wobei

$$t_{\nu;1-p} = -t_{\nu;p} \quad \text{und} \quad \gamma_{\nu;p} = t_{\nu;(1+p)/2}.$$



**Bild 24:**  $t$ - und Standardnormalverteilung

Für wachsenden Freiheitsgrad  $\nu \rightarrow \infty$  konvergiert die  $t$ -Verteilung mit ihren breiteren Enden gegen die Standardnormalverteilung, d.h.

$$\lim_{\nu \rightarrow \infty} t_{\nu;p} = u_p \quad \text{und} \quad \lim_{\nu \rightarrow \infty} \gamma_{\nu;p} = \lambda_p.$$



### 9.1.4 Stichprobenmittel und -varianz

Seien  $X_1, X_2, \dots, X_n$  iid  $N(\mu, \sigma^2)$ -verteilt. Gemäß Satz 2.34 über orthogonale Transformationen normalverteilter ZV gilt:

1. das Stichprobenmittel  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  und die  
 Stichprobenvarianz  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  sind unabhängig .
2. Beide Schätzer sind erwartungstreu, d.h.  $E(\bar{X}) = \mu$  und  $E(S^2) = \sigma^2$
3.  $\bar{X} \sim N(\mu, \sigma^2/n)$
4.  $(n-1) S^2/\sigma^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi_{n-1}^2$
5.  $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \bigg/ \sqrt{\frac{(n-1) S^2/\sigma^2}{n-1}} = \frac{\sqrt{n} (\bar{X} - \mu)}{S} \sim t_{n-1}$ .

Beim Vergleich der Streuungen in verschiedenen Gruppen benötigt man noch eine weitere wichtige Verteilung:

### 9.1.5 F-Verteilung

$X^2$  sei eine  $\chi^2$ -verteilte Zufallsgröße vom Freiheitsgrad  $m$  und  $Y^2$  eine davon unabhängige  $\chi^2$ -verteilte Zufallsgröße mit dem Freiheitsgrad  $n$ . Dann heißt die Verteilung der Zufallsgröße

$$F_{m,n} = \frac{X^2/m}{Y^2/n} \quad (9.7)$$

*F-Verteilung* mit dem *Zählerfreiheitsgrad*  $m$  und dem *Nennerfreiheitsgrad*  $n$ . Die Formel für die zugehörige Dichte entnehme man z.B. [Falk et al. (1995)], Satz 2.1.10. Eine Herleitung der Dichte, wieder über die Formel (2.30), findet man etwa in [Heinhold & Gaede (1979)], S. 247-248. Dort wird auch auf den Seiten 255-258 der enge Zusammenhang zur Betaverteilung hergestellt.

$$E(F_{m,n}) = \frac{n}{n-2}, \quad \text{für } n > 2.$$

$$\text{Var}(F_{m,n}) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, \quad \text{für } n > 4.$$

$$\text{Modus}(F_{m,n}) = \frac{n(m-2)}{m(n+2)}, \quad \text{für } n > 1 \text{ und } m > 2.$$

Die Definition der  $p$ -Quantile,  $0 < p < 1$ , lautet

$$P(F_{m,n} \leq F_{m,n;p}) = p.$$

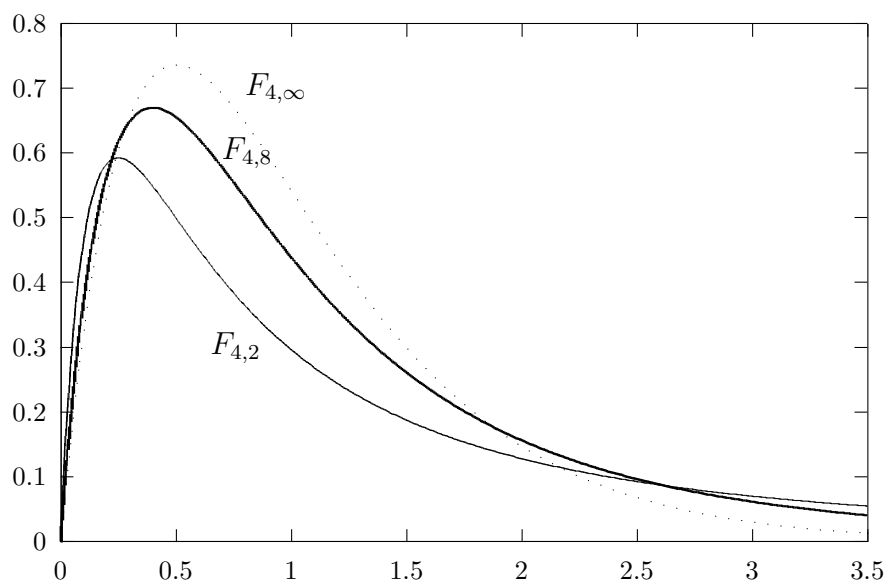
Außerdem gilt  $1/F_{m,n} = F_{n,m}$  und somit

$$P\left(\frac{1}{F_{m,n}} \geq F_{n,m;1-p}\right) = p \quad \text{oder} \quad P\left(F_{m,n} \leq \frac{1}{F_{n,m;1-p}}\right) = p,$$

woraus man

$$F_{m,n;p} = \frac{1}{F_{n,m;1-p}} \quad (9.8)$$

erhält. Diese Formel ermöglicht es, aus den  $p$ -Quantilen die dazu komplementären  $(1-p)$ -Quantile zu berechnen.



**Bild 25:** Dichten von  $F_{4,2}$ ,  $F_{4,8}$  und  $F_{4,\infty}$

Zusammenhang mit der  $t$ -Verteilung

$$\gamma_{n;1-\alpha} = t_{n;1-\alpha/2} = \sqrt{F_{1,n;1-\alpha}}, \quad 0 < \alpha < 1$$

oder

$$t_{n;p} = \sqrt{F_{1,n;2p-1}}, \quad \frac{1}{2} < p < 1.$$

Tabellen der Normal-,  $\chi^2$ -,  $t$ - und  $F$ -Verteilung findet man im Anhang, Abschnitt 11.3.

## 9.2 Konfidenzintervalle, Tests ( $\sigma^2$ bekannt)

Seien  $X_1, X_2, \dots, X_n$  iid  $N(\mu, \sigma^2)$ . Mit  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  ist  $U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ .  $u_p$  bezeichne die einseitigen und  $\lambda_p = u_{(1+p)/2}$  die zweiseitigen p-Quantile der Standardnormalverteilung. Die folgenden Aussagen beruhen auf

$$P\left(\frac{|\bar{X} - \mu|}{\sigma/\sqrt{n}} \leq \lambda_p\right) = P(|U| \leq \lambda_p) = P(-\lambda_p \leq U \leq \lambda_p) = p = P\% \quad (9.9)$$

oder entsprechenden einseitigen Ansätzen der Form  $P(U \geq -u_p)$  bzw.  $P(U \leq u_p)$ .

Tabellen für die Quantile der Standardnormalverteilung, der  $\chi^2$ -, der  $t$ - und der  $F$ -Verteilung findet man im Anhang sowie in den gängigen Statistik-Lehrbüchern.

### 9.2.1 Konfidenzintervalle

Aus dem Ansatz von (9.9) lassen sich sog. **Konfidenz-** oder **Vertrauensintervalle** ableiten; man sagt auch **Vertrauensbereiche** oder **Bereichsschätzungen** für den unbekanntem Erwartungswert  $\mu$ .

#### Def. 9.1 (Konfidenzintervall)

Mit den symmetrischen Quantilen  $\lambda_p$  von (9.4) und  $v_\sigma := \lambda_p \sigma / \sqrt{n}$  definiert man das (zufällige) **Konfidenzintervall**

$$KI_{n,p,\sigma} := [\bar{X} - v_\sigma, \bar{X} + v_\sigma] \subset \mathbb{R}. \quad (9.10)$$

Für dieses und  $0 < p < 1$  gilt

$$P(KI_{n,p,\sigma} \text{ enthält } \mu) = P(\bar{X} - v_\sigma \leq \mu \leq \bar{X} + v_\sigma) = p. \quad (9.11)$$

$$KI_{n,p,\sigma}(\omega) = [\bar{x} - v_\sigma, \bar{x} + v_\sigma] \subset \mathbb{R} \quad (9.12)$$

heißt **Realisierung** des Vertrauens- oder Konfidenzintervalls.

Entsprechend kann man einseitige Konfidenzintervalle, die nach oben oder unten unbeschränkt sind, definieren.

**Satz 9.2 (Notwendiger Stichprobenumfang)**

Oft ist man daran interessiert, dass das Konfidenzintervall höchstens eine Länge von  $l > 0$  hat. Für  $0 < p < 1$  muss dann gelten

$$2v_\sigma = \frac{2\sigma}{\sqrt{n}} \lambda_p \leq l.$$

Dies kann durch einen Stichprobenumfang von  $n \geq \frac{4\sigma^2 \lambda_p^2}{l^2}$  erreicht werden.

**Beispiel 9.3 (Porosität von Sandstein)**

Die Porosität  $X$ , d.h. der prozentuale Anteil des Porenraums eines Gesteins am Gesamtvolumen, sei  $N(\mu, \sigma^2)$ -verteilt,  $\sigma^2 = 9$ .  $n = 57$  Brocken wurden zufällig aus einer Gesteinsschicht herausgegriffen. Die Stichprobenrealisierung des Mittelwerts lieferte  $\bar{x} = 23.35\%$ ; vgl. [Kredler & Ritter (1995)].

Mit  $\lambda_{0.95} = 1.96$  und  $v_\sigma = \sigma \cdot \lambda_p / \sqrt{n} = 3 \cdot 1.96 / \sqrt{57} = 0.78$  ergibt sich die Realisierung des 95%-Konfidenzintervalls für  $\mu$  zu

$$[\bar{x} - v_\sigma, \bar{x} + v_\sigma] = [22.57, 24.13].$$

Mit einer Sicherheit von 95% enthält also das berechnete Intervall den unbekanntem Parameter  $\mu$ .

Um die Länge des Konfidenzintervalls auf  $l = 2v_\sigma = 1$  zu verkleinern, muss dann

$$n \geq \frac{4\sigma^2 \lambda_p^2}{l^2} = \frac{4 \cdot 9 \cdot 1.96^2}{1^2} = 138.3,$$

also  $n \geq 139$  gewählt werden.

**9.2.2 Tests für  $\mu$  bei bekanntem  $\sigma$** 

Das prinzipielle Vorgehen bei einem statistischen Test wird zunächst am einfachen Beispiel normalverteilter Daten mit bekannter Varianz erläutert. Dem praktisch wichtigen Fall mit unbekanntem  $\sigma^2$  ist der folgende Abschnitt gewidmet.

In vielen Fällen möchte man eine statistische Entscheidung nach dem Ausfall einer Stichprobe  $\mathbf{X} = (X_1, \dots, X_n)$  treffen. Ein Betonwerk produziere etwa zwei Sorten von Beton mit den mittleren Druckfestigkeiten von  $\mu_0 = 30 [N/mm^2]$  bzw.  $\mu_1 = 20 [N/mm^2]$ . Wir nehmen an, dass die Belastung  $X_i$  bis zum Bruch von Probewürfeln (z.B. von 20 cm Kantenlänge) aus einer Lieferung als normalverteilt mit Varianz  $\sigma^2 = 81 [N^2/mm^4]$  angesehen werden kann.  $\mathbf{x} = (x_1, \dots, x_n)$  sei die Realisierung einer Stichprobe  $\mathbf{X}$ . Für den Unterbau einer vielbefahrenen Autobahnstrecke benötigt

man unbedingt die bessere Betonsorte. Wir wissen, dass  $\bar{X} = \frac{1}{n} \sum X_i$  ein erwartungstreuer Schätzer mit minimaler Varianz für  $\mu$  ist. Falls also die Realisierung  $\bar{x}$  von  $\bar{X}$  kleiner ist als  $\mu_1$  bzw. größer als  $\mu_0$ , so wird man annehmen, dass die Sorte mit der geringeren bzw. mit der größeren Druckfestigkeit geliefert wurde. Schwieriger wird die Entscheidung schon, wenn etwa  $\bar{x} = 26$  gemessen wurde. Ein Instrumentarium für derartige Entscheidungen liefert die statistische Testtheorie.

Aufgrund des Ausfalls (Realisierung)  $\mathbf{x} = (x_1, \dots, x_n)$  einer Stichprobe  $\mathbf{X} = (X_1, \dots, X_n)$  soll eine Entscheidung (decision)  $d(\mathbf{x})$  getroffen werden zwischen einer **Hypothese**  $H_0$  (auch **Nullhypothese**) und einer **Alternative**  $H_1$  (auch **Gegenhypothese**). Wir beschränken uns zunächst auf den Fall, dass Hypothese und Alternative den unbekanntem Parameter  $\mu$  einer Normalverteilung betreffen und werden hier die beiden ersten der folgenden Beispiele erläutern

$$H_0 : \mu = \mu_0, \quad H_1 : \mu = \mu_1$$

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$$

$$H_0 : \mu = \mu_0, \quad H_1 : \mu > \mu_0$$

$$H_0 : \mu = \mu_0, \quad H_1 : \mu < \mu_0.$$

Abstrakt gesehen sind die Hypothese  $H_0$  und die Alternative  $H_1$  disjunkte Teilmengen des Parameterraums  $\Theta = \mathbb{R}$  für den unbekanntem Parameter  $\theta = \mu$ .

Alle Tests sind festgelegt durch einen **kritischen Bereich**  $K$  und ein **Signifikanzniveau**  $0 < \alpha < 1$ , z.B.  $\alpha = 0.1$ ,  $\alpha = 0.05$ ,  $\alpha = 0.01$ .  $P\% = p = 1 - \alpha$  nennt man **Sicherheitswahrscheinlichkeit**. Bei der folgenden Definition beschränken wir uns auf den parametrischen, eindimensionalen Fall.

#### Def. 9.4 (Parametrischer, statistischer Test)

$\mathbf{X} = (X_1, \dots, X_n)$  sei eine iid-Stichprobe mit Realisierung  $\mathbf{x} = (x_1, \dots, x_n)$ . Die Verteilungsfunktion  $F$  jedes der  $X_i$ ,  $i = 1, \dots, n$ , hänge (u.a.) von einem unbekanntem Parameter  $\theta \in \Theta$  ab. Die Hypothese  $H_0$  und die Alternative  $H_1$  seien disjunkte Teilmengen von  $\Theta$ .

Ein **statistischer Test** zum Signifikanzniveau  $0 < \alpha < 1$  ist eine Entscheidungsfunktion  $d$  vom Stichprobenraum (hier  $\mathbb{R}^n$ ) nach  $\{H_0, H_1\}$ , so dass

$$d(\mathbf{x}) = \begin{cases} H_1 & , \text{ falls } \mathbf{x} \in K \\ H_0 & , \text{ sonst,} \end{cases}$$

wobei der **kritische Bereich**  $K$  als Teilmenge des Stichprobenraumes so festgelegt ist, dass

$$P(d(\mathbf{X}) = H_1 \mid \theta \in H_0) = P(\mathbf{X} \in K \mid \theta \in H_0) \leq \alpha. \quad (9.13)$$

Offenbar sind zwei Arten von Fehlentscheidungen möglich

$d(\mathbf{x}) = H_1$ , aber $H_0$ ist richtig	<b>Fehler 1. Art</b>	$P(\text{Fehler 1. Art}) = \alpha$
$d(\mathbf{x}) = H_0$ , aber $H_1$ ist richtig	<b>Fehler 2. Art</b>	$P(\text{Fehler 2. Art}) = \beta$

Während die Wahrscheinlichkeit für den Fehler 1. Art durch die Wahl eines kleinen  $\alpha$ , z.B.  $\alpha = 0.05$  kontrolliert wird, ist die Wahrscheinlichkeit für den Fehler 2. Art bereits in den einfachsten Fällen nur sehr mühsam zu ermitteln; vgl. Beispiel unten.

### Konstruktion des kritischen Bereichs $K$

Wir demonstrieren die Festlegung des kritischen Bereichs am Beispiel

$$H_0 : \mu = \mu_0, \quad H_1 : \mu = \mu_1 \quad (\mu_1 < \mu_0).$$

Definition 9.4 legt den kritischen Bereich  $K$  nicht eindeutig fest. Für spezielle Verteilungen, u.a. die Normalverteilung, läßt sich der - hinsichtlich eines gleichmäßig kleinen Fehlers 2. Art - optimale kritische Bereich  $K$  einfach beschreiben. Für die hier zu untersuchende Alternative gilt  $K = \{ \mathbf{x} \in \mathbb{R}^n \mid \bar{x} < c \}$ . Unter der Hypothese  $H_0$  ist

$$U = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

standardnormalverteilt, wobei

$$\begin{aligned} \alpha &= \Phi(u_\alpha) = \Phi(-u_{1-\alpha}) = P(U \leq -u_{1-\alpha}) \\ &= P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq -u_{1-\alpha}\right) = P\left(\bar{X} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} u_{1-\alpha}\right). \end{aligned}$$

Also

$$K = \left\{ \mathbf{x} \in \mathbb{R}^n \mid \bar{x} < \mu_0 - \frac{\sigma}{\sqrt{n}} u_{1-\alpha} \right\},$$

was man übrigens auch bei Wahl der Alternative  $\mu < \mu_0$  erhalten hätte. Die Komplementärmenge  $\bar{K} = \{ \mathbf{x} \in \mathbb{R}^n \mid \bar{x} \geq \mu_0 - \frac{\sigma}{\sqrt{n}} u_{1-\alpha} \}$  heißt **Annahmehereich**.

Der kritische Bereich hängt also nur über die Funktion  $\bar{X}$  von  $X_1, \dots, X_n$  ab. In diesem Fall nennt man die Schätzfunktion  $\bar{X}$  (für  $\mu$ ) eine **Teststatistik**.

Seien im Beispiel  $n = 9$  Betonwürfel mit  $\bar{x} = 26$  abgedrückt worden und  $\alpha = 0.05$ . Es ergibt sich  $\sigma = 9$ ,  $u_{1-\alpha} = u_{0.95} = 1.64$  und

$$K = \{ \mathbf{x} \in \mathbb{R}^n \mid \bar{x} < c \} \quad \text{mit} \quad c = 30 - \frac{9}{\sqrt{9}} 1.64 = 25.08.$$

Da  $\bar{x} = 26 \geq 25.08 = c$ , ist die Hypothese  $H_0$  beizubehalten. In analoger Weise erhält man den kritischen Bereich  $K = \{ \mathbf{x} \in \mathbb{R}^n \mid \bar{x} > c \}$  für einpunktige Alternativen  $\mu_1 > \mu_0$  bzw. für  $H_1 : \mu > \mu_0$ .

### 9.2.3 Fehler 2. Art

Bisher wurde von der Alternative nur die Tatsache  $\mu_1 < \mu_0$  benützt. Der genaue Wert von  $\mu_1$  geht in den Fehler 2. Art ein. Unter  $H_1$  ist

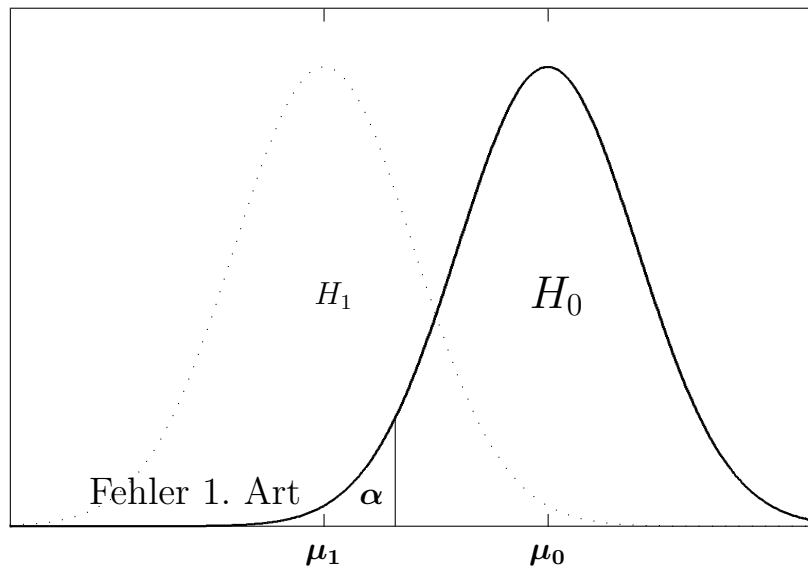
$$U = \frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}}$$

standardnormalverteilt. Wurde  $H_0$  nicht abgelehnt, so gilt

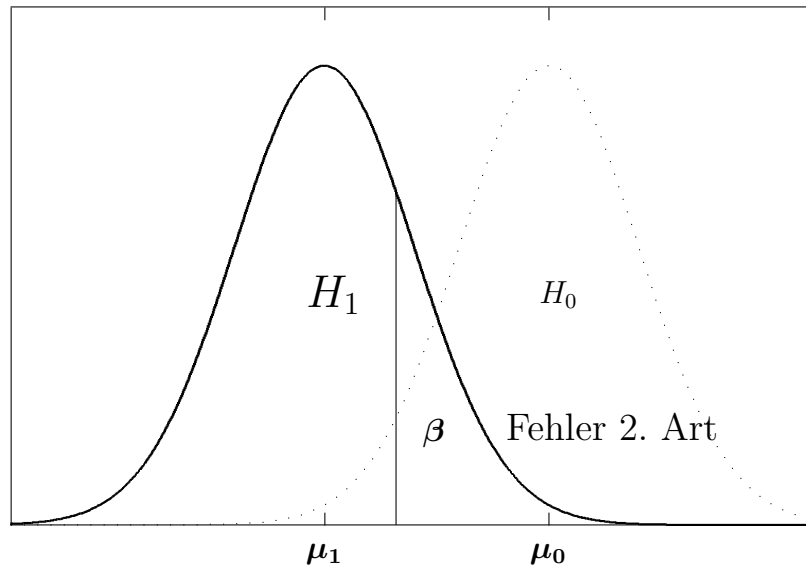
$$\begin{aligned} \beta = \beta(\mu_1) &= P(\text{Fehler 2. Art}) = P(d(\mathbf{X}) = H_0 \mid H_1) \\ &= P(\bar{X} \geq c \mid H_1) = P\left(\frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} \geq \frac{c - \mu_1}{\sigma/\sqrt{n}} \mid H_1\right) \\ &= 1 - \Phi\left(\frac{c - \mu_1}{\sigma/\sqrt{n}}\right) = 1 - \Phi\left(\frac{\mu_0 - u_{1-\alpha}\sigma/\sqrt{n} - \mu_1}{\sigma/\sqrt{n}}\right) \\ &= 1 - \Phi\left(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - u_{1-\alpha}\right) = \Phi\left(u_{1-\alpha} - \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right). \end{aligned}$$

Im Beispiel ist die Wahrscheinlichkeit, dass wir uns fälschlicherweise für  $H_0$  entschieden haben

$$P(\text{Fehler 2. Art}) = \Phi\left(1.64 - \frac{30 - 20}{9/\sqrt{9}}\right) = \Phi(-1.69) = 1 - \Phi(1.69) = 0.0455 = 4.55\%.$$



kritischer Bereich  $K \leftarrow | \rightarrow$  Annahmehereich



**Bild 26:** Fehler 1. und 2. Art



### 9.2.4 Zweiseitiger Test

Beim Test

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$$

ist  $H_0$  beim Niveau  $\alpha$  abzulehnen, falls  $\bar{x}$  "zu stark" von  $\mu_0$  nach oben bzw. nach unten abweicht. Das liefert den kritischen Bereich

$$\begin{aligned} K &= \left\{ \mathbf{x} \in \mathbb{R}^n \mid \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} > \lambda_{1-\alpha} \right\} \\ &= \left\{ \mathbf{x} \in \mathbb{R}^n \mid \bar{x} < \mu_0 - \frac{\sigma}{\sqrt{n}} \lambda_{1-\alpha} \text{ oder } \bar{x} > \mu_0 + \frac{\sigma}{\sqrt{n}} \lambda_{1-\alpha} \right\}. \end{aligned} \quad (9.14)$$

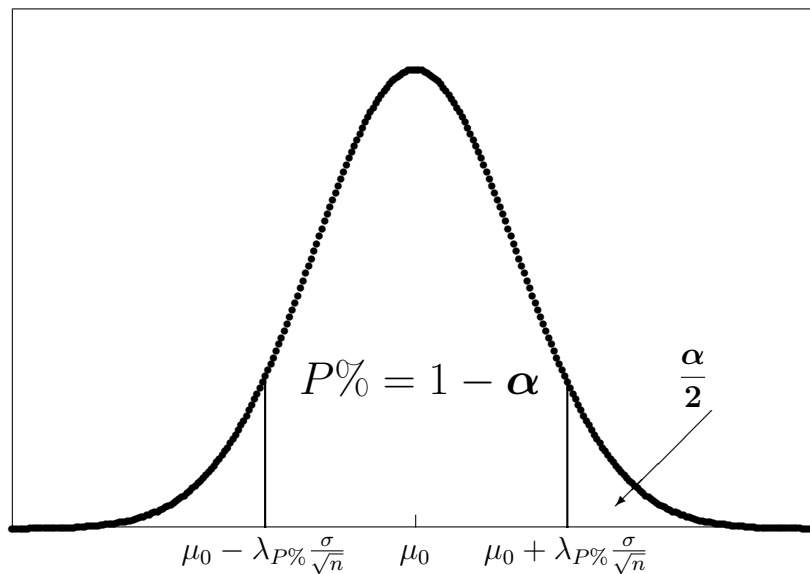


Bild 27: Zweiseitiger Test

Wiederum wird der Nullhypothese durch die Wahl eines kleinen  $\alpha$  aus  $[0.01, 0.1]$  eine große Chance eingeräumt, nicht abgelehnt zu werden. Falls aber  $\bar{x}$  in einen der nur mit  $\alpha/2$  Wahrscheinlichkeit auftretenden Außenbereiche fällt, so argumentiert man: "Wir werden doch bei der Stichprobe nicht gerade eine extrem seltene Beobachtung gezogen haben. Vielmehr liegt die Vermutung nahe, dass die Nullhypothese  $\mu = \mu_0$  verletzt ist."

Aus didaktischen Gründen wurde zunächst  $\sigma^2$  als bekannt vorausgesetzt. Damit waren auch kompliziertere Berechnungen wie die für Wahrscheinlichkeiten bzgl. des Fehlers 2. Art explizit und elementar durchführbar. Nun wollen wir den realistischen Fall betrachten, dass neben dem Mittelwert  $\mu$  auch die Varianz  $\sigma^2$  aus den **Daten**  $x_1, \dots, x_n$  zu schätzen ist.

## 9.3 Konfidenzintervalle und t-Test; $\sigma$ unbekannt

$X_1, X_2, \dots, X_n$  seien wieder iid  $N(\mu, \sigma^2)$ . Gemäß Abschnitt 9.1.4 sind

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

unabhängig und normal- bzw.  $\chi_{n-1}^2$ -verteilt. Weiterhin ist

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \quad t_{n-1} \text{ - oder Student-verteilt mit } n-1 \text{ Freiheitsgraden.}$$

### 9.3.1 Quantile der t-Verteilung

$u_p$  und  $\lambda_p$  seien wieder das p-Quantil bzw. das symmetrische p-Quantil der Standardnormalverteilung. Analog definiert man über

$$P(T \leq t_{\nu;p}) = p$$

das p-Quantil  $t_{\nu;p}$  der t-Verteilung mit  $\nu$  Freiheitsgraden.  $\gamma_{\nu;p}$  bezeichne das symmetrische p-Quantil der t-Verteilung. Also  $\gamma_{\nu;p} = t_{\nu;(1+p)/2}$ .

Gemäß Abschnitt 9.1.3 konvergiert für wachsenden Freiheitsgrad  $\nu \rightarrow \infty$  die t-Verteilung mit ihren breiteren Enden gegen die Standardnormalverteilung, d.h.

$$\lim_{\nu \rightarrow \infty} t_{\nu;p} = u_p \quad \text{und} \quad \lim_{\nu \rightarrow \infty} \gamma_{\nu;p} = \lambda_p.$$

Wir gehen nun analog zu den vorhergehenden Abschnitten bei bekanntem  $\sigma^2$  vor. Die folgenden Aussagen beruhen auf

$$P(|T| \leq \gamma_{\nu;p}) = p = P\%.$$

### 9.3.2 Konfidenzintervalle für $\mu$

$$P(\bar{X} - v \leq \mu \leq \bar{X} + v) = P\%,$$

$$v = \gamma_{n-1;p} \frac{S}{\sqrt{n}}, \quad \text{falls } \sigma^2 \text{ unbekannt.}$$

Um das Konfidenzintervall  $[\bar{x} - v, \bar{x} + v]$  auf eine gewünschte Länge  $l$  zu reduzieren, kann der **notwendige Stichprobenumfang**  $n_{neu}$  für eine "Ergänzungsstichprobe" näherungsweise folgendermaßen berechnet werden

$$n_{neu} + n \geq \frac{4s^2 \gamma_{n-1;p}^2}{l^2},$$

wobei  $s^2$  die empirische Varianz der bereits vorliegenden Stichprobe vom Umfang  $n$  ist, und  $p = P\%$  die gewählte Sicherheitswahrscheinlichkeit darstellt. Im Beispiel 9.3 (Sandsteinporosität) ergibt sich für  $n = 57$  mit der Stichprobenrealisierung  $s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 = 9$  bei einer Sicherheitswahrscheinlichkeit von 95%

$$v = \gamma_{n-1;p} \frac{s}{\sqrt{n}} = 2.00 \cdot 3/\sqrt{57} = 0.795$$

und die Realisierung des 95%-Vertrauensintervalls für  $\mu$ :

$$[22.56, 24.15],$$

die etwas breiter ausfällt als bei bekanntem  $\sigma^2$ .

Der notwendige Stichprobenumfang einer Ergänzungsstichprobe errechnet sich wiederum für  $l = 1$  aus

$$n_{neu} + n \geq \frac{4s^2\gamma_{n-1;p}^2}{l^2} = \frac{4 \cdot 9 \cdot 4.00}{1^2} = 144.$$

Bei unbekanntem  $\sigma^2$  müssen also  $n_{neu} = 87$  zusätzliche Stichprobenziehungen erfolgen. Insgesamt hat man 5 Elemente mehr zu ziehen (dort waren nur 139 notwendig) als unter gleichen Voraussetzungen bei bekanntem  $\sigma^2$ .

### 9.3.3 Konstruktion von $T$ -Statistiken, t-Tests

Bevor wir uns den wichtigen t-Tests zuwenden, soll nochmals der hier gewählte Ansatz unter einem anderen Blickwinkel, der sich weithin verallgemeinern läßt, betrachtet werden.

Gegeben:  $Y_1, \dots, Y_n \sim N(\mu, \sigma^2)$ , unabhängig

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad E(\bar{Y}) = \mu, \quad D(\bar{Y}) = \sigma/\sqrt{n} \quad (\text{Standardabweichung})$$

Nicht zugänglich:  $U = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$ , da  $\sigma^2$  unbekannt

Ersatz:  $T = \frac{\bar{Y} - \mu}{S/\sqrt{n}}$ , wobei  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$  Schätzer für  $\sigma^2$

Abstraktion:  $\bar{Y} = \hat{\mu}$  (Schätzer für  $\mu$ );  $\bar{Y}$  normalverteilt

$$S_{\hat{\mu}}^2 = S^2/n = \text{Schätzer für Varianz von } \bar{Y} = \hat{\mu}; \quad S_{\hat{\mu}}^2 \sim \chi^2\text{-verteilt}$$

$$\begin{aligned} T &= \frac{\bar{Y} - \mu}{S/\sqrt{n}} \\ &= \frac{\hat{\mu} - \mu}{S_{\hat{\mu}}} = \frac{\text{Schätzer} - (\text{unbekannter Parameter})}{\text{Schätzung für Standardabweichung des Schätzers}} \end{aligned}$$

$T \sim t_{n-1}$ , d.h.  $T$  ist Student-verteilt mit  $n-1$  Freiheitsgraden. Bei Verallgemeinerungen ist darauf zu achten, dass der Schätzer (hier  $\bar{Y}$ ) und der Schätzer für die Standardabweichung des Schätzers (hier  $S/\sqrt{n}$ ) unabhängig sind.

**Merke:**

**Freiheitsgrade = n - (für  $S^2$  benötigte Parameter; z.B.  $\bar{X}$  für  $\mu$ )**

Wir behandeln stets Nullhypothesen der Art

$$H_0 : \mu = \mu_0,$$

mit normalverteilten, erwartungstreuen Schätzern  $\hat{\mu}$ . Falls die Hypothese  $H_0$  richtig ist, lautet die Prüfgröße mit den Bezeichnungen des Abschnitts vorher

$$T = \frac{\hat{\mu} - \mu_0}{S_{\hat{\mu}}}.$$

Je nach Alternative ist die Hypothese  $H_0$  beim Signifikanzniveau  $\alpha$  abzulehnen, falls für die Stichprobenrealisierung  $t$  von  $T$  gilt

Alternative	Testvorschrift
$H_1 : \mu \neq \mu_0$	$ t  > \gamma_{n-1;1-\alpha} = t_{n-1;1-\alpha/2}$
$H_1 : \mu < \mu_0$	$t < t_{n-1;\alpha}$
$H_1 : \mu > \mu_0$	$t > t_{n-1;1-\alpha}$

Die Anzahl der Freiheitsgrade ist im einfachsten Fall  $\nu = n - 1$ .

Redeweisen:

Man sagt, der unbekannte Parameter  $\mu$  sei für  $\alpha = 0.05$  signifikant und für  $\alpha = 0.01$  hochsignifikant von  $\mu_0$  verschieden, falls die Hypothese  $H_0$  abgelehnt wird.

Merke:

Der t-Test ist robust gegen kleine Abweichungen von der Normalverteilungsannahme.

### 9.3.4 t-Test verbundener Stichproben; matched pairs

Beim sog. Einstichproben t-Test (Student's Test) geht es um die Wirkung  $W$  eines Einflusses, beispielsweise eines neuen Medikaments, auf  $n$  Probanden. Dabei wird bei jedem der  $n$  Individuen eine gewisse Kenngröße  $X$  vor dem Eintreten von  $W$  gemessen (z.B. Cholesteringehalt vor Einnahme des neuen Medikaments  $W$ ). Nach dem Einfluss  $W$  (z.B. Einnahme des Medikaments  $W$ ) wird die Kenngröße  $\tilde{X}$  bei demselben Individuum nochmals erhoben (z.B. Cholesteringehalt nach Einnahme von  $W$ ). Man erhält eine Folge von **matched pairs**

$$D = \left[ \begin{pmatrix} X_1 \\ \tilde{X}_1 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ \tilde{X}_n \end{pmatrix} \right],$$

z.B.  $X_i$  Gewicht Patient  $i$  vor Diätkur;  $\tilde{X}_i$  Gewicht Patient  $i$  nach Diätkur. Zwar sind  $\tilde{X}_i$  und  $X_i$  nicht unabhängig. Man kann aber in vielen Fällen davon ausgehen, dass die Unabhängigkeitsannahme für die Differenzen  $Y_i = X_i - \tilde{X}_i$  zutrifft. Falls die  $Y_i$  zusätzlich  $N(\mu, \sigma^2)$ -verteilt sind, dann lautet die Testgröße oder Teststatistik

$$T = \frac{\bar{Y} - \mu}{S_{\bar{Y}}} \sim t_{n-1}, \quad S_{\bar{Y}}^2 = \frac{S^2}{n} = \frac{1}{n(n-1)} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

$T$  ist also  $t$ -verteilt mit  $n - 1$  Freiheitsgraden.

### Beispiel 9.5 (Diätkur; matched pairs)

Neun Erwachsene testen eine neue Diätkur. Ihr Gewicht in Pfund betrug:

Tabelle 13: Diätkur; verbundene Stichproben									
	1	2	3	4	5	6	7	8	9
vorher:	132	139	126	114	122	132	142	119	126
nachher:	124	141	118	116	114	132	137	122	121
Differenz $y_i$	8	-2	8	-2	8	0	5	-3	5

Es werde angenommen, dass die Gewichte der einzelnen Personen vorher und nachher normalverteilte, verbundene Stichproben sind, d.h. die Differenzen  $Y_i$  sind iid  $N(\mu, \sigma^2)$ . Beim Niveau  $\alpha = 0.05$  ist die Nullhypothese, dass die Diät das erwartete Gewicht der einzelnen Personen nicht ändert, gegen die Alternative zu testen, dass das erwartete Gewicht verringert wird.

$$n = 9, \quad \alpha = 0.05, \quad p = 1 - \alpha = 0.95, \quad t_{8;0.95} = 1.86, \quad \gamma_{8;0.95} = t_{8;0.975} = 2.31$$

$$\bar{y} = 3, \quad s^2 = \frac{1}{9-1} \sum (y_i - \bar{y})^2 = 22.25, \quad s = 4.717.$$

Die Realisierung der Teststatistik  $T$  für die Hypothese  $H_0 : \mu = \mu_0 = 0$  lautet

$$t_0 = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{3 - 0}{4.717/\sqrt{9}} = 1.908.$$

Durchführung der Tests:

Hypothese	Alternative	Testvorschrift	Ergebnis
$H_0 : \mu = 0$	$H_1 : \mu \neq 0$	$ t_0  = 1.908 \leq 2.31 = \gamma_{n-1;p}$	$H_0$ beibehalten
$H_0 : \mu = 0$	$H_1 : \mu > 0$	$t_0 = 1.908 > 1.86 = t_{n-1;p}$	$H_0$ ablehnen

Während die Signifikanz der Daten für eine Ablehnung der zweiseitigen Alternative  $\mu \neq 0$  nicht ausreicht, kann die einseitige Gegenhypothese  $\mu > 0$  beim Niveau  $\alpha = 0.05$  abgelehnt werden. Im zweiten Fall liegt also eine signifikante Gewichtsabnahme durch die Diätkur vor.

### 9.3.5 t-Test zum Vergleich von Mittelwerten

Das folgende Vorgehen wird auch **Zweistichproben t-Test** genannt, da es sich wie bei Abschnitt 9.4 um den Vergleich zweier unabhängiger Stichproben handelt. Man betrachtet

$$\begin{aligned} X_1, \dots, X_n & \text{ iid } N(\mu_x, \sigma_x^2) \quad (\text{Gruppe 1}) \\ Y_1, \dots, Y_m & \text{ iid } N(\mu_y, \sigma_y^2) \quad (\text{Gruppe 2}), \end{aligned}$$

wobei  $X_1, \dots, X_n; Y_1, \dots, Y_m$  als unabhängig vorausgesetzt werden. Damit sind auch die abgeleiteten Größen

$$\bar{X} \quad \text{und} \quad \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{sowie} \quad \bar{Y} \quad \text{und} \quad \sum_{j=1}^m (Y_j - \bar{Y})^2 \quad (9.15)$$

unabhängig. Unter der Annahme

$$\sigma_x^2 = \sigma_y^2 = \sigma^2 \quad (\text{homogene Varianzen}) \quad (9.16)$$

kann die **gemeinsame ("pooled") Varianzschätzung**

$$S^2 = \frac{1}{n+m-2} \left[ \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2 \right]$$

verwendet werden. Der Fall

$$\sigma_x^2 = k^2 \sigma_y^2 = k^2 \sigma^2, \quad k \neq 0 \quad (\text{heterogene Varianzen}) \quad (9.17)$$

wird mit der Transformation

$$\tilde{X} := \frac{1}{k} X, \quad \tilde{Y} := Y \quad (9.18)$$

auf die Situation homogener Varianzen von (9.16) zurückgeführt. Auch beim Ansatz (9.17) ist noch der unbekannte Faktor  $k$  zu berücksichtigen. Bei dessen Wahl kann z.B. der F-Test von Abschnitt 9.4 hilfreich sein. Einen weiteren Zugang liefert der sog. Behrens-Fisher-Ansatz; vgl. etwa [Falk et al. (1995)], S. 63.

Mit der Voraussetzung homogener Varianz von (9.16) ist

$$T = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{S \sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{\hat{\mu}_x - \hat{\mu}_y - (\mu_x - \mu_y)}{S_{\hat{\mu}_x - \hat{\mu}_y}} \quad (9.19)$$

$t_{n+m-2}$ -verteilt. Wegen (9.15) sind nämlich  $\bar{X}, \bar{Y}$  und  $S^2$  unabhängig. Weiterhin sind nach dem Additionstheorem der Normalverteilung unter der Hypothese (9.20)

$$(\bar{X} - \mu_x) - (\bar{Y} - \mu_y) \sim N\left(0, \left(\frac{1}{n} + \frac{1}{m}\right) \sigma^2\right)$$

und nach dem Additionstheorem der  $\chi^2$ -Verteilung

$$\frac{1}{\sigma^2} \left[ \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2 \right] \sim \chi_{n+m-2}^2.$$

Wiederum bestätigt sich hier die **Faustregel**:

Freiheitsgrade = Anz. Beobachtungen  $(n + m)$  - Anz. geschätzte Parameter  $(\mu_x, \mu_y)$ .  
 $\sigma^2$  geht nicht in die  $\chi^2$ -Größe ein, zählt also nicht.

Unter der zusätzlichen Hypothese

$$H_0 : \mu_x = \mu_y \quad (9.20)$$

ist die Prüfgröße

$$T_0 = \frac{\bar{X} - \bar{Y}}{S\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

$t_{n+m-2}$ -verteilt.

### Beispiel 9.6 (Druckfestigkeit von zwei Betonsorten )

Gemäß [Heinhold & Gaede (1979)], S. 243) seien die Druckfestigkeiten in  $[N/mm^2]$  von zwei Betonsorten  $N(\mu_x, \sigma^2)$  bzw.  $N(\mu_y, \sigma^2)$ -verteilt. Zwei unabhängige Stichproben  $X_1, \dots, X_n$  der Sorte 1 und  $Y_1, \dots, Y_m$  der Sorte 2 ergaben folgende Realisierungen

Tabelle 14: Vergleich von Betonsorten											
$x_i$	14.2	14.7	13.9	14.3	13.7	13.7	14.2	14.5	14.0	14.3	$i=1, \dots, n = 10$
$y_j$	14.7	15.2	15.0	14.9	15.4	14.6	15.0	15.2	-	-	$j=1, \dots, m = 8$

Für den Test der Hypothese  $\mu_x = \mu_y$  gegen die Alternative  $\mu_x \neq \mu_y$  ergibt sich

$$\bar{x} = 14.15, \quad \bar{y} = 15;$$

und unter der Annahme homogener Varianzen

$$s^2 = \frac{1}{10 + 8 - 2} \left[ \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{j=1}^m (y_j - \bar{y})^2 \right] = 0.09156, \quad s = 0.3026,$$

erhält man

$$|t_0| = \frac{|\bar{x} - \bar{y}|}{s\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{|14.15 - 15|}{0.3026\sqrt{\frac{1}{10} + \frac{1}{8}}} = 5.922.$$

Da  $5.922 > \gamma_{16;0.99} = 2.921$ , ist die Hypothese  $\mu_x = \mu_y$  sogar beim Signifikanzniveau  $\alpha = 0.01$  abzulehnen. Die Mittelwerte der beiden Betonsorten sind also **hochsignifikant** verschieden.

## 9.4 Vergleich von Varianzen, F-Test

Beim Beispiel (9.6) wurde die Annahme homogener Varianz (9.16) in den beiden zu untersuchenden Gruppen nicht anhand der Daten verifiziert. Eine Möglichkeit dazu bietet der Vergleich von Varianzen mit der F-Verteilung von Abschnitt 9.1.5. Analog zu Abschnitt 9.3.5 seien  $X_1, \dots, X_n$  bzw.  $Y_1, \dots, Y_m$  unabhängig und  $N(\mu_x, \sigma_x^2)$  bzw.  $N(\mu_y, \sigma_y^2)$ -verteilt. Damit sind

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

bzw.

$$\frac{m-1}{\sigma_y^2} S_y^2 = \frac{1}{\sigma_y^2} \sum_{j=1}^m (Y_j - \bar{Y})^2$$

$\chi_{n-1}^2$ - bzw.  $\chi_{m-1}^2$ -verteilt und unabhängig. Also ist gemäß der Definition von (9.7)

$$F := \frac{S_x^2 / \sigma_x^2}{S_y^2 / \sigma_y^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 / (n-1) / \sigma_x^2}{\sum_{j=1}^m (Y_j - \bar{Y})^2 / (m-1) / \sigma_y^2} \sim F_{n-1, m-1}. \quad (9.21)$$

Mit der Hypothese

$$\tilde{H}_0 : \sigma_x = \sigma_y \iff \frac{\sigma_x^2}{\sigma_y^2} = 1 \quad (9.22)$$

gilt

$$F_0 := \frac{S_x^2}{S_y^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)}{\sum_{j=1}^m (Y_j - \bar{Y})^2 / (m-1)} \sim F_{n-1, m-1}. \quad (9.23)$$

$F_0$  ist also wie  $F$  von (9.21) F-verteilt mit Zählerfreiheitsgrad  $n-1$  und Nennerfreiheitsgrad  $m-1$ .  $F_{n-1, m-1; p}$  bezeichne das einseitige  $p$ -Quantil der entsprechenden F-Verteilung. Damit gilt für  $0 < \alpha < 1$

$$P\left(F_{n-1, m-1; \alpha/2} \leq \frac{S_x^2}{S_y^2} \leq F_{n-1, m-1; 1-\alpha/2}\right) = 1 - \alpha. \quad (9.24)$$

Wählt man als kritischen Bereich das Komplement des Intervalls von (9.24), dann ist die Hypothese  $\tilde{H}_0$  beim Niveau  $\alpha$  abzulehnen, falls für die entsprechenden Realisierungen  $s_x^2$  bzw.  $s_y^2$  von  $S_x^2$  bzw.  $S_y^2$  gilt:

$$\frac{s_x^2}{s_y^2} < F_{n-1, m-1; \alpha/2} \quad \text{oder} \quad \frac{s_x^2}{s_y^2} > F_{n-1, m-1; 1-\alpha/2}.$$

### Beispiel 9.7 (Betonarten; Fortsetzung)

Wir greifen die Untersuchung der Druckfestigkeit zweier Betonarten wieder auf; siehe auch [Heinhold & Gaede (1979)], S. 252. Mit den Daten von Beispiel 9.6 gilt  $n = 10$



und  $m = 8$ . Bei einem Niveau  $\alpha = 0.05$ , d.h.  $p = 0.95$  und  $(1 + p)/2 = 0.975$  liest man in den F-Tabellen zu den Freiheitsgraden  $9 = n - 1$  und  $7 = m - 1$  ab:

$$F_{9,7;0.975} = 4.82, \quad F_{9,7;0.025} = \frac{1}{F_{7,9;0.975}} = \frac{1}{4.20} = 0.238,$$

wobei für  $0 < p < 1/2$  die Umrechnungsformel (9.8) benutzt wurde. Ferner erhält man aus den Daten für  $S_x^2$  bzw.  $S_y^2$  die Realisierungen

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 10.72 \quad \text{bzw.} \quad s_y^2 = \frac{1}{m-1} \sum_{j=1}^m (y_j - \bar{y})^2 = 7.14.$$

Die Hypothese  $\tilde{H}_0$  kann also beibehalten werden, da

$$F_{9,7;0.025} = 0.238 \leq \frac{s_x^2}{s_y^2} = 1.50 \leq 4.82 = F_{9,7;0.975}.$$

Die Annahme homogener Varianz für den vergleichenden Zweistichprobentest von Beispiel 9.6 war also gerechtfertigt.

Wäre die Hypothese  $\tilde{H}_0$  verworfen worden, so kann das symmetrische  $(1-\alpha)$ -Konfidenzintervall Anhaltspunkte für die Wahl des Faktors  $k$  von (9.17) liefern. Aus (9.21) ergibt sich für  $p = 1 - \alpha$

$$P \left( \frac{S_y^2}{S_x^2} F_{n-1, m-1; (1-p)/2} \leq \frac{\sigma_y^2}{\sigma_x^2} \leq F_{n-1, m-1; (1+p)/2} \frac{S_y^2}{S_x^2} \right) = p. \quad (9.25)$$

Beim Beispiel 9.7 lauten die Grenzen der Realisierungen dieses Vertrauensintervalls für  $p = 0.95$

$$\frac{s_y^2}{s_x^2} F_{9,7;0.025} = \frac{7.14}{10.72} \cdot 0.238 = 0.16 \quad \text{und} \quad \frac{s_y^2}{s_x^2} F_{9,7;0.975} = \frac{7.14}{10.72} \cdot 4.82 = 3.21.$$

Diese Zahlen belegen, wie breit Konfidenzintervalle für Quotienten von Varianzen sein können, wenn nur wenige Daten vorliegen.

In den gängigen Statistik-Programmpaketen werden weitere Tests zum Vergleich von Mittelwerten angeboten. So verwendet etwa SAS den Behrens-Fisher-Ansatz; vgl. etwa [Falk et al. (1995)], S. 63 oder die nicht mehr  $F$ -verteilte Quotientenstatistik

$$F' := \frac{\max \{ S_X^2, S_Y^2 \}}{\min \{ S_X^2, S_Y^2 \}}; \quad (9.26)$$

siehe etwa [Falk et al. (1995)], S. 64. In obigem Beispiel erhielt man als Realisierung von  $F'$  den Wert  $f'_0 = \frac{10.72}{7.14} = 1.501$ .

Für weitere Aspekte, wie die varianzstabilisierende Wurzel- oder Logarithmustransformation, muss ebenso auf die weiterführende Literatur verwiesen werden (siehe etwa

[Falk et al. (1995)], S. 67) wie für den in der Praxis äußerst nützlichen Wilcoxon-Test zum Mittelwertvergleich. Dort wird keine Normalverteilung der Daten verlangt; vgl. [Falk et al. (1995)], S. 70-77.

Eine Frage soll am Ende dieses Abschnitts nochmals aufgegriffen werden, die mit der Festlegung des Signifikanzniveaus bei den diskutierten Tests zusammenhängt:

## 9.5 Der $p$ – Wert

Manchem Leser, der sich zum ersten Mal mit statistischen Tests befaßt, wird die Festlegung des Signifikanzniveaus  $\alpha$  willkürlich erscheinen. Er kann höchstens einsehen, dass  $\alpha < 10\%$  sinnvoll ist. Tatsächlich belegen die Daten von Beispiel 9.5, dass die Wahl des Signifikanzniveaus nicht ganz unproblematisch ist. Deshalb wird in den wichtigsten Statistik-Programmpaketen zu einer Teststatistik der sog.  $p$  – Wert angegeben.

### Def. 9.8 (p-Wert (p-value), Überschreitungswahrscheinlichkeit)

Gegeben sei ein Test mit kritischem Bereich  $K_\alpha$ , der nur über eine Teststatistik  $T = T(X_1, \dots, X_n)$  von der Stichprobe  $X_1, \dots, X_n$  abhängt.  $t_0$  sei die Realisierung von  $T$  unter der Hypothese  $H_0$ . Dann nennt man

$$p_{t_0} := \inf\{ \alpha \mid t_0 \in K_\alpha \} \quad (9.27)$$

den  **$p$  – Wert** ( **$p$  – value**) von  $t_0$ . Falls

$$K_{\alpha'} \subset K_\alpha \quad \text{für} \quad \alpha' \leq \alpha \quad (9.28)$$

ist der  $p$ -Wert jenes Signifikanzniveaus, bei dem die Nullhypothese gerade noch abgelehnt wird.

Die Hypothese  $H_0$  ist also beim Niveau  $\alpha$  abzulehnen, falls

$$p_{t_0} < \alpha. \quad (9.29)$$

$T$  ist dabei eine beliebige Teststatistik.  $H_0$  werde abgelehnt, falls  $t_0 \in K_\alpha$ . Für wachsendes  $\alpha$  bläht sich  $K_\alpha$  immer mehr auf, während es bei kleiner werdendem  $\alpha$  immer mehr schrumpft. Anstatt sich den Kopf zu zerbrechen, ob man  $\alpha = 0.05$  oder  $\alpha = 0.01$  von vorneherein festlegen soll, kann man auch erst die Realisierung  $t_0$  aus den Daten bestimmen und sehen, ob die Hypothese nicht beim Niveau 0.05 beibehalten oder beim Niveau 0.01 abgelehnt wird. Bei kritischen Bereichen, die die Bedingung (9.28) erfüllen, ändert sich dann in diesen Fällen die Entscheidung für  $0.01 \leq \alpha \leq 0.05$  nicht.

Nun soll der  $p$  – Wert an zwei Beispielen erläutert werden. Bei der Diätkur von Beispiel 9.5 galt  $t_0 = 1.908$ . Man erhält für die zweiseitige bzw. einseitige Alternative:

Alternative	kritischer Bereich	$p$ – Wert : $p_{t_0} := \inf_{\alpha} \{ t_0 \in K_{\alpha} \}$
$H_1 : \mu \neq 0$	$ t  > \gamma_{n-1; 1-\alpha}$	$p_{t_0} = P( T  >  t_0 ) = 9.28\%$ , da $\gamma_{8; 0.9072} = 1.908$
$H_1 : \mu > 0$	$t > t_{n-1; 1-\alpha}$	$p_{t_0} = P(T > t_0) = 4.64\%$ , da $t_{8; 0.9536} = 1.908$

Legt man das übliche Signifikanzniveau von  $\alpha = 0.05$  zugrunde, so sieht man an dem  $p$  – Wert von  $p_{1.908} = 4.64\%$  bei der einseitigen Alternative, dass bereits bei einem Niveau von  $\alpha = 4.5\%$  die Hypothese nicht abgelehnt worden wäre.

Dagegen liefert die Realisierung der T-Statistik beim Mittelwertvergleich des Beispiels 9.6 mit den Betonsorten bei 16 Freiheitsgraden einen  $p$  – Wert von

$$p_{5.922} = P(|T| > 5.922) = 0.002\%, \quad (9.30)$$

was z.B. bedeutet, dass die dort vorliegende Hypothese bei den beobachteten Daten sogar bei dem extrem kleinen Signifikanzniveau 0.01% abzulehnen wäre. Weitere Details zum  $p$  – Wert entnehme man etwa [Falk et al. (1995)], S. 66 und [Fahrmeir et al. (1997)], S. 408.

In Programmpaketen wie SAS wird übrigens der jeweilige  $p$  – Wert durch Bezeichnungen wie ”  $P > |T|$  ” ausgewiesen.

## 9.6 Einfache lineare Regression

$E_i$ ,  $i = 1, \dots, n$ , seien iid  $N(0, \sigma^2)$  Zufallsvariable (ZV) und  $x_1, \dots, x_n$  bekannte Zahlen (z.B. Meßstellen). Zu unbekanntem Parameter  $\alpha$  und  $\beta$  (auch Regressionskoeffizienten) beobachtet man Realisierungen der ZV

$$Y_i = \alpha + \beta x_i + E_i, \quad i = 1, \dots, n.$$

Aus den Voraussetzungen folgt für  $i = 1, \dots, n$ :  $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$  und unabhängig. Wie in der empirischen Regression definiert man

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \tilde{s}_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Man beachte jedoch, dass jetzt

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad \tilde{S}_Y^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad \text{und} \quad \tilde{S}_{xy} = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})(x_i - \bar{x})$$

Zufallsvariable sind.

### Schätzungen und deren Verteilung

Mit  $\tilde{s}_x^2 > 0$  (d.h. nicht alle  $x_i$  identisch) erhält man folgende Schätzungen für die unbekanntem Parameter

$$\hat{\beta} = B = \frac{\tilde{S}_{xy}}{\tilde{s}_x^2} = \frac{\sum (Y_i - \bar{Y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2},$$

$$\hat{\alpha} = A = \bar{Y} - B \bar{x},$$

$$S^2 := \frac{n}{n-2} (\tilde{S}_Y^2 - B^2 \tilde{s}_x^2) = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2,$$

wobei

$$\hat{Y}_i = A + Bx_i, \quad i = 1, \dots, n.$$

$y_i$ ,  $a$  und  $b$  seien die Realisierungen von  $Y_i$ ,  $i = 1, \dots, n$ ,  $A$  und  $B$ . Damit ergeben sich gerade die Ergebnisse der empirischen Regression. Weiterhin sei

$$s^2 := \frac{1}{n-2} \sum_{i=1}^n (y_i - a - bx_i)^2, \quad s = \sqrt{s^2}.$$

Analog zum Beweis der Unabhängigkeit des Stichprobenmittels und der Stichprobenvarianz zeigt man, dass die Schätzfunktionen  $\bar{Y}$ ,  $B$  und  $S^2$  unabhängig sind. Damit ist auch  $A = \bar{Y} - B \bar{x}$  von  $S^2$  unabhängig. Ferner sind  $A$  bzw.  $B$  normalverteilt mit Erwartungswerten  $\alpha$  bzw.  $\beta$ .  $(n-2) S^2 / \sigma^2$  ist  $\chi^2$ -verteilt mit  $n-2$  Freiheitsgraden.

Für die Schätzer  $A$  und  $B$  lauten die Varianzen

$$D^2(B) = \text{Var}(B) = \frac{\sigma^2}{n \tilde{s}_x^2} = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

$$D^2(A) = \text{Var}(A) = D^2(B)(\bar{x}^2 + \tilde{s}_x^2) = \sigma^2 \frac{\sum x_i^2/n}{\sum (x_i - \bar{x})^2}.$$

Mit  $S^2$  anstelle des unbekanntes  $\sigma^2$  erhält man Varianzschätzungen für  $A$  und  $B$

$$\hat{\sigma}_B^2 = \hat{D}^2(B) = \frac{S^2}{\sum (x_i - \bar{x})^2} \quad \text{und} \quad \hat{\sigma}_A^2 = \hat{D}^2(A) = S^2 \frac{\sum x_i^2/n}{\sum (x_i - \bar{x})^2}.$$

Damit sind

$$T_A = \frac{A - \alpha}{\hat{\sigma}_A} \quad \text{und} \quad T = T_B = \frac{B - \beta}{\hat{\sigma}_B}$$

$t$ -verteilt mit  $n - 2$  Freiheitsgraden (zur Berechnung von  $S^2$  benötigt man die Schätzer für die unbekanntes Parameter  $\alpha$  und  $\beta$ ).

### Signifikanztest und Vertrauensintervalle

Der folgende **Signifikanz-Test** wird häufig für den Spezialfall  $\beta_0 = 0$  angewendet

$$H_0 : \beta = \beta_0 \quad H_1 : \beta \neq \beta_0$$

$1 - p$  bezeichne das Testniveau.  $H_0$  ist mit einer Sicherheitswahrscheinlichkeit von  $p = P\%$  abzulehnen, falls für die Realisierung  $t_0$  der Testgröße

$$T_0 = \frac{B - \beta_0}{\hat{\sigma}_B}$$

unter der Nullhypothese gilt

$$|t_0| = \frac{|b - \beta_0|}{s/\sqrt{\sum (x_i - \bar{x})^2}} > \gamma_{n-2;P\%}.$$

Kann  $H_0 : \beta_0 = 0$ , etwa beim Signifikanzniveau  $1 - p = 0.05$  nicht abgelehnt werden, so ist der einfachere Ansatz

$$E(Y_i) = \alpha$$

(ohne  $x_i$ ) eher gerechtfertigt als

$$E(Y_i) = \alpha + \beta x_i.$$

Obige Verteilungen von  $A$ ,  $B$  und  $S^2$  liefern folgende **P%-Vertrauens-** oder **Konfidenzintervalle** für die **Regressionskoeffizienten**  $\alpha$  und  $\beta$  :

$$P(A - d_A \leq \alpha \leq A + d_A) = P\%, \quad d_A = \gamma_{n-2;P\%} S \sqrt{\frac{\sum x_i^2/n}{\sum (x_i - \bar{x})^2}}$$

$$P(B - d_B \leq \beta \leq B + d_B) = P\%, \quad d_B = \gamma_{n-2;P\%} S / \sqrt{\sum (x_i - \bar{x})^2}.$$

Mit ähnlichen Überlegungen wie bei den Regressionskoeffizienten (siehe etwa [Heinhold & Gaede (1979)], § 21 oder [Fahrmeir et al. (1996)], Kapitel 4.) erhält man zwei weitere wichtige Konfidenzintervalle:

**Prognoseintervall:**

Für festgelegtes  $x_0$  sei  $Y(x_0)$  eine neue ZV, unabhängig von  $Y_1, \dots, Y_n$ , mit  $E[Y(x_0)] = \alpha + \beta x_0$ .  $\hat{Y}(x_0) = A + Bx_0$  bezeichne die Prognose für  $Y(x_0)$ . Mit

$$d(x_0) = \gamma_{n-2; P\%} S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}.$$

erhält man

$$P \left( \hat{Y}(x_0) - d(x_0) \leq Y(x_0) \leq \hat{Y}(x_0) + d(x_0) \right) = P\%$$

ein  $P\%$ -Konfidenzintervall für eine neue Beobachtung  $Y(x_0)$  an der Stelle  $x_0$ .  $\hat{Y}(x_0) = A + Bx_0$  ist ein erwartungstreuer Schätzer für  $Y(x_0)$ . Die Prognose wird umso ungenauer, je weiter  $x_0$  von  $\bar{x}$  entfernt ist; siehe Bild 28. Das relativ weite Konfidenzintervall für die Prognose darf nicht verwechselt werden mit dem engeren für den Erwartungswert.

**Vertrauensintervall für den Erwartungswert**  $\eta(x) = \alpha + \beta x$ ,  $x$  bekannt:

Sei

$$c(x) = \gamma_{n-2; P\%} S \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2}},$$

dann gilt

$$P \left( \hat{Y}(x) - c(x) \leq \eta(x) \leq \hat{Y}(x) + c(x) \right) = P\%.$$

Die Konfidenzintervalle für Erwartungswert und Prognose unterscheiden sich formal nur im Term  $\sqrt{1 + \dots}$  und sind - grob gesprochen - folgendermaßen zu interpretieren:

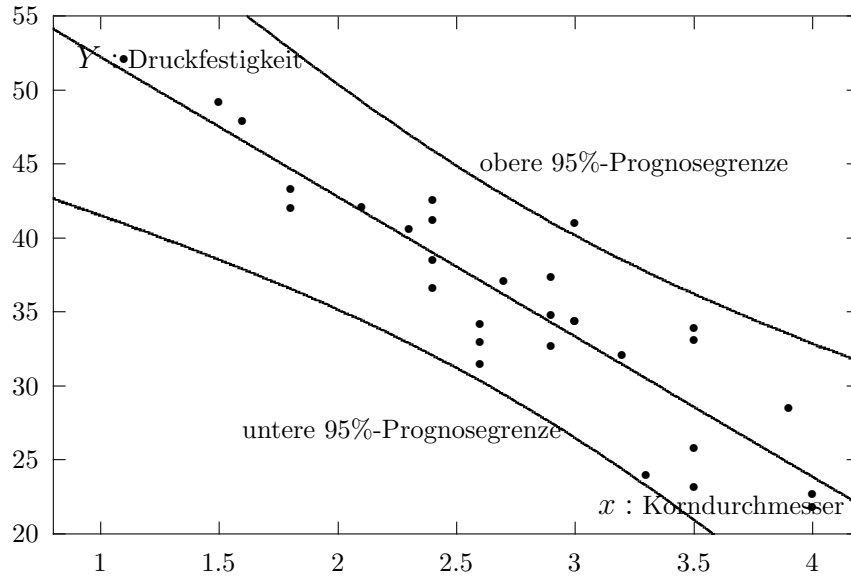
**1. Erwartungswert:** Würde man das dem linearen Modell zugrundeliegende Zufallsexperiment 1000 mal mit jeweils einer genügenden Zahl von Datenpunkten durchführen, so lägen bei einem Konfidenzniveau  $p = 0.95$  ca. 950 der berechneten Regressionsgeraden im engeren der beiden Konfidenzintervalle.

**2. Prognose:** Wiederum für  $p = 0.95$  liegen ca. 95% der Ergebnisse  $Y(x_0^{(j)})$  des gleichen Zufallsexperiments an 1000 neuen Meßpunkten  $x_0^{(j)}$ ,  $j = 1, \dots, 1000$ , im breiteren Prognoseintervall.

**Beispiel 9.9 (Korngröße und Druckfestigkeit)**

vgl. [Stoyan (1993)]; Abschnitt 1.3.

Tabelle 15: $Y$ Druckfestigkeit [ $N/mm^2$ ], $x$ Korndurchmesser [ $mm$ ], $n = 30$															
$x_i$	3.5	2.4	1.8	3.2	2.4	3.5	3.0	3.5	4.0	1.8	2.9	3.5	2.4	2.9	3.3
$y_i$	23.2	38.5	42.0	32.1	41.2	25.8	41.0	33.9	22.7	43.3	34.8	33.1	42.6	32.7	24.0
$x_i$	2.6	2.6	1.6	3.0	1.5	3.0	2.4	3.9	2.3	2.1	2.7	2.6	4.0	2.9	1.1
$y_i$	31.5	34.2	47.9	34.4	49.2	34.4	36.6	28.5	40.6	42.1	37.1	33.0	21.8	37.4	52.1



**Bild 28:** Regressionsgerade  $y = a + bx$  mit 95%-Prognoseintervall

Aus den Daten erhält man zunächst die Hilfsgrößen

$$\begin{aligned}\bar{x} &= 2.747, & \sum (x_i - \bar{x})^2 &= 16.054, \\ \bar{y} &= 35.723, & \sum (y_i - \bar{y})^2 &= 1750.43, \\ n &= 30 & \text{und} & \sum (x_i - \bar{x})(y_i - \bar{y}) &= -151.69.\end{aligned}$$

Daraus ergeben sich folgende Schätzungen für die Regressionskoeffizienten

$$b = \frac{-151.69}{16.054} = -9.45, \quad a = 35.723 - (-9.45)2.747 = 61.68.$$

Als Restvarianz erhält man dann

$$s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - a - bx_i)^2 = 11.327, \quad s = 3.366.$$

Bestimmtheitsmaß  $R^2$  und Korrelationskoeffizient  $r$  lauten hier

$$R^2 = \frac{(-151.69)^2}{16.054 \cdot 1750.43} = 0.819, \quad r = -0.905.$$

Die Hypothese  $H_0 : \beta = 0$  ist beim Niveau 0.05 abzulehnen, da

$$|t_0| = \frac{|-9.45|}{3.366/\sqrt{16.054}} = 11.25 > 2.048 = \gamma_{28;0.95} = t_{28;0.975}.$$

Der  $p$ -Wert  $p_{t_0}$  (vgl. Abschnitt 9.5) zu  $t_0 = -11.25$ , also jenes Signifikanzniveau, bei dem  $H_0$  gerade noch abgelehnt würde, lautet  $p_{t_0} = p_{-11.25} = 0.000\%$ ; ist also praktisch gleich Null.

**95%-Konfidenzintervalle:**

Mit dem symmetrischen Quantil  $\gamma_{28;0.95} = 2.048$  erhält man

$$d_B = 2.048 \frac{3.366}{\sqrt{16.054}} = 1.72, \quad d_A = 1.72 \sqrt{16.054/30 + 2.747^2} = 4.89,$$

wobei die Identität  $\sum x_i^2/n = \sum (x_i - \bar{x})^2/n + \bar{x}^2$  benutzt wurde. Somit lauten die Realisierungen der 95%-Vertauensintervalle

$$56.78 \leq a \leq 66.57, \quad -11.17 \leq b \leq -7.73.$$

$x$	Prognose $\hat{y}(x)$	$\frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2}$	$c(x) = \gamma s \sqrt{\frac{1}{n} + \sum \dots}$	$d(x) = \gamma s \sqrt{1 + \frac{1}{n} + \sum \dots}$
1.1	51.28	0.169	3.10	7.56
$\bar{x}$	$\bar{y} = 35.72$	0	1.26	7.01
4	23.88	0.098	2.50	7.33

$x$	$\hat{y}(x)$	Erwartungswert $\eta(x)$	Prognose $y(x)$
1.1	51.28	[48.18,54.38]	[43.72,58.84]
$\bar{x} = 2.747$	$\bar{y} = 35.72$	[34.46,36.98]	[28.71,42.73]
4	23.88	[21.38,26.38]	[16.55,31.21]

Das Schaubild zeigt die sog. **Regressionstrompete** für das 95%-Prognoseintervall. Die Prognose ist also bei  $\bar{x}$  am genauesten und wird umso unschärfer, je weiter man sich von  $\bar{x}$  entfernt. Übrigens liegt  $(\bar{x}, \bar{y})$  stets auf der Regressionsgeraden.

## 9.7 Multiple Regression

Hier setzt man

$$E(Y_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \quad i = 1, \dots, n,$$

an, wobei

$$\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T, \quad p \leq n,$$



nicht zufallsabhängige, bekannte Vektoren sind, die sog. **Regressor-Variablen**.

$$\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$$

ist ein unbekannter Parametervektor.

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix}$$

heißt **Daten-** oder **Design-Matrix**.

Die erste Spalte von  $\mathbf{X}$  wird meist als  $(1, \dots, 1)^T$  gewählt und repräsentiert im Ansatz ein konstantes Glied (z.B.  $\alpha$  in der einfachen linearen Regression). Mit

$$\mathbf{x}_i = (1, x_i)^T \quad \text{und} \quad \boldsymbol{\beta} = (a, \beta)^T$$

ist die einfache lineare Regression ein Spezialfall der multiplen Regression.

Seien weiterhin

$$\mathbf{Y} = (Y_1, \dots, Y_n)^T \quad \text{und} \quad \mathbf{E} = (E_1, \dots, E_n)^T,$$

dann erhält man in Matrixschreibweise

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{E}, \quad E(\mathbf{E}) = 0, \quad \text{Cov}(\mathbf{E}) = \sigma^2 \mathbf{I}.$$

Die Schätzungen  $\hat{\boldsymbol{\beta}}$  bzw.  $S^2$  für die unbekannt Parameter  $\boldsymbol{\beta}$  und  $\sigma^2$  erhält man aus den sog. **Normalgleichungen**

$$\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{Y}$$

und

$$S^2 = \frac{1}{n - \text{Rang}(\mathbf{X})} \|\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}\|_2^2.$$

Bei Maximalrang, also  $\text{Rang}(\mathbf{X}) = p$ , gilt

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

und

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} =: \sigma^2 \mathbf{C}.$$

Falls  $\mathbf{E} \sim N_n(0, \sigma^2 \mathbf{I})$ -verteilt ist, gilt:

1.  $\hat{\boldsymbol{\beta}}$  und  $S^2$  sind unabhängig
2.  $\frac{(n-p) S^2}{\sigma^2}$  ist  $\chi_{n-p}^2$ -verteilt
3.  $\hat{\boldsymbol{\beta}}$  ist  $N_p(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1})$ -verteilt.

Daraus lassen sich analog zur einfachen linearen Regression Konfidenzintervalle für die  $\beta_i$  und für Prognosen ableiten. Wir gehen hier nur noch ein auf den **Signifikanz-Test**

$$H_0 : \beta_i = 0, \quad H_1 : \beta_i \neq 0.$$

$H_0$  ist abzulehnen, falls

$$\frac{|\hat{\beta}_i|}{\sqrt{c_{ii}} S} > \gamma_{n-p;P\%},$$

wobei  $c_{ii}$  das  $i$ -te Diagonalelement von  $\mathbf{C} = (\mathbf{X}^T \mathbf{X})^{-1}$  und  $\gamma_{n-p;P\%}$  das symmetrische  $P\%$ -Quantil der  $t$ -Verteilung mit  $n - p$  Freiheitsgraden bezeichnet.

### Bemerkungen:

1. Falls mehrere Regressor-Variablen (man sagt auch: **Einflussgrößen** oder **Regressoren**) vorliegen, läßt sich der oben genannte Signifikanz-Test zur sog. **Variablenauswahl** heranziehen. Wird nämlich  $H_0$  für ein  $\beta_i$  abgelehnt, so übt diese Variable (statistisch gesehen) keinen Einfluss auf den Erwartungswert von  $Y$  aus und kann im Ansatz weggelassen werden. Das fortgesetzte Aufnehmen und Weglassen von Regressorvariablen nennt man **Schrittweise Regression**. Für Details sei auf die weiterführende Literatur verwiesen, z.B. [Fahrmeir et al. (1996)], Kapitel 4 oder [Seber (1977)].
2. Die Berechnungen zur multiplen Regression werden normalerweise mit Standardstatistikpaketen wie SPSS, SAS, S-Plus etc. durchgeführt.
3. Wie schon erwähnt, kann die einfache lineare Regression als Spezialfall der multiplen Regression mit  $p = 2$  angesehen werden.

Für eine eingehende Diskussion weiterer statistischer Fragestellungen sei auf Lehrbücher wie [Sachs (1984)], [Witting (1985)] und [Fahrmeir et al. (1996)] hingewiesen.

# Kapitel 10

## Explorative Datenanalyse

Die Darstellung folgt in wichtigen Teilen [Falk et al. (1995)]. Eine leicht verständliche Einführung in die Datenanalyse findet man auch in [Fahrmeir et al. (1997)].

Eine wesentliche Grundlage der explorativen Datenanalyse ist die Tatsache, dass sich eine unbekannte Verteilungsfunktion durch Realisierung entsprechend vieler zugehöriger iid-Zufallsvariablen beliebig genau approximieren läßt.

### Def. 10.1 (Empirische Verteilungsfunktion)

$X_1, \dots, X_n \in \mathbb{R}$  sei eine Stichprobe. Dann heißt

$$F_n(t) = \frac{|\{X_i, i = 1, \dots, n : X_i \leq t\}|}{n} = \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, t]}(X_i), \quad t \in \mathbb{R},$$

**empirische Verteilungsfunktion.** Je nach Sachlage werden wir obigen Zugang mit den Zufallsvariablen (ZV)  $X_i$  oder die der Definition 7.12 entsprechende Schreibweise von  $F_n$  bzgl. der Realisierungen  $x_1, \dots, x_n$  wählen.  $1_A(\cdot)$  bezeichnet die Indikatorfunktion einer Menge  $A$ , vgl. Def. 1.44.

### Lemma 10.2 (von Glivenko-Cantelli)

Für iid-Variablen  $X_1, \dots, X_n$  mit Verteilungsfunktion  $F$  gilt

$$P \left( \lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| = 0 \right) = 1, \quad d. h.$$

$$\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| = 0 \quad \text{fast sicher (f.s.),}$$

wobei  $F_n(t) = \sum_{i=1}^n 1_{(-\infty, t]}(X_i)/n$ .

Das Glivenko-Cantelli-Lemma wird oft **Hauptsatz der Statistik** genannt.

Beweis: siehe etwa [Chung (1974)], Theorem 5.5.1.

## 10.1 Kerndichte-Schätzer

Sind die  $X_i$  stetig und ist  $F$  differenzierbar, so lautet die Dichte  $f = F'$  (wir nehmen wieder an, dass  $f$  stückweise stetig ist). Wir erhalten somit f.s.

$$\frac{F_n(t+h) - F_n(t)}{h} \xrightarrow{n \rightarrow \infty} \frac{F(t+h) - F(t)}{h} \xrightarrow{h \rightarrow 0} F'(t) = f(t)$$

für  $t \in \mathbb{R}$  und  $h > 0$ . Wenn  $n$  also groß genug ist und  $h$  klein, so wird gelten:

$$\frac{F_n(t+h) - F_n(t)}{h} \sim f(t),$$

wobei das Symbol " $\sim$ " hier soviel wie *asymptotisch gleich* bedeutet. Mit den Bezeichnungen von Abschnitt 7.3 und  $t = a_{j-1}$ ,  $h = a_j - a_{j-1}$ , folgt daher für  $x \in I_j$

$$\frac{F_n(t+h) - F_n(t)}{h} = \frac{F_n(a_j) - F_n(a_{j-1})}{a_j - a_{j-1}} = \frac{n_j}{n} \frac{1}{a_j - a_{j-1}} = f_n(x) \sim f(t).$$

Die empirische Dichte  $f_n$  ist also ein Schätzer für die zugrundeliegende Dichte  $f(\cdot)$ .

### Def. 10.3 (Kern)

Es sei  $K : \mathbb{R} \rightarrow [0, 1]$  eine Verteilungsfunktion, der sogenannte **Kern**. Setze für  $h > 0$  und  $t \in \mathbb{R}$

$$\hat{F}_n(t) = \int_{-\infty}^{\infty} K\left(\frac{t-x}{h}\right) dF_n(x) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{t-X_i}{h}\right).$$

Der Schätzer  $\hat{F}_n(t)$  ist die **Faltung** von  $K(\cdot/h)$  und der empirischen Verteilungsfunktion  $F_n$  und heißt **Kern-Schätzer** von  $F$ , der Verteilungsfunktion von iid-ZV  $X, X_1, \dots, X_n$ . Der Parameter  $h$  heißt **Bandbreite** (bandwidth) oder **Fensterbreite**.

Je nachdem, ob die Schätzfunktion (Schätzer)  $\hat{F}_n$  oder die reelle Funktion  $F_n(t)$  betrachtet wird, ist jeweils die ZV  $X_i$  oder deren Realisierung  $x_i$ ,  $i = 1, \dots, n$ , einzusetzen. Da  $K$  eine Verteilungsfunktion ist, gilt  $\lim_{x \rightarrow \infty} K(x) = 1$  und  $\lim_{x \rightarrow -\infty} K(x) = 0$ .

Damit ist  $\hat{F}_n$  auch eine Verteilungsfunktion, die bei einer kleinen Fensterbreite  $h$  in der Nähe der empirischen Verteilungsfunktion liegt, denn

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{t-X_i}{h}\right) \xrightarrow{h \rightarrow 0} F_n(t)$$

für alle  $t \notin \{X_1, \dots, X_n\}$ , d. h.  $\hat{F}_n(t)$  ist wiederum ein Schätzer für  $F(t)$ . Für differenzierbares  $K$  mit  $K' = k$  folgt:

$$\hat{F}'_n(t) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{t - X_i}{h}\right) = \frac{1}{h} \int_{-\infty}^{\infty} k\left(\frac{t - x}{h}\right) dF_n(x) =: \hat{f}_n(t),$$

wobei die Differentiation so durchzuführen ist als würde  $x_i$  anstelle der  $X_i$  verwendet.

#### Def. 10.4 (Kern-Dichteschätzer)

Sei  $F$  die Verteilungsfunktion von stetigen iid ZV  $X, X_1, \dots, X_n$  mit Dichte  $f$  und  $k: \mathbb{R} \rightarrow \mathbb{R}$  eine nichtnegative Funktion mit  $\int_{-\infty}^{\infty} k(x) dx = 1$ , der **Kern**. Die Abbildung

$$\hat{f}_n(t) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{t - X_i}{h}\right) = \frac{1}{h} \int_{-\infty}^{\infty} k\left(\frac{t - x}{h}\right) dF_n(x), \quad t \in \mathbb{R}, \quad h > 0,$$

heißt **univariater Kern-Dichteschätzer** für  $f(t)$  mit Fensterbreite oder Bandbreite  $h$ .

Die Verwandtschaft mit der Histogramm-Dichte von Abschnitt 7.3 zeigt sich bei der speziellen Wahl  $k(x) = \frac{1}{2} 1_{[-1,1]}(x)$

$$\begin{aligned} \hat{f}_n(t) &= \frac{1}{nh} \sum_{i=1}^n k\left(\frac{t - X_i}{h}\right) = \frac{F_n(t+h) - F_n(t-h)}{2h} \\ &= \frac{\text{Anzahl der Beobachtungen im Intervall } (t-h, t+h]}{n \cdot \text{Länge von } (t-h, t+h]}. \end{aligned} \quad (10.1)$$

Wichtige Beispiele für Kerndichteschätzer sind:

#### 1. Epanechnikov-Kern

Dieser populäre Kern ist definiert als

$$k_E(x) = \begin{cases} \frac{3}{4\sqrt{5}} \left(1 - \frac{x^2}{5}\right), & \text{falls } |x| \leq \sqrt{5} \\ 0, & \text{sonst.} \end{cases} \quad (10.2)$$

Unter allen Kernen  $k$ , die außerhalb des Intervalls  $[-\sqrt{5}, \sqrt{5}]$  den Wert 0 annehmen und für die gilt

$$\int k(x) dx = 1, \quad \int x^2 k(x) dx = 1,$$

minimiert er  $\int k^2(x) dx$ .

## 2. Rechtecks-Kern

$$k_u(x) := \begin{cases} \frac{1}{2} & , \text{ falls } |x| \leq 1 \\ 0 & , \text{ falls } |x| > 1. \end{cases}$$

## 3. Dreiecks-Kern

$$k_\Delta(x) := \begin{cases} 1 - |x| & , \text{ falls } |x| \leq 1 \\ 0 & , \text{ falls } |x| > 1. \end{cases}$$

## 4. Normalverteilung (Gauß)-Kern

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

Für eine ausführlichere Darstellung siehe z.B. [Falk et al. (1995)].

Kerndichteschätzer sind sehr sensitiv gegenüber der Wahl der Bandbreite  $h$ . Dies steht in direkter Analogie zum Einfluss der Intervallbreiten bei Histogrammen. Das folgende Bild zeigt Kerndichteschätzer desselben Typs angewendet auf Jahresminima der Wintertemperaturen in der Oberpfalz bzgl. der Jahre 1946 bis 1997. Man beachte die unterschiedliche Skalierung der Ordinatenachse.

**Bild 29:** Kerndichteschätzer mit Bandbreite  $h = 0.5$  (links) und  $h = 0.125$ .

Für eine Diskussion optimaler Bandbreiten muss auf die Spezialliteratur verwiesen werden. Statistikpakete wie SAS bieten eine interaktive Wahl der Bandbreite an, was für den Benutzer in vielen Fällen ausreicht.

Das sog. nichtparametrische Modell der Kerndichteschätzer läßt sich in der Regel sehr gut an gegebene Daten anpassen. Dieses Verfahren eignet sich aber beispielsweise nicht,

um Extremwerte zu schätzen. Denn alle genannten Kerndichteschätzer haben kompakten Träger. Über diesen beschränkten "Tellerrand" können sie nicht hinaussehen, was aber bei der Behandlung von Extremwerten unerlässlich ist. Interessant ist etwa die Berechnung der Wahrscheinlichkeit dafür, dass die Jahresextremtemperatur unter  $-40^\circ$  Celsius sinkt. Gemäß obigen Grafiken verschwindet die Kerndichteschätzung "links" von  $-37^\circ$ , womit die gesuchte Wahrscheinlichkeit 0 ist.

Die übliche Alternative zu Kerndichteschätzern ist die historisch ältere - und nach wie vor häufig angewendete - Methode der parametrischen Dichteschätzverfahren. Dort ist ein gewisser Verteilungstyp a priori festzulegen (z.B. Normal-, Exponentialverteilung etc.). Dann sind nur noch die unbekannt Parameter (z.B.  $\mu, \sigma^2$  oder  $\lambda$ ) dieser Verteilung zu schätzen; etwa mit der ML-Methode.

## 10.2 QQ-Plots

Wie schon in dem vorangegangenen Abschnitt unterscheiden wir hier nicht zwischen den Zufallsvariablen  $X_1, \dots, X_n$  und deren Realisierungen  $x_1, \dots, x_n$ . Der besseren Übersicht halber seien einige Resultate aus dem Abschnitt 1.5.4 hier nochmals zusammengestellt. Notation und Darstellung folgen weitgehend [Falk et al. (1995)].

**Def. 10.5 (Verallgemeinerte Inverse, Quantilfunktion)**

$$F^{\leftarrow}(q) := \inf \{ x \in \mathbb{R} : F(x) \geq q \}, \quad q \in (0, 1).$$

heißt **verallgemeinerte Inverse** oder **Quantilfunktion** zu  $F$ .

**Lemma 10.6 (Eigenschaften der Quantilfunktion)**

1. Für beliebige Verteilungsfunktionen gilt

$$F^{\leftarrow}(q) \leq t \iff q \leq F(t) \quad \text{für alle } t \in \mathbb{R} \text{ und } q \in (0, 1).$$

2. Für streng monotonen  $F : \mathbb{R} \rightarrow (0, 1)$  erhält man  $F^{\leftarrow} = F^{-1}$ .

**Korollar 10.7 (Quantiltransformation)**

$X$  sei eine ZV mit Verteilungsfunktion  $F$  und  $U$  sei  $(0, 1)$ -gleichverteilt.

1. Für beliebiges  $X$  hat die Zufallsvariable  $Y := F^{\leftarrow}(U)$  die Verteilungsfunktion  $F$ , d. h.

$$P(Y \leq y) = P(F^{\leftarrow}(U) \leq y) = F(y), \quad y \in \mathbb{R}.$$

2. Für stetiges  $F$  ist  $Z := F(X)$  gleichverteilt in  $(0, 1)$ , d. h.

$$P(Z \leq z) = P(F(X) \leq z) = z, \quad z \in (0, 1).$$

Wegen dieser Transformation kann man bei der Untersuchung einer Zufallsvariablen  $X$  mit Verteilungsfunktion  $F$  von der Darstellung

$$X = F^{\leftarrow}(U)$$

ausgehen, wobei  $U \sim U(0, 1)$ .



**Lemma 10.8 (Normierung)**

Es sei  $Y$  eine Zufallsvariable mit der Verteilungsfunktion  $G$ ;  $\sigma > 0$ ,  $\mu \in \mathbb{R}$ . Weiter sei  $F$  eine Verteilungsfunktion zu  $X := \mu + \sigma Y$ , d. h.  $F(t) = G((t - \mu)/\sigma)$ ,  $t \in \mathbb{R}$ . Dann gilt für  $q \in (0, 1)$ :

$$F^{\leftarrow}(q) = \mu + \sigma G^{\leftarrow}(q).$$

**Lemma 10.9**

Für auf  $(0, 1)$  gleichverteilte und unabhängige Zufallsvariablen  $U_1, \dots, U_n$  gilt mit Wahrscheinlichkeit 1

$$\max_{1 \leq k \leq n} |U_{k:n} - k/n| \leq \sup_{t \in [0,1]} |F_n(t) - t|,$$

wobei  $F_n$  die **empirische Verteilungsfunktion** zu  $U_1, \dots, U_n$  ist und  $U_{1:n} < U_{2:n} < \dots < U_{n:n}$  die in Abschnitt 2.54 eingeführten Ordnungsstatistiken bezeichnen. Wegen der Stetigkeit der Gleichverteilung folgt  $P(U_{k:n} = U_{k+1:n}) = 0$ .

**Korollar 10.10**

In Analogie zum starken Gesetz der großen Zahlen (vgl. Satz 4.8 und wegen Formel (2.43) gilt für die Ordnungsstatistiken  $U_{k:n}$ ,  $k = 1, \dots, n$ , mit Wahrscheinlichkeit 1

$$\max_{1 \leq k \leq n} \left| U_{k:n} - \frac{k}{n+1} \right| \xrightarrow{n \rightarrow \infty} 0.$$

Die Beweise entnehme man [Falk et al. (1995)], S. 31 ff.

**Korollar 10.11**

Die Zufallsvariablen  $X_1, \dots, X_n$  seien iid mit der Verteilungsfunktion  $F$  und es sei  $F^{\leftarrow}$  stetig auf  $(a, b) \subset (0, 1)$ . Dann gilt für die Ordnungsstatistiken  $X_{k:n}$  mit Wahrscheinlichkeit 1

$$\max_{k_1 \leq k \leq k_2} \left| X_{k:n} - F^{\leftarrow}\left(\frac{k}{n+1}\right) \right| \xrightarrow{n \rightarrow \infty} 0,$$

falls  $k_1 = k_1(n) \leq k_2 = k_2(n)$  Zahlenfolgen sind, welche die Bedingung

$$a < \liminf_{n \in \mathbb{N}} \frac{k_1}{n} \leq \limsup_{n \in \mathbb{N}} \frac{k_2}{n} < b$$

erfüllen.

**Beweis:**

Die Funktion  $F^{\leftarrow}$  ist gleichmäßig stetig auf  $[a + \epsilon, b - \epsilon]$ , falls  $\epsilon > 0$  hinreichend klein ist. Setzen wir  $X_{k:n} = F^{\leftarrow}(U_{k:n})$ , so folgt die Behauptung aus dem vorherigen Korollar.

**Quantil-Plots (QQ-Plots)**

Wir nehmen im folgenden an, dass die Zufallsvariablen  $X_1, \dots, X_n$  iid sind mit identischer Verteilungsfunktion  $F$  der Form

$$F(t) = G\left(\frac{t - \mu}{\sigma}\right), t \in \mathbb{R}.$$

Dann gilt gemäß den Korollaren 10.7 und 10.10 sowie mit Lemma 10.8

$$X_{i:n} = F^{\leftarrow}(U_{i:n}) = \mu + \sigma G^{\leftarrow}(U_{i:n}), i = 1, \dots, n.$$

Plotten wir nun  $X_{k:n}$  gegen  $G^{\leftarrow}(k/(n+1))$ , d. h. tragen wir in einem Koordinatensystem die Punkte

$$\left(G^{\leftarrow}\left(\frac{k}{n+1}\right), X_{k:n}\right), k = 1, \dots, n,$$

ab, so erhalten wir einen **Quantile-Quantile Plot** oder **QQ-Plot**. Man beachte noch, dass die der Größe nach geordneten  $X_{k:n}$  entlang der senkrechten Ordinaten-Achse aufgetragen werden. Aufgrund des obigen Korollars 10.10 wird im Fall einer stetigen Quantilfunktion  $G^{\leftarrow} : (0, 1) \rightarrow \mathbb{R}$  die Approximation

$$X_{k:n} \sim F^{\leftarrow}\left(\frac{k}{n+1}\right) = \mu + \sigma G^{\leftarrow}\left(\frac{k}{n+1}\right)$$

gelten, so dass die Punkte

$$(G^{\leftarrow}(k/(n+1)), X_{k:n}) \sim (G^{\leftarrow}(k/(n+1)), \mu + \sigma G^{\leftarrow}(k/(n+1)))$$

in etwa auf der Geraden  $s = \mu + \sigma t \simeq \bar{X}_n + S_n t$ ,  $t \in \mathbb{R}$  liegen werden. Dabei bezeichnet  $S_n = s_n(X_1, \dots, X_n)$  die Standardabweichung zu  $X_1, \dots, X_n$  und  $\bar{X}_n$  deren arithmetisches Mittel.

Wir wollen nun untersuchen, ob die Gewichts-Daten von Beispiel 7.14 in etwa normalverteilt sind. Bild 30 zeigt den QQ-Plot bzgl.  $G = \Phi$ , wobei  $\Phi$  wieder die Standardnormalverteilungsfunktion bezeichnet. Die geplotteten Punkte liegen in etwa auf einer Geraden. Der QQ-Plot spricht also nicht gegen eine Normalverteilungsannahme. Achsenabschnitt (beim Abszissenwert 0) und Steigung der KQ-Geraden durch die Daten

$$\left(\Phi^{-1}\left(\frac{k}{n+1}\right), X_{k:n}\right)_{k=1, \dots, n}$$

stimmen sehr gut mit den Schätzungen  $\bar{x} = 61.49$  und  $s = 2.58$  für Mittelwert und Standardabweichung überein. Bei der Standardnormalverteilung gilt gemäß Punkt 2.

von Lemma 10.6:  $\Phi^{\leftarrow} = \Phi^{-1}$ .

**Bild 30:** QQ-Plot der Daten von Beispiel 7.14 für  $G = \Phi$

Die folgenden Grafiken zeigen, dass QQ-Plots bei der Festlegung des korrekten Verteilungsmodells ein mächtigeres Hilfsmittel sind als der optische Eindruck, den Histogramme vermitteln. Bild 31 stellt das Histogramm und den QQ-Plot für 100 simulierte standardnormalverteilte Daten dar.

**Bild 31:** QQ-Plot 100 standardnormalverteilter Daten gegen Quantile von  $G = \Phi$

Das Histogramm der 1000  $t$ -verteilten Daten (mit Freiheitsgrad 4) des folgenden Bildes 32 wirkt optisch symmetrischer und legt eher eine Verwandtschaft mit der Gaußschen Glockenkurve nahe als das Histogramm der immerhin 100 normalverteilten Daten von Bild 31. Beim QQ-Plot zeigt sich aber an den Enden jene typische Abweichung der Daten nach unten (linkes Ende) bzw. nach oben (rechtes Ende) von der angepaßten KQ-Geraden, das für Verteilungen mit *heavy tails* charakteristisch ist. Die Masse solcher Verteilungen ist (bei geeigneter Standardisierung) "nicht so eng" um den Mittelwert konzentriert wie etwa die Masse der Standardnormalverteilung.

**Bild 32:** QQ-Plot 1000  $t_4$ -verteilter Daten gegen Quantile von  $G = \Phi$

## 10.3 Box-Plots

Hier stellen wir ein anschauliches Instrument zu einer ersten Datenbegutachtung vor. Folgende Fragen sind interessant:

1. Ist die empirische Verteilung der Daten symmetrisch ?
2. Spricht etwas dagegen, dass die beobachteten Daten unabhängige Realisierungen einer normalverteilten Zufallsvariablen (ZV) sind ?
3. Wie eng (bezogen auf die Standardabweichung) streuen die Beobachtungen um den Mittelwert ?
4. Gibt es Ausreißer, d. h. Beobachtungen, die "sehr weit" vom Mittelwert entfernt liegen ?

Wir wollen nun diese Fragen einer quantifizierbaren Untersuchung zugänglich machen. Als Referenz dient (wie könnte es anders sein) die Normalverteilung mit Mittelwert  $\mu \in \mathbb{R}$ , Varianz  $\sigma^2 > 0$  und der Dichte

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

Die Dichte der  $N(0, 1)$ -Standardnormalverteilung bezeichnen wir wieder mit  $\varphi$  und die zugehörige Verteilungsfunktion mit  $\Phi$ .

Für eine beliebige Zufallsvariable  $X$  mit Verteilungsfunktion  $F$  bezeichne gemäß Def. 1.50

$$x_p := \inf \{ x \in \mathbb{R} \mid F(x) \geq p \}, \quad p \in (0, 1) \quad (10.3)$$

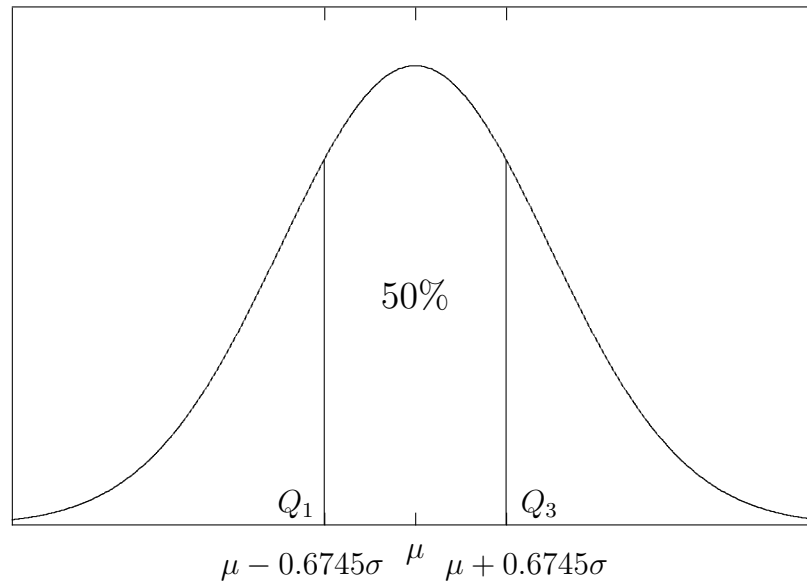
das  $p$ -Quantil.  $Q_1 = x_{0.25}$  bzw.  $Q_3 = x_{0.75}$  heißen **erstes** bzw. **drittes Quartil** und  $Q_2 = x_{0.5}$  **Median**. Dementsprechend lautet dann der **Interquartilsabstand**  $IQR = Q_3 - Q_1$ .

### Quantile und Quartile der Standardnormalverteilung

Nun sei  $u_p$  speziell das  $p$ -Quantil für die Standardnormalverteilung mit

$$\Phi(u_p) = p, \quad 0 < p < 1.$$

Aus den  $N(0, 1)$ -Tabellen des Anhangs 11.3 entnimmt man  $Q_3 = u_{0.75} \simeq 0.675$ . Der genaue Wert lautet  $u_{0.75} = 0.6745$ . Weiterhin gilt hier aus Symmetriegründen  $\Phi(-u) = 1 - \Phi(u)$ , und daher  $u_{1-p} = -u_p$ . Also  $Q_1 = -0.6745$ ,  $Q_2 = 0$  und  $IQR = 2\tau = 1.349$ , was  $\tau = 0.6745$  zur Folge hat.



**Bild 33:** Quartile der  $N(\mu, \sigma^2)$ -Verteilung

Man beachte, dass die Normalverteilung stärker um den Mittelwert konzentriert ist als viele andere Verteilungen. Dies drückt sich darin aus, dass der Semiquartilsabstand  $\tau$  einen kleineren Bruchteil der Standardabweichung ausmacht als beispielsweise bei der Gleichverteilung.

Verteilung	$\tau$	$\sigma$
Normalverteilung	$0.675 \sigma$	$1.483 \tau$
Gleichverteilung	$0.865 \sigma$	$1.156 \tau$

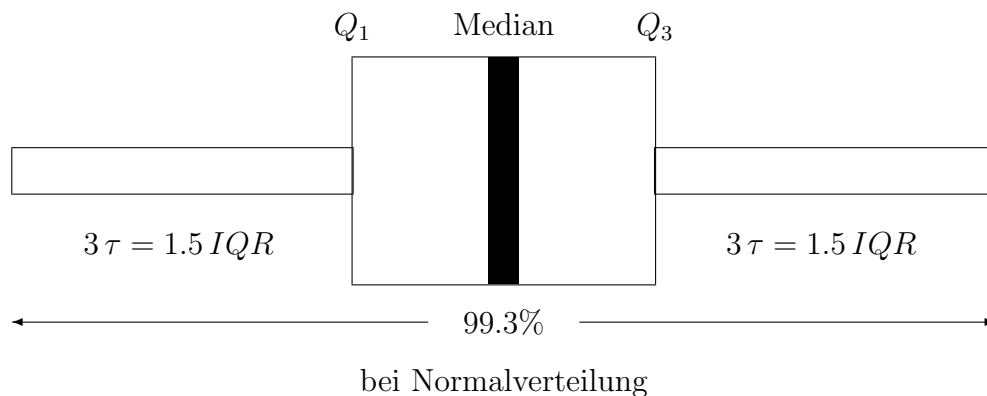
Der Übersichtlichkeit halber seien hier nochmals wichtige symmetrische p-Quantile  $\lambda_p := u_{(1+p)/2}$  der Standardnormalverteilung zusammengestellt; vgl. auch Tabelle 12

Fläche $P\%$	symmetrisches Intervall $[\mu - \sigma \cdot \lambda_{P\%}; \mu + \sigma \cdot \lambda_{P\%}]$	symmetrisches Quantil $\lambda_{P\%}$
50%	$[\mu - 0.6745 \sigma; \mu + 0.6745 \sigma]$	0.6745
68.26%	$[\mu - \sigma; \mu + \sigma]$	1
95.46%	$[\mu - 2 \sigma; \mu + 2 \sigma]$	2
99%	$[\mu - 2.576 \sigma; \mu + 2.576 \sigma]$	2.576
99.31%	$[\mu - 4 \tau; \mu + 4 \tau]$	2.698
99.73%	$[\mu - 3 \sigma; \mu + 3 \sigma]$	3

Das Intervall  $[\mu - 4\tau; \mu + 4\tau]$  kann wegen  $Q_1 = \mu - \tau$  und  $Q_3 = \mu + \tau$  auch in der Form

$$[Q_1 - 3\tau; Q_3 + 3\tau] = [Q_1 - 1.5 IQR; Q_3 + 1.5 IQR] \quad (10.4)$$

dargestellt werden. Gemäß obiger Tabelle müssen über 99% normalverteilter Daten in diesem Intervall liegen. Ist dies nicht der Fall, so muss von einer Abweichung der Daten von der Normalverteilung ausgegangen werden. Eine einfache grafische Darstellung dieses Zusammenhangs sind sog. **Box-Plots**.



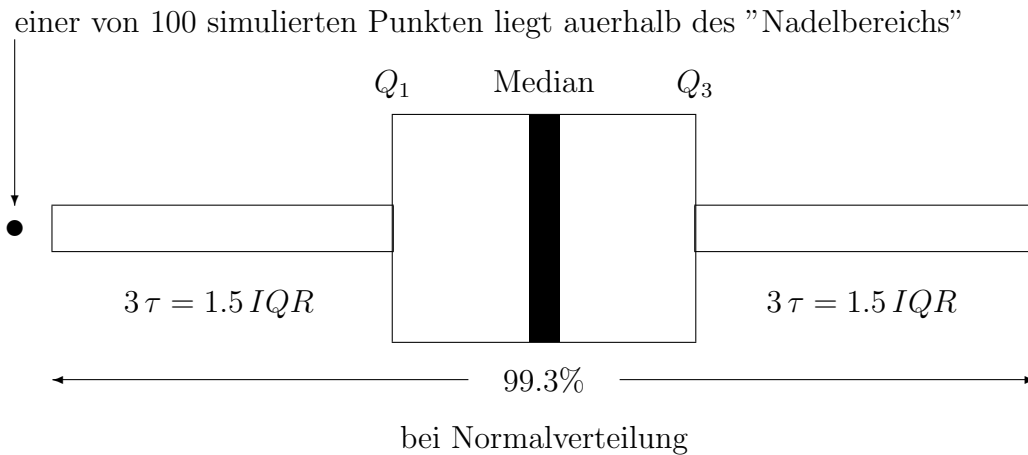
**Bild 34: Idealer Box-Plot für  $N(\mu, \sigma^2)$**

Bei ideal normalverteilten Daten wird der Median (dicker senkrechter Balken) genau in der Mitte der Box (linkes Ende  $Q_1$ , rechtes Ende  $Q_3$ ) liegen. Die Nadeln (whiskers) links und rechts der Box sind höchstens  $3\tau = 1.5 IQR$  lang. Falls der kleinste Datenpunkt  $x_{min}$  innerhalb der Spanne  $Q_1 - 3\tau$  liegt, so endet die linke Nadel bereits bei  $x_{min}$ ; analog rechts bei  $x_{max}$ .

In der Praxis sind die in Def. 1.50 erklärten Werte  $Q_1$ ,  $Q_2$  und  $Q_3$  sowie  $IQR = Q_3 - Q_1$  normalerweise nicht bekannt. Deswegen ersetzt man diese Größen durch ihre empirischen Analoga von Def. 7.18. Für die folgenden Überlegungen sind also immer die empirischen Quartile oder die davon abgeleiteten Größen  $\tau$  bzw.  $IQR$  heranzuziehen.

Der Box-Plot ist so eingerichtet, dass bei normalverteilten Daten nur *sehr selten* Werte links und rechts außerhalb der Nadeln liegen. Natürlich ist dieser Fall auch bei exakt normalverteilten Daten nicht ganz ausgeschlossen.





**Bild 35: Box-Plot für 100  $N(\mu, \sigma^2)$  -Zufallszahlen**

Bild 35 zeigt den Box-Plot von 100 simulierten normalverteilten Daten. Einer dieser Werte liegt außerhalb des Nadelbereichs, der damit 99% der Daten überdeckt. Übrigens kommt die theoretisch vorliegende Symmetrie auch grafisch im Box-Plot schön zum Ausdruck.

Die bereits in Beispiel 7.14 untersuchten Körpergewichte

Tabelle 20: Urliste Körpergewichte										
57.8	61.3	60.1	60.0	63.7	60.5	64.8	62.2	58.1	65.9	61.1
63.2	56.2	64.4	61.9	59.7	60.2	61.3	65.6	63.2	58.7	62.8

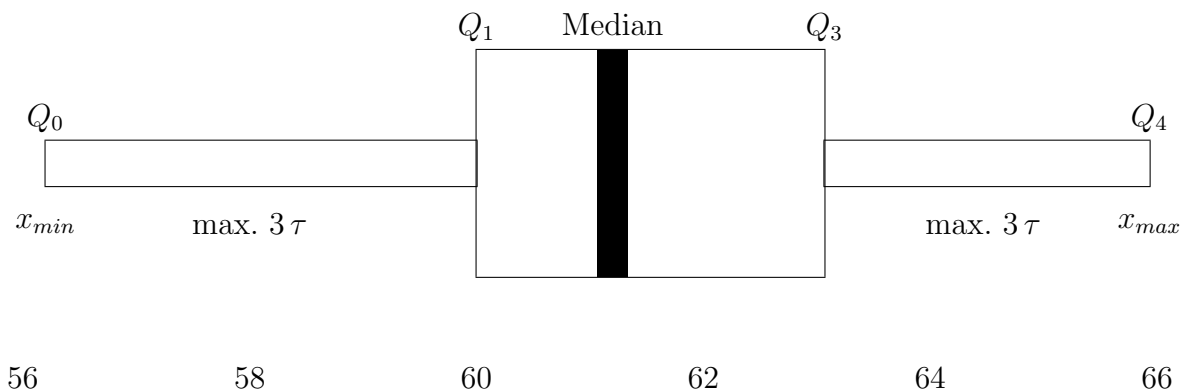
mit  $Q_0 = 56.2 = x_{min}$ ;  $Q_1 = 60$ ;  $Q_2 = 61.3 = \text{Median}$ ;  $Q_3 = 63.2$ ;  
 $Q_4 = 65.9 = x_{max}$ ;  $\tau = (Q_3 - Q_1)/2 = 1.6$ ;

$$Q_1 - 3\tau = 60 - 4.8 = 55.2 < x_{min} = 56.2$$

$$Q_3 + 3\tau = 63.2 + 4.8 = 68 > x_{max} = 65.9$$

liefern folgende Grafik:

alle 22 Datenpunkte liegen innerhalb des Nadelsbereichs



**Bild 36: Box-Plot für das Beispiel Körpergewichte**

Da  $x_{max}$  näher bei  $Q_3$  liegt als  $x_{min}$  bei  $Q_1$ , ist die rechte Nadel kürzer als die linke.

Bewertung eines Box-Plots. Sind die Daten normalverteilt?

1. Ein Box-Plot kann nur Anhaltspunkte und Indizien für oder gegen die Normalverteilungsannahme liefern. In der Regel sind Verteilungstests zur Prüfung der Normalverteilungshypothese durchzuführen. Außerdem sollten QQ-Plots (vgl. Abschnitt 10.2) und Größen wie Schiefe und Exzess zur Beurteilung der Daten mit herangezogen werden. Falls Normalverteilung vorliegt, dürfen die empirischen Werte für Schiefe und Exzess nicht zu weit von 0 abweichen.
2. Für eine Normalverteilung der Daten von Bild 36 spricht, dass kein Wert weit links oder rechts außerhalb der Nadelspitzen liegt.
3. Eher gegen die Normalverteilungsannahme spricht, dass weder der Median symmetrisch in der Box liegt noch die Nadeln links und rechts einigermaßen gleichlang sind.

Für weitere Details sei auf Lehrbücher wie [Falk et al. (1995)] und [Fahrmeir et al. (1997)] hingewiesen.

# Kapitel 11

## Anhang, Tabellen, Approximationen

### 11.1 Asymptotische Formeln

Zur Berechnung der Verteilungsfunktionen bzw. der Quantile der Normal-,  $\chi^2$ -,  $t$ - bzw.  $F$ -Verteilung liegen in den wichtigen Statistikprogramm Paketen Routinen vor. Trotzdem wollen wir hier noch einige Naherungs- bzw. asymptotische Formeln angeben, die gelegentlich nutzlich sind.

#### 1. Reihendarstellung der Standardnormalverteilungsfunktion $\Phi$

Durch gliedweise Integration der Reihe der Standardnormalverteilungsdichte erhalt man:

$$\Phi(u) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k u^{2k+1}}{2 \cdot 4 \cdot 6 \cdot \dots \cdot 2k \cdot (2k+1)} \quad (11.1)$$

Fur numerische Zwecke eignet sich die folgende Darstellung, die man durch partielle Integration gewinnt, besser:

$$\Phi(u) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \sum_{k=0}^{\infty} \frac{u^{2k+1}}{1 \cdot 3 \cdot 5 \cdot \dots \cdot (2k+1)} \quad (11.2)$$

#### 2. Asymptotische Approximationen der $\chi^2$ -Verteilung

$$\chi_{\nu;p}^2 \simeq \frac{1}{2}(\sqrt{2\nu-1} + u_p)^2 \quad \text{fur Freiheitsgrade } \nu > 100$$

$$\chi_{\nu;p}^2 \simeq \nu(1 - e + u_p\sqrt{e})^3, e = \frac{2}{9\nu} \quad \text{fur Freiheitsgrade } \nu > 50.$$

Fur  $\nu = 85$  und  $p = 0.95$  gilt also mit  $u_{0.95} = 1.645$

<b>Tabelle 21:</b> Approximation der $\chi^2$ -Quantile		
$\chi_{85;0.95}^2$	Formel	Art der Berechnung
107.521	—————	exakt
107.5	$(101.9 + 113.1)/2$	aus Tabelle interpoliert
107.24	$0.5(\sqrt{2 \cdot 85 - 1} + 1.645)^2$	grobe Formel (1)
107.52	$85 \cdot (1 - 0.002641 + 1.645 \cdot \sqrt{0.002641})^3$ $e = 2/(9 \cdot 85) = 0.002641$	genauere Formel (2)

### 3. Asymptotik der $t$ -Verteilung

$$t_{\nu;p} \simeq u_p + \frac{u_p}{4\nu}(u_p^2 + 1) + \frac{1}{96\nu^2}(5u_p^5 + 16u_p^3 + 3u_p) = t_{\nu;0.95}^{appr}.$$

Mit dieser Näherung erhalten wir beispielsweise für  $p = 0.95$  und  $u_{0.95} = 1.645$

<b>Tabelle 22:</b> $t$ -Approximation		
$\nu$	$t_{\nu;0.95}^{appr}$	$t_{\nu;0.95}^{exakt}$
15	1.75275	1.75305
65	1.66863	1.66864
85	1.66298	1.66298

### 4. Asymptotik der $F$ -Verteilung

Zur näherungsweisen Berechnung der Quantile können folgende Grenzeigenschaften der Verteilungsfunktionen herangezogen werden

$$\lim_{n \rightarrow \infty} F_{m,n}(x) = \chi_m^2(mx),$$

$$\lim_{m \rightarrow \infty} F_{m,n}(x) = 1 - \chi_n^2\left(\frac{n}{x}\right).$$

Sind  $g_m$  bzw.  $g_n$  die Dichten von  $\chi_m^2$ - bzw.  $\chi_n^2$ -verteilten Zufallsvariablen, so bedeutet dies für die Dichten

$$\lim_{n \rightarrow \infty} f_{m,n}(x) = mg_m(mx),$$

$$\lim_{m \rightarrow \infty} f_{m,n}(x) = \frac{n}{x^2} g_n\left(\frac{n}{x}\right).$$

## 11.2 Rationale Bestapproximationen für $\Phi$ und $\Phi^{-1}$

Die Standardnormalverteilungsfunktion und deren Quantile können beliebig genau approximiert werden. Wir begnügen uns hier mit ca. 6 Stellen. Eine für die Praxis ausreichende Näherung  $\tilde{\Phi}(u)$  von

$$\Phi(u) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{t^2}{2}} dt$$

für  $u \geq 0$  erhält man mit

$$\Psi(u) := e^{-u^2} \frac{a_0 + a_1 u + a_2 u^2}{b_0 + b_1 u + b_2 u^2 + u^3}, \quad \text{wobei}$$

$$\begin{aligned} a_0 &= 2.6938, & a_1 &= 1.9273, & a_2 &= 0.56656, \\ b_0 &= 2.6938, & b_1 &= 4.9678, & b_2 &= 3.4710, & (b_3 = 1), \end{aligned}$$

und

$$\tilde{\Phi}(u) = 1 - \frac{1}{2} \Psi\left(\frac{u}{\sqrt{2}}\right).$$

Für negative Argumente nützt man die Identität:  $\Phi(-u) = 1 - \Phi(u)$ . Übrigens gilt:

$$|\tilde{\Phi}(u) - \Phi(u)| \leq 5.5 \cdot 10^{-6} \text{ für alle } u \geq 0.$$

Bei der Berechnung der Umkehrfunktion  $\Phi^{-1}(p)$  sucht man für  $p \in (0, 1)$  ein  $u \in \mathbb{R}$ , so dass  $\Phi(u) = p$ . Für  $p \leq 0.5$  erhält man eine Näherung  $\tilde{u}$  für  $u$  durch:

$$\tilde{u}(p) := \frac{a_0 + a_1 t + a_2 t^2}{1 + b_1 t + b_2 t^2 + b_3 t^3} - t, \quad \text{wobei } t = \sqrt{-2 \ln(p)} \text{ und}$$

$$\begin{aligned} a_0 &= 2.515517, & a_1 &= 0.802853, & a_2 &= 0.010328, \\ b_1 &= 1.432788, & b_2 &= 0.189269, & b_3 &= 0.001308. \end{aligned}$$

Für Argumente  $0.5 < p < 1$  nutzt man wiederum  $\Phi(-u) = 1 - \Phi(u)$  und erhält jetzt  $\Phi^{-1}(p) = -\Phi^{-1}(1 - p)$  und die Fehlerabschätzung:

$$|\tilde{u}(p) - \Phi^{-1}(p)| \leq 4.5 \cdot 10^{-4} \text{ für alle } p \in (0, 1).$$

Algorithmen für genauere Approximationen findet man in:

*Abramowitz, M. and Stegun, I.A.* (eds.): Handbook of Mathematical Functions. Dover Publications, New York 1965.

*Hart J.F.* (ed.): Computer Approximations, 2nd edition. Krieger Publ. Comp., Huntington 1978.

## 11.3 Tabellen

Es folgen Tabellen für die Standardnormalverteilung und für Stichprobenverteilungen, die aus der Normalverteilung abgeleitet sind.

$u$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57534
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79954	.80234	.80510	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84613	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87697	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89616	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91773
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96637	.96711	.96784	.96856	.96926	.96994	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98213	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98777	.98809	.98840	.98870	.98899
2.3	.98928	.98955	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99157
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99491	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99830	.99836	.99841	.99846	.99851	.99856	.99860
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99909	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99967	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99977	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99991	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99994	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99996	.99997

Normalverteilung

$\nu$	$p$	0.005	0.01	0.025	0.05	0.10	0.20	0.50	0.80	0.90	0.95	0.975	0.99	0.995
1		.04393	.03157	.02398	.01793	.01579	.06418	.4549	1.642	2.706	3.841	5.024	6.635	7.879
2		.01003	.02010	.05064	.1026	.2107	.4463	1.386	3.219	4.605	5.991	7.378	9.210	10.60
3		.07172	.1148	.2158	.3518	.5844	1.005	2.366	4.642	6.251	7.815	9.348	11.34	12.84
4		.2070	.2971	.4844	.7107	1.064	1.649	3.357	5.989	7.779	9.488	11.14	13.28	14.86
5		.4117	.5543	.8312	1.145	1.610	2.343	4.351	7.289	9.236	11.07	12.83	15.09	16.75
6		.6757	.8721	1.237	1.635	2.204	3.070	5.348	8.558	10.64	12.59	14.45	16.81	18.55
7		.9893	1.239	1.690	2.167	2.833	3.822	6.346	9.803	12.02	14.07	16.01	18.48	20.28
8		1.344	1.646	2.180	2.733	3.490	4.594	7.344	11.03	13.36	15.51	17.53	20.09	21.95
9		1.735	2.088	2.700	3.325	4.168	5.380	8.343	12.24	14.68	16.92	19.02	21.67	23.59
10		2.156	2.558	3.247	3.940	4.865	6.179	9.342	13.44	15.99	18.31	20.48	23.21	25.19
11		2.603	3.053	3.816	4.575	5.578	6.989	10.34	14.63	17.28	19.68	21.92	24.72	26.76
12		3.074	3.571	4.404	5.226	6.304	7.807	11.34	15.81	18.55	21.03	23.34	26.22	28.30
13		3.565	4.107	5.009	5.892	7.042	8.634	12.34	16.98	19.81	22.36	24.74	27.69	29.82
14		4.075	4.660	5.629	6.571	7.790	9.467	13.34	18.15	21.06	23.68	26.12	29.14	31.32
15		4.601	5.229	6.262	7.261	8.547	10.31	14.34	19.31	22.31	25.00	27.49	30.58	32.80
16		5.142	5.812	6.908	7.962	9.312	11.15	15.34	20.47	23.54	26.30	28.85	32.00	34.27
17		5.697	6.408	7.564	8.672	10.09	12.00	16.34	21.61	24.77	27.59	30.19	33.41	35.72
18		6.265	7.015	8.231	9.390	10.86	12.86	17.34	22.76	25.99	28.87	31.53	34.81	37.16
19		6.844	7.633	8.907	10.12	11.65	13.72	18.34	23.90	27.20	30.14	32.85	36.19	38.58
20		7.434	8.260	9.591	10.85	12.44	14.58	19.34	25.04	28.41	31.41	34.17	37.57	40.00
21		8.034	8.897	10.28	11.59	13.24	15.44	20.34	26.17	29.62	32.67	35.48	38.93	41.40
22		8.643	9.542	10.98	12.34	14.04	16.31	21.34	27.30	30.81	33.92	36.78	40.29	42.80
23		9.260	10.20	11.69	13.09	14.85	17.19	22.34	28.43	32.01	35.17	38.08	41.64	44.18
24		9.886	10.86	12.40	13.85	15.66	18.06	23.34	29.55	33.20	36.42	39.36	42.98	45.56
25		10.52	11.52	13.12	14.61	16.47	18.94	24.34	30.68	34.38	37.65	40.65	44.31	46.93
26		11.16	12.20	13.84	15.38	17.29	19.82	25.34	31.79	35.56	38.89	41.92	45.64	48.29
27		11.81	12.88	14.57	16.15	18.11	20.70	26.34	32.91	36.74	40.11	43.19	46.96	49.64
28		12.46	13.56	15.31	16.93	18.94	21.59	27.34	34.03	37.92	41.34	44.46	48.28	50.99
29		13.12	14.26	16.05	17.71	19.77	22.48	28.34	35.14	39.09	42.56	45.72	49.59	52.34
30		13.79	14.95	16.79	18.49	20.60	23.36	29.34	36.25	40.26	43.77	46.98	50.89	53.67
31		14.46	15.66	17.54	19.28	21.43	24.26	30.34	37.36	41.42	44.99	48.23	52.19	55.00
32		15.13	16.36	18.29	20.07	22.27	25.15	31.34	38.47	42.58	46.19	49.48	53.49	56.33
33		15.82	17.07	19.05	20.87	23.11	26.04	32.34	39.57	43.75	47.40	50.73	54.78	57.65
34		16.50	17.79	19.81	21.66	23.95	26.94	33.34	40.68	44.90	48.60	51.97	56.06	58.96
35		17.19	18.51	20.57	22.47	24.80	27.84	34.34	41.78	46.06	49.80	53.20	57.34	60.27
36		17.89	19.23	21.34	23.27	25.64	28.73	35.34	42.88	47.21	51.00	54.44	58.62	61.58
37		18.59	19.96	22.11	24.07	26.49	29.64	36.34	43.98	48.36	52.19	55.67	59.89	62.88
38		19.29	20.69	22.88	24.88	27.34	30.54	37.34	45.08	49.51	53.38	56.90	61.16	64.18
39		20.00	21.43	23.65	25.70	28.20	31.44	38.34	46.17	50.66	54.57	58.12	62.43	65.48
40		20.71	22.16	24.43	26.51	29.05	32.34	39.34	47.27	51.81	55.76	59.34	63.69	66.77
41		21.42	22.91	25.21	27.33	29.91	33.25	40.34	48.36	52.95	56.94	60.56	64.95	68.05
42		22.14	23.65	26.00	28.14	30.77	34.16	41.34	49.46	54.09	58.12	61.78	66.21	69.34
43		22.86	24.40	26.79	28.96	31.63	35.07	42.34	50.55	55.23	59.30	62.99	67.46	70.62
44		23.58	25.15	27.57	29.79	32.49	35.97	43.34	51.64	56.37	60.48	64.20	68.71	71.89
45		24.31	25.90	28.37	30.61	33.35	36.88	44.34	52.73	57.51	61.66	65.41	69.96	73.17
46		25.04	26.66	29.16	31.44	34.22	37.80	45.34	53.82	58.64	62.83	66.62	71.20	74.44
47		25.77	27.42	29.96	32.27	35.08	38.71	46.34	54.91	59.77	64.00	67.82	72.44	75.70
48		26.51	28.18	30.75	33.10	35.95	39.62	47.34	55.99	60.91	65.17	69.02	73.68	76.97
49		27.25	28.94	31.55	33.93	36.82	40.53	48.33	57.08	62.04	66.34	70.22	74.92	78.23
50		27.99	29.71	32.36	34.76	37.69	41.45	49.33	58.16	63.17	67.50	71.42	76.15	79.49
60		35.53	37.48	40.48	43.19	46.46	50.64	59.33	68.97	74.40	79.08	83.30	88.38	91.95
70		43.28	45.44	48.76	51.74	55.33	59.90	69.33	79.71	85.53	90.53	95.02	100.4	104.2
80		51.17	53.54	57.15	60.39	64.28	69.21	79.33	90.41	96.58	101.9	106.6	112.3	116.3
90		59.20	61.75	65.65	69.13	73.29	78.56	89.33	101.1	107.6	113.1	118.1	124.1	128.3
100		67.33	70.06	74.22	77.93	82.36	87.95	99.33	111.7	118.5	124.3	129.6	135.8	140.2

Ablesebeispiele:  $P\{\chi_{\nu}^2 \leq 4.404\} = 0.025 = 2.5\%$   $\nu = 12$ ;  $.043927 = .00003927$

Asymptotische Formel:  $\chi_{\nu,p}^2 \simeq \frac{1}{2}(\sqrt{2\nu-1} + u_p)^2$  für Freiheitsgrade  $\nu > 100$ .

**$\chi^2$ -Verteilung**

$\nu$	0.90	0.95	0.975	0.99	0.995	0.999	0.9995
1	3.078	6.314	12.71	31.82	63.66	318.3	636.6
2	1.886	2.920	4.303	6.965	9.925	22.33	31.60
3	1.638	2.353	3.182	4.541	5.841	10.21	12.92
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.708	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.859	2.306	2.896	3.355	4.500	5.041
9	1.383	1.833	2.262	2.821	3.250	4.296	4.781
10	1.372	1.812	2.228	2.764	3.169	4.143	4.587
11	1.363	1.796	2.201	2.718	3.106	4.024	4.437
12	1.356	1.782	2.179	2.681	3.054	3.929	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.732	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.645	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.551	3.849
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.466	3.745
25	1.316	1.708	2.059	2.485	2.787	3.450	3.725
26	1.315	1.706	2.055	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	1.296	1.671	2.000	2.390	2.660	3.231	3.460
80	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	1.290	1.660	1.984	2.364	2.626	3.174	3.390
120	1.289	1.658	1.980	2.358	2.617	3.159	3.373
200	1.286	1.653	1.972	2.345	2.601	3.131	3.340
500	1.283	1.648	1.965	2.334	2.586	3.106	3.310
$\infty$	1.282	1.645	1.960	2.326	2.576	3.090	3.291

 **$t$ -Verteilung**



$f_2$	$f_1$	1	2	3	4	5	6	7	8	9	10	15	20	30	40	50	100	200	500	$\infty$
1	161	200	216	225	230	234	237	239	241	242	246	248	250	251	252	253	254	254	254	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	19.5
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.62	8.59	8.58	8.55	8.54	8.53	8.53	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.75	5.72	5.70	5.66	5.65	5.64	5.63	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.50	4.46	4.44	4.41	4.39	4.37	4.36	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.81	3.77	3.75	3.71	3.69	3.68	3.67	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.38	3.34	3.32	3.27	3.25	3.24	3.23	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.08	3.04	3.02	2.97	2.95	2.94	2.93	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.86	2.83	2.80	2.76	2.73	2.72	2.71	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.70	2.66	2.64	2.59	2.56	2.55	2.54	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.57	2.53	2.51	2.46	2.43	2.42	2.40	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.47	2.43	2.40	2.35	2.32	2.31	2.30	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.38	2.34	2.31	2.26	2.23	2.22	2.21	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.31	2.27	2.24	2.19	2.16	2.14	2.13	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.25	2.20	2.18	2.12	2.10	2.08	2.07	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28	2.19	2.15	2.12	2.07	2.04	2.02	2.01	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23	2.15	2.10	2.08	2.02	1.99	1.97	1.96	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19	2.11	2.06	2.04	1.98	1.95	1.93	1.92	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23	2.16	2.07	2.03	2.00	1.94	1.91	1.89	1.88	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.04	1.99	1.97	1.91	1.88	1.86	1.84	1.84
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.09	2.01	1.92	1.87	1.84	1.78	1.75	1.73	1.71	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.84	1.79	1.76	1.70	1.66	1.64	1.62	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.74	1.69	1.66	1.59	1.55	1.53	1.51	1.51
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.87	1.78	1.69	1.63	1.60	1.52	1.48	1.46	1.44	1.44
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75	1.65	1.59	1.56	1.48	1.44	1.41	1.39	1.39
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.81	1.72	1.62	1.57	1.53	1.45	1.40	1.37	1.35	1.35
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.79	1.70	1.60	1.54	1.51	1.43	1.38	1.35	1.32	1.32
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.78	1.69	1.59	1.53	1.49	1.41	1.36	1.33	1.30	1.30
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.77	1.68	1.57	1.52	1.48	1.39	1.34	1.31	1.28	1.28
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.73	1.64	1.54	1.48	1.44	1.34	1.29	1.25	1.22	1.22
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.72	1.62	1.52	1.46	1.41	1.32	1.26	1.22	1.19	1.19
300	3.87	3.03	2.63	2.40	2.24	2.13	2.04	1.97	1.91	1.86	1.70	1.61	1.50	1.43	1.39	1.30	1.23	1.19	1.15	1.15
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.69	1.59	1.48	1.42	1.38	1.28	1.21	1.16	1.11	1.11
$\infty$	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.67	1.57	1.46	1.39	1.35	1.24	1.17	1.11	1.00	1.00

## 95%-Quantile der F-Verteilung

$$f_1 = m, f_2 = n$$

## 11.4 Approximation der Binomial- durch die Normal- bzw. Poissonverteilung

Seien  $n \in \mathbb{N}$ ,  $0 < p < 1$  und  $\lambda = np$  und weiterhin

$$b_{k,n,p} := \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n,$$

$$v_{k,\lambda} := \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathbb{N}_0,$$

$$NV_{k,n,p} = \Phi\left(\frac{k+0.5-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-0.5-np}{\sqrt{np(1-p)}}\right), \quad \Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-t^2/2} dt.$$

Die Approximationsgüte sowohl der Normal- als auch der Poissonverteilung wächst für großes  $n$ . Die Approximationsgüte der Normalverteilung ist für  $p \approx 1/2$  am genauesten, während die Poissonapproximation für kleines  $p$  und kleines  $k$  die besseren Ergebnisse liefert. Dies wird durch folgende Abschätzung motiviert:

Für eine beliebige Teilmenge  $A \subset \mathbb{N}_0$  sowie jedes  $0 < p < 1$  und jedes  $n \in \mathbb{N}$  gilt

$$\left| \sum_{k \in A} b_{k,n,p} - \sum_{k \in A} v_{k,np} \right| \leq p.$$

Vergleich der Binomialwahrscheinlichkeiten  $b_{k,n,p}$  mit der Normalverteilungsapproximation  $NV_{k,n,p}$  bzw. der Näherung  $v_{k,np}$  durch die Poissonverteilung.

$n = 12, \quad p = 0.4, \quad \lambda = np = 4.8$					
Binomial		Normal		Poisson	
$k$	$b_{k,n,p}$	$NV_{k,n,p}$	$ b_{k,n,p} - NV_{k,n,p} $	$v_{k,np}$	$ b_{k,n,p} - v_{k,np} $
0	0.0022	0.0047	0.0026	0.0082	0.0061
1	0.0174	0.0203	0.0029	0.0395	0.0221
2	0.0639	0.0617	0.0021	0.0948	0.0310
3	0.1419	0.1342	0.0077	0.1517	0.0098
4	0.2128	0.2080	0.0048	0.1820	0.0308
5	0.2270	0.2302	0.0031	0.1747	0.0523
6	0.1766	0.1818	0.0052	0.1398	0.0368

$n = 50, \quad p = 0.01, \quad \lambda = np = 0.5$					
Binomial		Normal		Poisson	
$k$	$b_{k,n,p}$	$NV_{k,n,p}$	$ b_{k,n,p} - NV_{k,n,p} $	$v_{k,np}$	$ b_{k,n,p} - v_{k,np} $
0	0.6050	0.4224	0.1826	0.6065	0.0015
1	0.3056	0.4224	0.1168	0.3033	0.0023
2	0.0756	0.0754	0.0002	0.0758	0.0002
3	0.0122	0.0022	0.0100	0.0126	0.0004
4	0.0015	0.0000	0.0014	0.0016	0.0001
5	0.0001	0.0000	0.0001	0.0001	0.0000

# Literaturverzeichnis

- [Becker (1993)] B.: Statistik. Oldenbourg, München.
- [Chatterjee et al. (1995)] Chatterjee S., Handcock M.S. and Simonoff J.S.: A Casebook for a First Course in Statistics and Data Analysis. Wiley, New York.
- [Chung (1979)] K.L.: Elementary Probability Theory with Stochastic Processes, 3rd Edition. Springer, New York.
- [Chung (1974)] K.L.: A Course in Probability Theory, 2nd ed. Academic Press, New York.
- [Dufner et al. (1992)] Dufner J., Jensen U. und Schumacher E.: Statistik mit SAS. Teubner, Stuttgart.
- [Embrechts & Klüppelberg & Mikosch (1997)] Embrechts P., Klüppelberg C. and Mikosch T.: Modelling Extremal Events for Insurance and Finance. Springer, Berlin.
- [Fahrmeir et al. (1997)] Fahrmeir L., Künstler R., Pigeot I. und Tutz G.: Statistik. Der Weg zur Datenanalyse. Springer, Berlin.
- [Falk et al. (1995)] Falk M., Becker R. und Mahrohn F.: Angewandte Statistik mit SAS. Springer, Berlin.
- [Fahrmeir et al. (1996)] Fahrmeir. L, Hamerle A. und Tutz G. (Hrsg.): Multivariate statistische Verfahren, 2., erweiterte Auflage. De Gruyter, Berlin.
- [Hartung (1982)] J.: Statistik. Lehr- und Handbuch der angewandten Statistik. Oldenbourg, München.
- [Heinhold & Gaede (1979)] Heinhold, J. und Gaede, K.-W.: Ingenieur-Statistik, 2. Auflage. Oldenbourg, München.
- [Henze (1997)] N.: Stochastik für Einsteiger. Vieweg, Braunschweig.
- [Isaac (1995)] R.: The Pleasures of Probability. Springer, New York.
- [Johnson & Kotz (1970)] Johnson, N.L and Kotz, S.: Continuous Univariate Distributions. Wiley, Boston.
- [Kredler & Ritter (1995)] Kredler Ch. und Ritter K.: Einführung in die Statistik. Schriftenreihe des Inst. f. Angewandte Mathematik und Statistik Nr. 4, TU München.

- [Krengel (1988)] U.: Einführung in die Wahrscheinlichkeitstheorie und Statistik, Vieweg, Braunschweig.
- [Lehn & Wegmann (1992)] Lehn, J. und Wegmann, H.: Einführung in die Statistik, 2. Auflage. Teubner, Stuttgart.
- [Müller (1975)] P.H. (Hrsg.): Lexikon der Stochastik, 2. Auflage. Akademie Verlag, Berlin.
- [Pfanzagl (1988)] J.: Elementare Wahrscheinlichkeitsrechnung. De Gruyter, Berlin.
- [Resnick (1994)] S.: Adventures in Stochastic Processes, 2nd Edition. Birkhäuser, Boston.
- [Ross (1972)] S.M.: Introduction to Probability Models. Academic Press, New York.
- [Sachs (1984)] L.: Angewandte Statistik. Springer, Berlin.
- [Seber (1977)] G.A.F.: Linear Regression Analysis. Wiley, New York.
- [Stirzaker (1994)] D.: Elementary Probability. Cambridge University Press, Cambridge.
- [Stoyan (1993)] D.: Stochastik für Ingenieure und Naturwissenschaftler. Akademie Verlag, Berlin.
- [Williams (1991)] D.: Probability with Martingales. Cambridge University Press. Cambridge.
- [Witting (1985)] H.: Mathematische Statistik 1. Teubner, Stuttgart.

# Index

- $1_A$ , 25  
 $A^c$ , 2  
 $B(n, p)$ , 109, 110  
 $CV$ , 138  
 $Cov(X, Y)$ , 55  
 $D^2(X)$ , 38  
 $E(X)$ , 36, 55  
 $ED(\lambda)$ , 27, 113  
 $F(-\infty)$ , 22  
 $F(\infty)$ , 22  
 $F_{X_1, \dots, X_n}$ , 47  
 $H(n, N, M)$ , 110  
 $IQR$ , 137, 154  
     empirisch, 137  
 $M(n, \boldsymbol{\pi})$ , 111  
 $Poi(\lambda)$ , 20, 109  
 $Q_0, Q_1, \dots$ , 154–156  
 $Q_1, Q_3$ , 30, 154, 207  
     empirisch, 133  
 $Q_2$ , 30, 133, 154, 207  
 $S^2$ , 62, 66, 159  
 $T_n$ , 160  
 $Var(X)$ , 38  
 $W(r, \lambda)$ , 118  
 $X_{(k)}$ , 79  
 $X_{k:n}$ , 79  
 $\bar{F}$ , 21, 27, 113  
 $\mathbb{N}_0$ , 1  
 $\Phi$ , 98  
 $\bar{X}_n$ , 62, 66, 159, 160  
 $\mathcal{P}$ , 3  
 $\det(C)$ , 119  
 $\epsilon$ , 134  
 $\hat{y}_i$ , 145  
 $\infty$ , 123, 131  
 $\det(C)$ , 50  
 $\mu$ , 26, 36, 207  
 $\rho(X, Y)$ , 56  
 $\sigma$ , 37, 207  
 $\sigma^2$ , 37, 207  
 $\sim$ , 25, 29  
 $\tau$ , 207  
 $T$ , 46  
 $dF(x)$ , 41  
 $n!$ , 8  
 $s, s_x$ , 136  
 $s^2$ , 136  
 $s_L$ , 137  
 $x^T$ , 50, 119  
 $x_p$ , 30  
 $x_{(1)}$ , 133  
 $x_{(n)}$ , 133  
 $x_{0.5}$ , 134  
 $x_{1:n}$ , 133  
 $x_\alpha$ , 133  
 $x_{n:n}$ , 133  
 $\binom{n}{k}$ , 8  
 $\mathfrak{F}$ , 3  
 $E(X | Y)$ , 76  
 $Y_n \xrightarrow{d} Y$ , 97  
 $Y_n \xrightarrow{P} Y$ , 95  
 $Y_n \xrightarrow{f.s.} Y$ , 96  
 $\Phi$ , 27, 214  
     Approximation, 214  
     Tabelle, 215  
 $\Phi^{-1}$ , 214  
 $\chi^2$ -Verteilung, 168, 216  
     Tabelle, 216  
 $\sigma$ -Algebra, 3  
 $\sigma$ -field, 3  
 $\varphi$ , 27  
 $f_{X, Y}$ , 74  
 Übergangszeit, 106  
 Überlebenswahrscheinlichkeit, 21, 27, 113  
 Überschreitungswahrscheinlichkeit, 187  
 $\lambda_p$ , 166  
 $\tau$ , 208–210

- $\bar{X}_n$  , 170
- IQR*
  - empirisch , 154
- IQR* , 209, 210
- $S^2$  , 170
- $u_p$  , 165, 166
  
- a.e. (almost everywhere), 96
- Abhängigkeit
  - lineare, 150
- absolute Häufigkeit, 126
- absolute Momente, 42
- Abstandsquadrate, 145
- Abweichung
  - mittlere lineare, 137
  - mittlere quadratische, 38, 136
- Achsenabschnitt, 144
- Additionssatz, 6
- Algebra
  - $\sigma$ -, 4
  - Borel- $\sigma$ -, 4
- Alternative, 174
- Annahmehereich, 175
- arithmetic mean, 135
- arithmetisches Mittel, 135, 136
- Ausreißer, 134, 135, 137, 207
- Axiome von Kolmogorov, 4
  
- Balkendiagramm, 126
- Bandbreite, 197
- bandwidth, 197
- Bayes, 11
- bedingte Erwartung, 73, 76, 77
  - $L_2$ -Approximation, 78
  - Projektionseigenschaft, 78
- bedingte Normalverteilung, 75
- bedingte Verteilung, 74
- bedingte Wahrscheinlichkeit, 10
- bedingter Erwartungswert, 72
- Beispiele
  - diskreter ZV, 39, 109
  - stetiger ZV, 39, 112
- Bereichsschätzung, 172
- Bernoulli, 7
- Bernoulli Walk, 101
- Bernoulli-Variable, 39
- Bernoullisches Experiment, 19
  
- Bestimmtheitsmaß, 147
- Betaverteilung, 170
- Beton, 184, 185
- Bias, 159
- Binomialkoeffizient, 8
- Binomialverteilung, 19, 39, 98, 99, 109, 219
  - Approx.d.Normalvert., 98
  - Approx.d.Poisson-Vert., 99
  - erzeugende Funktion, 84, 109
  - EW, Varianz, 39, 109
  - ML-Schätzer, 163
  - MV-Schätzer, 161
  - Summe von ZV, 70, 86
- bivariate Normalverteilung, 119
- BLUE-Schätzer, 161
- Boole's inequalities, 6
- Borel- $\sigma$ -Algebra, 4
- Box-Plot, 155, 156, 209, 210
- Breakdownpoint, 134
  
- Cauchy-Verteilung, 39, 40, 70, 169
- charakteristische Funktion, 87
- coefficient of variation, 43, 138
- conditional probability, 10
- conditioning, 11
- consistent, 160
- convolution, 85
- cumulants, 44, 91
  
- Daten, 125, 178
  - gruppiert, 140
- Daten-Matrix, 194
- Datenbank, 124
- de Moivre-Laplace, 97
- de Morgan, 3
- density, 23
- Design-Matrix, 194
- Determinante, 50, 119
- Diätkur, 182
- Diagramm
  - Balken-, 126
  - Kreissektoren-, 126
  - räumliches Säulen-, 126
  - Säulen-, 126
  - Streu-, 142
  - Torten-, 126

- Dichte, 23
  - der Cauchyverteilung, 116
  - der Erlangverteilung, 115
  - der Exponentialverteilung, 27, 113
  - der Gammaverteilung, 28, 115
  - der Logistischen Verteilung, 116
  - der Lognormalverteilung, 34, 117
  - der Normalverteilung, 112, 119, 165, 207
  - der Pareto-Verteilung, 116
  - der Rechteckverteilung, 26, 113
  - der Standardnormalverteilung, 27
  - der Weibull-Verteilung, 118
  - empirisch, 131
  - n-dim., 49
  - Transformationssatz, 63
  - Wahrscheinlichkeits-, 23
- Dichtefunktion, 23
  - empirisch, 130
- disjunkt, 2
- diskret, 18, 47
  - Merkmal, 122, 125
- diskrete Gleichverteilung, 74
- distribution, 17, 18, 46, 48
- distribution function, 21
  - n-dim., 47
- Dreiecks-Kern, 199
- Drift, 105
- Druckfestigkeit, 184, 185
- duration of the game, 106
- Durchschnitt, 135
- Eigenwert-Eigenvektorzerlegung, 65
- einfache Irrfahrt, 15, 101
- Einflussgröße, 195
- Einstichproben t-Test, 181
- Element, 121
- Elementarereignis, 2
- empirisch
  - Exzess, 139
  - IQR, 137
  - Median, 134
  - Quantil, 133
  - Quartil, 133
  - Schiefe, 139
- empirische
  - Dichtefunktion, 130, 131
  - Kovarianz, 142
  - Regression, 142
  - Regressionsgerade, 145
  - Stichprobe, 125
  - Varianz, 136, 142
  - Verteilungsfunktion, 128, 132
- Epanechnikov-Kern, 198
- Ereignis, 2
  - Elementar-, 2
  - sicheres, 2
  - unabhängiges, 12
- Ergebnis, 1
- Ergebnismenge, 1
- Erlangverteilung, 89, 115
  - Summe von ZV, 89, 115
- erwartungstreu, 159
- Erwartungswert, 36, 37
  - bedingt bzgl.  $B$ , 72
  - einer diskreten ZV, 36
  - einer stetigen ZV, 36
  - n-dim., 54
  - Vektor, 57
- erzeugende Funktion, 83, 84
  - Summen, 86
- estimator, 159
  - consistent, 160
  - unbiased, 159
- event space, 3
- EW, 36
  - n-dim., 54, 55
- Excel, 154
  - Analyse-Funktionen, 154
  - Extras, 154
  - Histogramm, 154
  - Mittelwerte, 154
  - Populationskenngrößen, 154
  - Streudiagramm, 154
- Exponentialfamilie, 111
- Exponentialverteilung, 27, 40, 89, 113
  - EW, Varianz, 39, 113
  - Kumulanten, 45
  - ML-Schätzer, 163
  - momenterzeugende Fkt, 88, 113
  - MV-Schätzer, 161
  - Summe von ZV, 70, 89, 113
- Exzess, 43, 139, 154–156

- F-Test
  - Zweistichproben, 185, 186
- F-Verteilung, 218
  - Tabelle, 218
- f.s. (fast sicher), 96, 98, 196
- Faltung, 68, 85
- fast sichere Konvergenz, 96
- Fehler 1. Art, 175
- Fehler 2. Art, 175, 176
- Fehlerquadratsumme, 145
- Fermat, 7
- Fisher-Verteilung, 170
- Fouriertransformierte, 87
- Freiheitsgrad, 179, 180
- function
  - distribution, 21, 47
- Funktion
  - Gamma-, 28, 115
  - Wahrscheinlichkeits-, 18, 48
- Galton, 142
- gambler's ruin, 13
- Gammafunktion, 28, 115
- Gammaverteilung, 40, 89, 115
  - EW, Varianz, 39, 115
  - Kumulanten, 45
  - Momente, 45, 115
  - momenterzeugende Fkt, 88, 115
  - Summe von ZV, 70, 89, 115
- Gauß
  - Dichte, 207
  - Glockenkurve, 207
  - Verteilung, 207
- Gaußsche Glockenkurve, 27
- Gaußverteilung, 26, 112, 165
- Gegenhypothese, 174
- generating function, 83, 84, 86
  - moments, 87
  - of the cumulants, 90
- Geometrische Verteilung, 21, 39, 110
  - EW, Varianz, 110
  - erzeugende Funktion, 84, 110
- Gesetz der großen Zahlen, 92
  - schwach, 95
  - stark, 96
- gestutztes Mittel, 138
- getrimmtes Mittel, 138
- Gleichverteilung, 26, 40, 113
  - diskret, 74
  - EW, Varianz, 39, 113
  - Kumulanten, 45
  - ML-Schätzer, 163
  - momenterzeugende Fkt, 88, 113
  - MV-Schätzer, 161
  - zentrale Momente, 45, 113
- Glivenko-Cantelli, 98, 196
- Grundgesamtheit, 121
- Grundproblem der Statistik, 158
- gruppierte Daten
  - Mittelwert und Streuung, 140
- Häufigkeit
  - absolut, 126, 127
  - kumuliert, 126, 127
  - Rechenregeln, 127
  - relativ, 126, 127
  - relative Summen-, 126
- Hauptsatz der Differential- und Integralrechnung, 23
- heavy tails, 205
- Histogramm, 130, 131
- hitting time, 106
- hochsignifikant, 184
  - von  $\mu_0$  verschieden, 181
- Hypergeometrische Verteilung, 110
  - EW, Varianz, 110
- Hypothese, 174
- identically distributed, 25
- identisch verteilt, 25
- iid, 61, 101, 125, 158
- Indikatorfunktion, 25, 128
- inter quartile range, 30, 137
- Interquartilsabstand, 137
- Intervallskala, 123
- IQR, 30, 154, 207
- Irrfahrt
  - einfach, 101
- kardinal, 123
- Kern, 197
  - Dreieck, 199
  - Epanechnikov, 198
  - Normalverteilungs, 199



- Rechteck, 199
- Kerndichteschätzer, 197, 198
- Kernschätzer, 197
- Klassenbildung
  - Regeln zur, 129
- Klassenbreite, 130
- Kleinst-Quadrat-Methode, 144, 162
- Kolmogorov, 4
- Kombinationen
  - ohne Wiederholung, 8
- Konfidenzintervall, 178
  - $\alpha$  und  $\beta$ ; einf. lin. Regr., 190
  - 95%-, 193
  - EW; einf. lin. Regr., 191
  - für  $\mu$ , 172, 179
  - Prognose; einf. lin. Regr., 191
  - Realisierung, 172
- Konfidenzniveau, 191
- konsistent, 160
- Konvergenz
  - fast sicher, 96
  - in Verteilung, 97
  - in Wahrscheinlichkeit, 95, 160
  - mit Wkt 1, 96
  - schwach, 97
  - stark, 96
  - stochastisch, 95, 160
- Korrelationskoeffizient, 56
  - empirischer, 143
- Kovarianz, 55
  - empirische, 142
  - Vektor, 58
- Kovarianzmatrix, 50, 55, 57, 58
- KQ
  - Gerade, 144
  - Methode, 144, 162
  - Schätzung, 145
- Kreissektorendiagramm, 126
- kritischer Bereich, 174, 175
  - bei zweiseitigem Test, 178
  - Konstruktion, 175
- Kumulanten, 44
  - Normalverteilung, 91, 112
  - Poisson-Verteilung, 91, 109
- Kumulantenfunktion, 90
- kurtosis, 43, 139, 154–156
- Lageparameter, 30, 36, 132
- Laplace, 7
- Laplace-Annahme, 7
- Laplace-Raum, 7
- Lebensdauer, 27, 40, 113
- Lemma
  - Glivenko-Cantelli, 98, 196
- Likelihoodfunktion, 162
- Lineare Prognose, 78
  - $L_2$ -Approximation, 78
  - 2-dim. Normalverteilung, 78
- lineare Regression, 142
- lineare Transformation, 65
- Log-Likelihoodfunktion, 162
- logarithmischer Maßstab, 130
- Logarithmustransformation, 187
- Logistische Verteilung, 116
- Lognormalverteilung, 34, 40, 117
  - EW, Varianz, 39, 117
- Lokationsmaße, 132
- marginal distribution, 51
- Massenerscheinungen, 120
- matched pairs, 181
- mathematische Stichprobe, 159
- Maximum
  - von gleichverteilten ZV, 71
  - von ZV (allgemein), 70
- Maximum-Likelihood-Methode, 162
- Maßzahl
  - statistische, 132
- mean
  - arithmetic, 135
- Median, 30, 134, 154–156, 209, 210
  - empirisch, 134
- Menge, 121
- Merkmal, 1, 122
  - Ausprägung, 121
  - diskret, 122, 123, 125
  - kardinal, 123
  - metrisch, 123, 126, 132, 134–137, 140
  - nominal, 123, 125, 132
  - ordinal, 123, 126, 132, 134, 135, 137
  - qualitativ, 122
  - quantitativ, 122
  - Spannweite, 135
  - stetig, 122, 129

- Variationsbreite, 135
- messbar, 32
- Messreihe, 125
- Methode
  - der kleinsten Quadrate, 144, 162
  - KQ, 144, 162
  - Maximum-Likelihood, 162
  - ML, 162
  - Momenten-, 162
- metrisch, 123, 126, 132, 134–137, 140
- Minimum
  - von exponentialverteilten ZV, 71
  - von ZV (allgemein), 70
- Mittel, 62, 142
  - arithmetisch, 135, 136
  - gestutzt, 138
  - getrimmt, 138
  - gruppierte Daten, 140
  - Median, 134, 136
  - Modalwert, 132
  - Modus, 132
- mittlere lineare Streuung, 137
- ML-Methode, 162
- ML-Schätzer
  - Binomialverteilung, 163
  - Exponentialverteilung, 163
  - Gleichverteilung, 163
  - Normalverteilung, 163
  - Poisson-Verteilung, 163
- Modalwert, 30, 132
- Modus, 30, 132
- moment generating function, 87, 88
- Momente, 42
  - absolut, 42
  - Existenz, 43
  - zentrale, 42
- Momentenmethode, 162
- momenterzeugende Funktion, 87
  - Summen, 88
- Monotonie, 131
- Multinomialverteilung, 111
  - EW, Kovarianz, 111
- Multiplikationsregel d.Kombinatorik, 7
- MV-Schätzer, 161
  - Binomialverteilung, 161
  - Exponentialverteilung, 161
- Gleichverteilung, 161
- Normalverteilung, 161
- Poisson-Verteilung, 161
- Nadeln in Box-Plot, 209
- negative Binomialverteilung, 110
  - EW, Varianz, 110
- nichtlinearer Zusammenhang, 151
- nominal, 123, 125, 132
- Normalgleichungen, 194
- Normalverteilung, 26, 39, 89, 112, 165, 207
  - $\sigma$ , 207
  - 2-dim., 50, 57
  - bedingt, 75
  - EW, Varianz, 39, 112
  - Exzess, 43
  - IQR, 207, 208
  - Kumulanten, 44, 45, 91, 112, 165
  - logarithmische, 34, 40, 117
  - ML-Schätzer, 163
  - momenterzeugende Fkt, 88, 112
  - MV-Schätzer für  $\mu$ , 161
  - MV-Schätzer für  $\sigma^2$ , 161
  - n-dim., 50, 119
  - Quantil, 112, 165–167, 207, 208
  - Quartil, 207
  - Schiefe, 43
  - Standard-, 26, 98, 112
  - Summe von ZV, 69, 70, 89, 112, 165
  - Tabelle, 215
  - zentrale Momente, 45, 112
  - zweidim., 119
- Normalverteilungs-Kern, 199
- Nullhypothese, 174
- Nullmenge, 76
- order statistics, 79
- ordinal, 123, 126, 132, 134, 135, 137
- Ordnungsstatistiken, 79
  - Dichte, 80
  - Gleichverteilung, 80
  - Randverteilung, 80
- p-Quantil, 30
  - der t-Verteilung, 179
- p-value, 187

- p-Wert, 187
- Parameter
  - Lage, 36
  - Streuung, 38
- Parameterraum, 101
- Pareto-Verteilung, 40, 116
  - EW, Varianz, 39, 116
- Pascal, 7
- passage time, 106
- Passierzeit, 106
- Permutation, 8
- Pfad, 15, 101
- Pictogramm, 126
- Poisson-Verteilung, 20, 39, 99, 109, 219
  - Approx.d.Normalvert., 98
  - erzeugende Funktion, 84, 109
  - EW, Varianz, 39, 109
  - Kumulanten, 45, 91, 109
  - ML-Schätzer, 163
  - momenterzeugende Fkt, 88, 89, 109
  - MV-Schätzer, 161
  - Summe von ZV, 70, 85, 86
- Polarkoordinaten, 63
- Population, 1, 121
- Porosität von Sandstein, 173, 179
- positiv definit, 50, 65, 119
- probability
  - conditional, 10
  - total, 11
- probability mass function, 18
  - 2-dim., 48
- Prognose, 144, 150
  - linear, 78
- Prognoseintervall, 191
- Programme
  - Statistik, 154
- Prozentpunkt, 29, 132
- Punktewolke, 142
- QQ-Plot, 158, 203
- Quantil, 27, 29
  - Approximation von  $u_p$ , 214
  - der  $\chi^2$ -Verteilung, 216
  - der F-Verteilung, 218
  - der t-Verteilung, 179, 217
  - empirisch, 132, 133
  - Normalverteilung, 30, 166, 167, 207, 208
    - symmetrisch, 166, 167, 208
  - quantile, 29
  - quantile-quantile plot, 203
  - Quantilfunktion, 29, 201
  - Quartil, 30, 154–156, 207, 208
    - empirisch, 133
    - Normalverteilung, 207
- Rückkehr zur 0, 106
- Rückkehrzeit, 106
- random variable, 17
- random vector, 46
- Random Walk, 15, 101, 102
  - simple, 101
  - symmetrisch, 101
- Randverteilung, 51, 59
  - diskret, 51
  - stetig, 51
- range, 135
- Realisierung, 17, 46
  - eines Vertrauensintervalls, 172
- Rechenregeln
  - für Erwartungswerte, 37
  - für Streuungen, 38
- Rechtecks-Kern, 199
- Rechteckverteilung, 26, 40, 113
  - IQR, 208
- recurrence, 106
- recurrence time, 106
- Reduktionslage, 129
- Regelfunktion, 23
- Regeln
  - zur Klassenbildung, 129
- Regression
  - einfach linear, 189
  - einfach linear emp., 142
  - empirisch, 142
  - multiple, 193
  - nichtlinear, 151
  - schrittweise, 195
- Regressionsgerade
  - empirische, 145
- Regressionskoeffizient, 189
- Regressionstrompete, 193
- Regressor, 195

- Variable, 194
- Rekurrenzzeit, 106
- relative
  - Häufigkeit, 126, 127
  - Summenhäufigkeit, 126
  - Summenhäufigkeitsfunktion, 127
- Riemann-Stieltjes-Integral, 41
- robust, 134, 137
- Ruin des Spielers, 13, 101, 103, 105
  - Spieldauer, 106
- S-Plus, 154
- Säulendiagramm, 126
  - raumlich, 126
- Saison, 148
- sample space, 1
- SAS, 154–156, 188
- Satz
  - der totalen Wkt., 11
  - Glivenko-Cantelli, 98, 196
  - von Bayes, 11
  - von Tschebyschov, 95
  - Zentraler Grenzwert-, 97
- Scatterplot, 142
- Schädlingsbekämpfungsmittel, 152
- Schätzer, 159, 197
  - erwartungstreu, 159
  - konsistent, 160
  - minimaler Varianz, 161
  - MV, 161
  - unverzerrter, 159
- Schätzfunktion, 158, 159, 175, 197
  - Konstruktion, 161
  - minimaler Varianz, 160
  - MV-, 160
  - wirksamste, 160
- Schätzung, 189
- Schiefheit (skewness), 43, 139, 154–156
- schwache Konvergenz, 97
- schwaches Gesetz der großen Zahlen, 95
- Semiquartilsabstand, 155, 156, 208
- Sicherheitswahrscheinlichkeit, 174
- signifikant
  - von  $\mu_0$  verschieden, 181
- Signifikanz-Test, 190, 195
- Signifikanzniveau, 174
- singleton, 2
- Skala
  - Intervall-, 123
  - nominal, 123
  - ordinal, 123
- skewness, 43, 139, 154–156
- Spektralzerlegung, 65
- Spieldauer, 103, 106
- SPSS, 154
- standard deviation, 136
- Standardabweichung, 38, 62, 155, 156
  - empirisch, 136
  - gruppierte Daten, 140
- Standardisierung, 39
- Standardnormalverteilung, 26, 27, 98, 112, 165
  - Approximation, 214
  - Tabelle, 215
- starke Konvergenz, 96
- starkes Gesetz der großen Zahlen, 96
- Statistik, 120, 158, 159
  - beschreibend, 121
  - Datenerfassung, 120
  - Datenpräsentation, 120
  - Datenverarbeitung, 120
  - Definition, 120, 158, 159
  - deskriptiv, 120, 121
  - empirisch, 121
  - grafische Darstellung, 120
  - induktiv, 121
  - Massenerscheinung, 120
  - Merkmalsausprägung, 121
  - schließend, 121
- Statistik-Programme, 154
- Steigung, 144
- Sterbetafel, 120
- stetig, 49
  - Merkmal, 122, 123, 129
- Stetigkeitssatz, 7
- Stichprobe, 125, 129, 196
  - empirisch, 196
  - geordnet, 133
  - iid, 61, 196
  - mathematisch, 61, 159
  - Ziehen mit Zurücklegen, 8
  - Ziehen ohne Zurücklegen, 9, 110
- Stichprobenmittel, 62, 170

- Stichprobenraum, 1
- Stichprobenumfang, 159
  - bei Normalverteilung, 173
  - bei t-Verteilung, 179
  - notwendiger, 173
- Stichprobenvarianz, 170
- stochastische Konvergenz, 95, 160
- stochastischer Prozeß, 101
- Stoppzeit, 107
- Streudiagramm, 142
- Streuung, 38
  - des Fehlers, 146
  - empirisch, 142
  - gruppierte Daten, 140
  - mittlere lineare, 137
  - mittlere quadratische, 136
- Streuungsparameter, 38, 132
- Student's t-Test, 181
- Student-Verteilung, 169, 179
- Summe von Zufallsvariablen, 68, 85, 88
- Summenhäufigkeit, 126
- Summenhäufigkeitsfunktion
  - empirisch, 132
- symmetrisch, 50, 65, 119
- t-Test, 181
  - $\sigma^2$  unbekannt, 178
  - verbundener Stichproben, 181
  - zum Vergleich von Mittelwerten, 183
  - Zweistichproben, 183, 186
- t-Verteilung, 169, 179
  - Tabelle, 217
- Tafel
  - der  $\chi^2$ -Verteilung, 216
  - der F-Verteilung, 218
  - der Standardnormalverteilung, 215
  - der t-Verteilung, 217
- tail probability, 21, 27, 28, 113
- Teilerhebung, 124
- Test
  - $\sigma^2$  bekannt, 174
  - einseitiger, 174
  - Signifikanz-, 190, 195
  - Statistik, 175, 181, 182, 187
  - statistischer, 174
  - Wilcoxon, 187
  - zweiseitiger, 178
- Tests
  - für  $\mu, \sigma^2$  bekannt, 173
- Teststatistik, 175, 181, 182, 187
- Tortendiagramm, 126
- total probability, 11
- totale Wahrscheinlichkeit, 11
- Transformation, 151
  - linear, 32, 65
  - Logarithmus, 187
  - Maximum, 32
  - Minimum, 32
  - quadratisch, 32
  - varianzstabilisierend, 187
  - Wurzel, 187
- Transformationsregel für Polarkoordinaten, 63
- Transformationssatz für Dichten, 63
- Transposition, 50, 119
- Trendgerade, 148
- Tschebyscheff, 94
- Tschebyschov
  - Satz von, 95
  - Ungleichung, 94
- unabhängig, 12
  - Funktionen von ZV, 61
  - unkorreliert, 61
  - ZV, 59
- unabhängig identisch verteilt, 61, 101
- Unabhängigkeit, 10
  - von ZV, 59
- unbiased, 159
- Ungleichung
  - en von Boole, 6
  - Basis, 93
  - Jensen, 92
  - Markov, 93
  - Tschebyschov, 93, 94
- uniform distribution, 26, 113
- unkorreliert, 55
  - unabhängig, 61
- unvereinbar, 2
- unverzerrt, 159
- Urliste, 125, 129
- Variable
  - Regressor-, 194

- Zufalls-, 17
- Variablenauswahl, 195
- variance, 136
- Varianz, 38, 62
  - einer diskreten ZV, 38
  - empirisch, 136, 142
  - gruppierte Daten, 140
  - Summe, 58
- Varianzschätzung
  - gemeinsame, 183
- varianzstabilisierend, 187
- Variationen mit Wiederholung, 8
- Variationsbreite, 129, 135
- Variationskoeffizient, 43, 138
- verallgemeinerte Inverse, 29, 201
- Verschiebungsregel, 38, 56, 136, 150
- Verteilung, 17, 18, 46, 48, 49
  - $\chi^2$ , 29, 115
  - $\chi^2$ -, 216
  - $\chi^2$ , 168
  - $F$ -, 170
  - $t$ -, 169
  - bedingt bzgl.  $B$ , 72
  - affine Trafo, 33
  - bedingt, 74
  - bedingt bzgl.  $B$ , 72
  - Beta-, 170
  - Binomial-, 19, 39, 98, 99, 109, 219
  - Cauchy-, 70, 169
  - eingipfelig, 30
  - Erlang-, 29, 89, 115
  - Exponential-, 27, 40, 89, 113
  - $F$ -, 218
  - Fisher-, 170
  - Gamma-, 40, 89, 115
  - Gauß-, 26, 112, 165, 207
  - geometrisch, 21, 39, 110
  - Gleich-, 26, 40, 113
  - hypergeometrisch, 110
  - Lebensdauer, 27, 113
  - logistisch, 116
  - Lognormal-, 34, 117
  - Multinomial-, 111
  - negative Binomial-, 110
  - Normal-, 26, 39, 89, 91, 112, 165, 207
  - Pareto, 40, 116
  - Poisson-, 20, 39, 91, 98, 99, 109, 219
  - Rechteck-, 26, 40, 113
  - stabil, 33
  - stetig, 112
  - Student-, 169, 179
  - $t$ -, 179, 217
  - unimodal, 30
  - Weibull-, 118
- Verteilungsfunktion, 21
  - der Gleichverteilung, 26
  - empirisch, 128, 131, 132, 196
  - Exponentialverteilung, 28, 113
  - Standardnormalverteilung, 27
- Verteilungskonvergenz, 97
- Vertrauensbereich, 172
- Vertrauensintervall
  - $\alpha$  und  $\beta$ ; einf. lin. Regr., 190
  - EW; einf. lin. Regr., 191
  - für  $\mu$ , 172, 179
  - Prognose; einf. lin. Regr., 191
  - Realisierung, 172
- VF, 158
- Vollerhebung, 124
- Vorhersage, 144
- W-Maß, 4
- W-Raum, 5
- Würfeln, 19, 49
- Wahrscheinlichkeit, 4
  - bedingte, 10
  - total, 11
- Wahrscheinlichkeitsfunktion, 18, 48
  - 2-dim., 48
  - Binomialverteilung, 20
  - Poisson-Verteilung, 20
- Wahrscheinlichkeitsmaß, 4
- Wahrscheinlichkeitsraum, 5
  - diskret, 5
- Walk
  - Bernoulli, 101
  - Random, 101
- Weibull-Verteilung, 118
  - EW, Varianz, 118
- whiskers, 209
- Wilcoxon-Test, 187
- Wurzeltransformation, 187

- Zeitreihe, 148
  - Quartal, 156
- zentrale Momente, 42
- Zentraler Grenzwertsatz, 97, 158
- Zentralwert, 134
- Ziehen
  - mit Zurücklegen, 8
  - ohne Zurücklegen, 9, 110
- Zufallsexperiment, 1
- Zufallsgröße, 17
- Zufallsvariable, 17
  - diskret, n-dim., 47
  - diskrete, 18
  - Maximum, 70
  - Minimum, 70
  - n-dim., 46
  - Produkt, 68
  - Quotient, 68
  - stetig, n-dim., 49
  - stetige, 23
  - Summe, 68, 85, 86, 88, 89
  - Transformation, 32, 63
- Zufallsvektor, 46
- Zustandsraum, 101
- ZV, 17
  - n-dim., 46
- Zweistichproben
  - F-Test, 185, 186
  - t-Test, 183, 186
  - Wilcoxon-Test, 187