

*Elementare  
Wahrscheinlichkeitstheorie*

Stochastik I

Prof. Dr. Uwe Küchler  
Institut für Mathematik  
Humboldt-Universität zu Berlin

Sommersemester 2007

16. Juli 2007

e-mail: [kuechler@mathematik.hu-berlin.de](mailto:kuechler@mathematik.hu-berlin.de)  
[www.mathematik.hu-berlin.de/~kuechler](http://www.mathematik.hu-berlin.de/~kuechler)



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>3</b>
<b>2</b>	<b>Zufällige Versuche und zufällige Ereignisse</b>	<b>9</b>
2.1	Zufällige Versuche . . . . .	9
2.2	Zufällige Ereignisse . . . . .	11
2.3	Verknüpfung von Ereignissen . . . . .	13
2.4	Ereignisse und Mengen . . . . .	15
2.5	Beispiel: Münzenwurf . . . . .	16
<b>3</b>	<b>Wahrscheinlichkeiten und Zufallsgrößen</b>	<b>21</b>
3.1	Axiomensystem und erste Folgerungen . . . . .	22
3.2	Laplace-Experimente . . . . .	32
3.3	Münzenwurf, zum Zweiten . . . . .	34
3.4	Was sagen uns Wahrscheinlichkeiten? . . . . .	40
3.5	Elemente der Kombinatorik* . . . . .	41
3.6	Rein zufällige Wahl eines Punktes aus $[0, 1)$ . . . . .	48
3.7	Zufallsgrößen . . . . .	50
3.8	Verteilungsfunktionen . . . . .	55
3.9	Verteilungsdichten . . . . .	63
<b>4</b>	<b>Diskrete Verteilungen und Zufallsgrößen</b>	<b>75</b>
4.1	Definitionen und Beispiele . . . . .	75
4.2	Die hypergeometrische Verteilung . . . . .	81
4.3	Erwartungswert und Varianz . . . . .	83
4.4	Kovarianz und Korrelation . . . . .	96
4.5	Regressionsgerade . . . . .	99
4.6	Erzeugende Funktionen . . . . .	102
4.7	Mehrstufige zufällige Versuche . . . . .	106

<b>5</b>	<b>Bedingte Wahrscheinlichkeiten und Unabhängigkeit</b>	<b>117</b>
5.1	Definition und Eigenschaften . . . . .	117
5.2	Unabhängigkeit . . . . .	127
<b>6</b>	<b>Bernoullischemata und Irrfahrten</b>	<b>141</b>
6.1	Bernoullischemata . . . . .	141
6.2	Irrfahrten . . . . .	150
<b>7</b>	<b>Erwartungswert und Integral</b>	<b>163</b>
7.1	Definitionen . . . . .	163
7.2	Einige Eigenschaften des Erwartungswertes . . . . .	168
7.3	Dichten eindimensionaler Verteilungen . . . . .	176
7.4	Die Kovarianzmatrix eines zufälligen Vektors . . . . .	186
7.5	Dichten mehrdimensionaler Verteilungen . . . . .	190
<b>8</b>	<b>Produktmaße und Summen unabhängiger Zufallsgrößen</b>	<b>197</b>
8.1	Der Satz von Fubini . . . . .	197
8.2	Faltungsformeln . . . . .	201
<b>9</b>	<b>Charakteristische Funktionen</b>	<b>209</b>
<b>10</b>	<b>Gesetze der großen Zahlen</b>	<b>219</b>
10.1	Einführung . . . . .	219
10.2	Schwache Gesetze der großen Zahlen . . . . .	220
10.3	Starke Gesetze der großen Zahlen . . . . .	224
10.4	Anwendungen des starken Gesetzes der großen Zahlen . . . . .	234
<b>11</b>	<b>Zentrale Grenzwertsätze</b>	<b>237</b>
11.1	Lokaler Grenzwertsatz von Moivre-Laplace . . . . .	237
11.2	Der zentrale Grenzwertsatz von Feller-Lévy . . . . .	241
11.3	Der zentrale Grenzwertsatz von Lindeberg-Feller . . . . .	249
<b>12</b>	<b>Elemente der Mathematischen Statistik</b>	<b>255</b>
12.1	Der Hauptsatz der mathematischen Statistik . . . . .	256
12.2	Statistische Schätzungen . . . . .	261
12.3	Elemente der Testtheorie . . . . .	278



# Kapitel 1

## Einleitung

Diese Vorlesung handelt vom Zufall, genauer, von der Mathematik des Zufalls. Mit dem Zufall haben wir täglich zu tun. Wir sagen, ein Ereignis hänge vom Zufall ab und sprechen von einem *zufälligen Ereignis*, wenn es nicht gewiss ist, ob dieses Ereignis eintritt oder nicht. Eine Größe, deren Wert nicht genau vorhergesagt werden kann, bezeichnen wir als eine *zufällige Größe* oder *Zufallsgröße*.

So wird zum Beispiel die zeitliche Dauer des Weges von zu Hause zur Universität vom Zufall beeinflusst. Der finanzielle Schaden, den ein PKW-Unfall verursacht, hängt vom Zufall ab, Messungen physikalischer Größen werden vom Zufall beeinträchtigt.

Bedeutende Größen zufälliger Natur sind weiterhin die Lebenszeit von Menschen, das Geschlecht Neugeborener, Niederschlagsmengen eines Monats, aber auch Kurse am Aktienmarkt und das Ergebnis von Fußballspielen. Die Reihe der Beispiele läßt sich mühelos fortsetzen. Der Zufall ist überall.

Zufall in reiner Form findet man bei Glücksspielen. Die Augenzahl beim Werfen eines Würfels, die Zahl, bei der die Roulettekugel nach dem Ausrollen liegen bleibt, das Skatblatt, das man nach gutem Mischen und Austeilen erhalten hat, sind rein zufällig. Hier ist der Zufall erwünscht.

Der Einfluss des Zufalls wird allerdings häufig als unangenehm und störend empfunden. Er verursacht *Risiken*, also die Gefahr, dass Schäden entstehen. Wir versuchen ihn deshalb zurückzudrängen, möglichst auszuschließen. Das Ergebnis einer Klausur oder Prüfung wollen wir möglichst nicht vom Zufall in negativer Hinsicht beeinflusst wissen, also bereiten wir uns möglichst gut vor (ein Restrisiko bleibt natürlich: Aufgaben, die wir nicht lösen können, nervli-

che Anspannung usw.) Unfällen, die häufig durch "Verkettung unglücklicher Umstände" entstehen, beugt man durch technische Überwachung, Schulungen usw. vor.

Völlig unterdrücken lässt sich der schädliche Zufall meist nicht oder nur mit extrem großem Aufwand. Wenn man ihn aber schon nicht eliminieren kann, so möchte man die durch ihn verursachten Risiken aber einschätzen, um vorbereitet zu sein. Ein Mittel dafür ist die Stochastik, die Mathematik des Zufalls.

Der Zufall macht zukünftige Ereignisse ungewiss. Er schafft Risiken, aber auch Chancen. Die Stochastik stellt mathematische Verfahren zur Verfügung, mit deren Hilfe man zufällige Erscheinungen, Chancen und Risiken, rechnerisch bewerten kann. Sie gliedert sich in *Wahrscheinlichkeitstheorie* und *Mathematische Statistik*.

Unter der *Wahrscheinlichkeit eines Ereignisses* verstehen wir, kurz gesagt, den *Grad der Gewissheit seines Eintretens*. Sie kann Werte zwischen Null und Eins annehmen. Dabei ordnet man Ereignissen, die praktisch nicht eintreten können, den Wert Null zu, Ereignissen, die mit Sicherheit eintreten, den Wert Eins. Am meisten unbestimmt ist das Eintreten von Ereignissen mit der Wahrscheinlichkeit  $1/2$ .

Mit dem Zufall haben die Menschen seit jeher zu tun. Wetter, Krankheit, Nahrungssuche, Tod waren fundamentale Größen, von denen die Menschen abhängen und die durch Zufall geprägt waren. Es gab vielfältige Versuche, hier dem Zufall auf die Spur zu kommen, seine Herrschaft einzudämmen bzw. sein Wirken aufzuklären. Die ersten mathematischen Ansätze haben ihre Ursprünge in Problemen, die sich bei der Organisation von Versicherungsgesellschaften (z.B. Berechnung von Versicherungsprämien) ergaben und solchen, die bei der Beurteilung von Fragen von Glücksspielen entstanden.

Im Laufe der Zeit hat die Bedeutung des richtigen Umganges mit dem Zufall, insbesondere seine angemessene quantitative Beschreibung, noch zugenommen. In modernen technischen Produkten wie Flugzeugen, Schiffen, Eisenbahnen sind außerordentlich viele Einzelteile vereint und müssen für einen reibungslosen Ablauf zuverlässig funktionieren. Kleine zufällige Störungen können Unfälle mit größten Sach- und Personenschäden verursachen. Ohne eine genaue mathematische Analyse aller auftretenden Risiken wäre der Betrieb solcher technische Produkte nicht mehr denkbar, Unglücke würden wesentlich öfter auftreten. Aber auch für das Verständnis vieler Naturvorgänge kommt man heute

nicht mehr ohne Einbeziehung des Zufalls aus.

Denken wir an die Evolution im Pflanzen- und Tierreich, die man auf zufällige Mutationen mit anschließender natürlicher Auslese zurückführt, oder an die Begriffswelt der Quantenphysik, wo man den Ort eines Teilchens nicht mehr exakt bestimmen kann und als Wahrscheinlichkeitsverteilung modelliert.

Zufall, Chance und Risiko sind Begriffe, die im Gegensatz zu vielen physikalischen Größen wie Temperatur, Länge, Gewicht nicht sinnlich wahrnehmbar sind. Für viele Menschen haben sie heute noch etwas Unheimliches, Mystisches und auch Reizvolles an sich. Die Abhängigkeit vom Zufall weckt Hoffnung und Angst. Die Menschen spielen Lotto bzw. tragen ein Maskottchen. Das muss nicht schlimm sein. Problematisch ist es erst, wenn man Risiken und Chancen, auch für das persönliche Leben, nicht richtig abschätzt und damit eventuell Gefahren für Hab und Gut oder gar die Gesundheit und das Leben eingeht.

Was ist nun eigentlich Zufall? Diese alte und schwierige Frage gehört zur Philosophie und soll hier nicht behandelt werden.

Die Vorlesung soll die Hörer mit einigen grundlegenden Begriffen der Stochastik bekannt machen, typische Denk- und Schlussweisen vorstellen und einige wichtige Gesetzmäßigkeiten des Zufalls nahe bringen.

Das Wissen um den Zufall, die damit verbundenen Begriffe und Methoden des Umganges mit dem Zufall ist heute Allgemeingut aller Wissenschaftszweige und gehört auch um notwendigen Alltagswissen. Vielen Menschen ist das genauere Wissen um den Zufall jedoch noch fremd.

Das vorliegende Skript entstand auf der Grundlage von Vorlesungen über Elemente der Stochastik, die ich in Abständen von Jahren mehrfach für Studierende der Mathematik gehalten habe. Dennoch ist die schriftliche Ausarbeitung eines Skriptes immer auch eine aufwändige Arbeit. Ich danke unserer Institutssekretärin Frau S. Bergmann für die umfangreiche Arbeit am Skript und die unendliche Geduld gegenüber meinen zahlreichen Änderungswünschen.

Mein Dank geht weiterhin an unsere wissenschaftlichen Mitarbeiter und Studierenden, die durch kritische und konstruktive Hinweise zur Vorjahresversion dieses Skriptes erheblich zu einer gründlichen Überarbeitung beigetragen haben. Dazu gehören u. a. Dr. Markus Riedle und die Diplommathematiker Thomas Knispel, Katja Krol, Hagen Gilsing sowie die Studierenden Andrea

Konieczny und Friedrich Bolz. Herr H. Gilsing hat für das vorliegende Skript die Graphiken erstellt. Auch dafür danke ich ihm.

Dieses Skript enthält sicher noch eine Reihe von Fehlern, insbesondere Druckfehler, aber auch andere Unzulänglichkeiten. Dafür bin ich allein verantwortlich. Ich bitte die Hörer der Vorlesung und andere Leser um Nachsicht und bin sehr dankbar für kritische und helfende Hinweise.

Vieles wird in der Vorlesung vorkommen, zum Beispiel weitere Bilder, Graphiken, interessante Beispiele und Anwendungen, das im Skript nicht enthalten ist. Andererseits werden in der Vorlesung manche Ausführungen des Skriptes nur gestreift werden, zum Beispiel gewisse Elemente der Maß- und Integrationstheorie. Sie wurden teilweise nur aufgenommen, um eine gewisse Geschlossenheit der Darstellung und eine Festlegung der Terminologie zu erreichen.

An der Lehrveranstaltung Stochastik I werden mehrere erfahrene wissenschaftliche Mitarbeiter und Studierende höherer Semester als Übungsleiter und Korrektoren der schriftlich anzufertigenden Übungen beteiligt sein. Wir sind gespannt, freuen uns auf die Arbeit mit den Studierenden im Sommersemester 2007 und wünschen uns allen viel Erfolg!

Uwe Küchler

Berlin, 11. April 2007

# Literaturverzeichnis

- [1] Bauer, H.: Maß- und Integrationstheorie, de Gruyter, 1990
- [2] Bauer, H.: Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie. 4. Auflage, de Gruyter, Berlin, 1991
- [3] Dehling, H., Haupt, B.: Einführung in die Wahrscheinlichkeitstheorie und Statistik, Springer, 2004
- [4] Elstrodt, J.: Maß- und Integrationstheorie, Springer, 1999
- [5] Henze, N.: Stochastik für Einsteiger, 6. Auflage, 2006
- [6] Hesse, Ch.: Angewandte Wahrscheinlichkeitstheorie, Vieweg, 2003
- [7] Krenzel, U.: Einführung in die Wahrscheinlichkeitstheorie und Statistik, Vieweg, 6. Auflage, 2002
- [8] Jacod, J. and Protter, Ph.: Probability Essentials, Springer 2000
- [9] Müller, P.H. u.a.: Lexikon der Stochastik, Akademie-Verlag, Berlin, 5. Auflage, 1991
- [10] Müller, P.H.; Neumann, R. und Storm, R.: Tafeln zur Mathematischen Statistik, Fachbuchverlag, Leipzig 1973
- [11] Pfanzagl, J.: Elementare Wahrscheinlichkeitsrechnung, de Gruyter Berlin, 2. Auflage, 1991
- [12] Pitmann, J.: Probability, Springer 1993
- [13] Renyi, A.: Briefe über die Wahrscheinlichkeit, Dt. Verlag der Wiss., Berlin, 1969

- [14] Siraeu, A.N.: Wahrscheinlichkeit, Dt. Verlag der Wiss., Berlin, 1988
- [15] Winkler, M.: Vorlesungen zur Mathematischen Statistik, Teubner, Leipzig, 1983

# Kapitel 2

## Zufällige Versuche und zufällige Ereignisse

In diesem Kapitel führen wir zunächst anschaulich die grundlegenden Begriffe des zufälligen Versuchs und des zufälligen Ereignisses ein und stellen danach eine Verbindung zur Mengenlehre her. Damit wird die Grundlage einer mathematischen Theorie des Zufalls, der Wahrscheinlichkeitstheorie gelegt. Wir erfahren, dass die mit einem zufälligen Versuch verbundenen zufälligen Ereignisse eine  $\sigma$ -Algebra bilden.

### 2.1 Zufällige Versuche

**Definition 2.1** *Unter einem zufälligen Versuch versteht man einen Versuch (im weitesten Sinne des Wortes), dessen Ausgang unter bestimmten wesentlichen und fixierten Bedingungen im Rahmen bestimmter Möglichkeiten ungewiss ist.*

Die einzelnen möglichen Versuchsausgänge (-ergebnisse) werden häufig mit  $\omega$ , die Menge aller möglichen Versuchsausgänge des betrachteten zufälligen Versuches mit  $\Omega$  bezeichnet.

#### Beispiel 2.2

- a) Werfen einer Münze: Die möglichen Versuchsausgänge  $\omega$  sind gleich  $Z$  und  $W$ , d. h. Zahl oder Wappen. Folglich ist  $\Omega = \{Z, W\}$ .

- b) Werfen zweier unterschiedlicher Münzen:  $\omega = (\omega_1, \omega_2)$   
 $\omega_i =$  Wurfresultat der  $i$ -ten Münze,  
 $\omega_i \in \{Z, W\}, \quad i = 1, 2$   
 $\Omega = \{(Z, Z), (Z, W), (W, Z), (W, W)\}$
- c)  $n$ -maliges Werfen einer Münze,  $\omega = (\omega_1, \dots, \omega_n), \omega_i \in \{Z, W\}, i = 1, 2, \dots, n$
- d) Werfen eines Würfels:  
 $\Omega = \{1, 2, \dots, 6\}$
- e) Werfen zweier unterscheidbarer Würfel:  $\omega = (i, j), \quad i$  Augenzahl des ersten,  $j$  Augenzahl des zweiten Würfels,  
 $\Omega = \{(i, j) : i, j \in \{1, 2, \dots, 6\}\}$
- f) Sonntagsziehung im Lotto "6 aus 49" (ohne Zusatzzahl):  
 $\Omega = \{\omega = \{i_1, \dots, i_6\} : i_1, \dots, i_6 \in \{1, \dots, 49\}, \quad i_j \neq i_k (j \neq k)\}$
- g) Geburt eines Kindes, es werden registriert Gewicht  $\eta$  in  $g$ , Größe  $\xi$  in  $cm$ , Geschlecht  $\tau$ :  
 $\omega = (\eta, \xi, \tau), \quad \Omega = (0, \infty) \times (0, \infty) \times \{m, w\}$
- h) Niederschlagsmenge  $\omega$  pro Quadratmeter in  $mm$  am 30. 10. 2007 auf dem Alexanderplatz:  $\Omega = [0, \infty)$
- i) Schadenhöhe  $\omega$  bei einem PKW-Unfall, die der Versicherer in Euro zu zahlen hat:  
 $\Omega = [0, \infty)$
- j) Anzahl  $\omega$  aller polizeilich gemeldeten Kfz-Unfälle an einem bestimmten Tag auf der Rudower Chaussee in Adlershof:  $\Omega = \{0, 1, \dots, n, \dots\}$

Diskussion:

1.  $\Omega$  ist nicht eindeutig festgelegt. Einzige Bedingung: nach Ausführung des Versuches muss genau ein  $\omega$  aus  $\Omega$  als Versuchsergebnis feststehen. Insbesondere ist es nicht notwendig, dass alle  $\omega \in \Omega$  auch tatsächlich auftreten können,  $\Omega$  kann also größer gewählt werden als unbedingt notwendig (vgl. Beispiele g) - j)).  
 Erweiterungen von  $\Omega$  sind aus mathematischen Gründen oder wegen

Übersichtlichkeit häufig vorteilhaft.

2. Vor dem Versuch ist der tatsächlich auftretende Ausgang  $\omega \in \Omega$  des zufälligen Versuches ungewiss. Sicher ist nur, dass genau eines der  $\omega$  aus  $\Omega$  auftreten wird. Nach dem Versuch liegt der aufgetretene Ausgang  $\omega$  fest. Die Ungewissheit ist verschwunden. Der Versuch wurde *realisiert, verwirklicht*. Das nach dem Versuch erschienene  $\omega$ , eine Zahl in a), d), h), i), j) oder allgemeinere Ergebnisse in den anderen Beispielen, heißt *Realisierung dieses Versuches*. Bei erneuter Ausführung des Versuches tritt i.A. ein anderer Ausgang in Erscheinung, es erscheint eine andere Realisierung.

Wird der Versuch mehrmals durchgeführt, ergibt sich eine Folge von Realisierungen, eine sogenannte  $(\omega, \eta, \dots, \kappa)$  *Stichprobe*, ein *Datensatz*.

## 2.2 Zufällige Ereignisse

**Definition 2.3** Ein zufälliges Ereignis (oder kurz Ereignis) ist ein Ereignis, das (im Rahmen eines bestimmten zufälligen Versuches und in Abhängigkeit vom Versuchsausgang) eintreten kann, aber nicht eintreten muss.

Zufällige Ereignisse beschreibt man häufig verbal durch eine logische Aussage und symbolisch durch große Buchstaben  $A, B, C, \dots$ , meist vom Anfang des Alphabetes.

Betrachten wir einige Ereignisse im Zusammenhang mit Beispielen aus Abschnitt 2.1.

### Beispiel 2.4 (Fortsetzung von 1.2)

- a)  $A$  : = "Es erscheint das Wappen"
- e)  $A$  : = "Die Summe der Augenzahlen ist gerade"
- f)  $A$  : = "Bei der Ziehung erscheint mindestens ein Zahlenzwilling"  
 $B$  : = "Der abgegebene Tippschein enthält 3 Richtige"

h)  $A :=$  "Es regnet mehr als 10 mm"

i)  $A :=$  "Der Schaden ist größer als 100 000 EUR"

**Definition 2.5** *Man sagt, das Ereignis  $A$  tritt (bei Versuchsdurchführung mit dem Versuchsausgang  $\omega$ ) ein, wenn die zugehörige logische Aussage bei diesem  $\omega$  wahr ist, es tritt nicht ein, wenn sie bei diesem  $\omega$  falsch ist. Wenn ein Ereignis  $A$  beim Versuchsausgang  $\omega$  eintritt, so sagt man auch, dieses  $\omega$  führt zum Eintreten von  $A$ .*

**Definition 2.6** *Das Ereignis  $A$  zieht das Ereignis  $B$  nach sich oder ist ein Teil von  $B$ , falls aus dem Eintreten von  $A$  folgt, dass auch  $B$  eintritt. Symbolisch:  $A \subseteq B$ .*

Im Beispiel e) aus 1.2 gilt mit

$C :=$  "Die Summe der Augenzahlen ist fünf" und

$D :=$  "Es erscheint eines der Paare (1,4), (2,3), (3,2), (4,1)" die Beziehung  $C \subseteq D$ .

**Definition 2.7** *Zwei Ereignisse  $A$  und  $B$  heißen einander gleich (symbolisch:  $A = B$ ), wenn das Eintreten des einen Ereignisses das Eintreten des anderen nach sich zieht, d. h. falls  $A \subseteq B$  und  $B \subseteq A$  gelten. Einander gleiche Ereignisse treten entweder beide ein oder beide nicht ein.*

Im Beispiel d) aus 1.2 gilt mit

$C :=$  "Es erscheint eine ungerade Zahl"

$D :=$  "Die gewürfelte Augenzahl ist nicht 2, 4 oder 6"  
die Beziehung  $C = D$ .

**Definition 2.8** *Ein zufälliges Ereignis  $A$  heißt mit einem gegebenen zufälligen Versuch verbunden, falls man für jeden möglichen Versuchsausgang  $\omega \in \Omega$  entscheiden kann, ob er zum Eintreten von  $A$  führt oder nicht.*

Das Ereignis  $A$  ist also mit dem zufälligen Versuch  $\Omega$  genau dann verbunden, wenn man nach Ausführung des Versuches entscheiden kann, ob  $A$  eingetreten ist oder nicht.

Das Ereignis  $B :=$  "Der abgegebene Tippschein enthält drei Richtige" ist mit dem zufälligen Versuch einer Sonntagsziehung im Zahlenlotto verbunden. Das Ereignis "Morgen scheint die Sonne mindestens zwei Stunden" ist nicht mit dem zufälligen Versuch des Werfens eines Würfels verbunden.

**Definition 2.9** *Im Rahmen eines zufälligen Versuches heißt ein Ereignis  $S$  ein sicheres Ereignis, falls es bei jedem Versuchsausgang eintritt. Ein Ereignis  $U$  nennt man ein unmögliches Ereignis, wenn es bei keinem Versuchsausgang eintritt.*

Offenbar gelten für jedes mit dem Versuch verbundene Ereignis  $A$  die Relationen  $U \subseteq A \subseteq S$ .

Im Beispiel d) aus 1.2 ist das Ereignis "Es erscheint eine der Zahlen "1, 2, ..., 6" ein sicheres Ereignis, und das Ereignis "Es erscheint eine Zahl, die größer als 10 ist" ein unmögliches Ereignis.

Wir erinnern daran, dass wir mit jedem zufälligen Versuch eine Menge  $\Omega$  festlegen, die alle möglichen Ausgänge des Versuches enthält.

Die Menge aller Ereignisse  $A$ , die mit einem gegebenen zufälligen Versuch verbunden sind, wird mit  $\mathfrak{A}$  bezeichnet.

Das Paar  $(\Omega, \mathfrak{A})$  ist für uns das vorläufige Modell eines zufälligen Versuches.

## 2.3 Verknüpfung von Ereignissen

Es sei  $(\Omega, \mathfrak{A})$  ein zufälliger Versuch, d. h.  $\Omega$  enthalte die Menge aller möglichen Versuchsausgänge  $\omega$  und  $\mathfrak{A}$  sei das System der mit dem Versuch verbundenen Ereignisse. Dabei seien  $S$  das sichere und  $U$  das unmögliche Ereignis.

Aus gegebenen Ereignissen  $A, B \in \mathfrak{A}$  lassen sich weitere Ereignisse bilden, die ebenfalls mit dem zufalligen Versuch verbunden sind:

**Definition 2.10** Das Ereignis  $A \cup B$  tritt ein, falls  $A$  eintritt oder  $B$  eintritt (oder beide).  $A \cup B$  nennt man die Vereinigung von  $A$  und  $B$ . Es gilt  $A \cup S = S$ ,  $A \cup U = A$ .

Das Ereignis  $A \cap B$  tritt ein, falls  $A$  und  $B$  beide eintreten.  $A \cap B$  nennt man den Durchschnitt von  $A$  und  $B$ . Es gilt  $A \cap S = A$ ,  $A \cap U = U$ .

$\bar{A}$  tritt genau dann ein, falls  $A$  nicht eintritt.  
 $\bar{A}$  heit das zu  $A$  komplementare Ereignis. Es gilt  $\overline{\bar{U}} = S$  und  $\bar{S} = U$ .

$A \setminus B$  tritt genau dann ein, wenn  $A$  eintritt und  $B$  nicht eintritt.  
 $A \setminus B$  heit die Differenz von  $A$  und  $B$ . Es gilt  $A \setminus B = A \cap \bar{B}$ .

Das Ereignis  $A \Delta B := (A \setminus B) \cup (B \setminus A)$  heit *symmetrische Differenz* von  $A$  und  $B$ . Es tritt genau dann ein, wenn entweder  $A$  oder  $B$  eintritt.

Wenn  $A \cap B = U$  gilt, so heien  $A$  und  $B$  *disjunkt* oder *unvereinbar*.  
 Es gilt stets:  $A \cap \bar{A} = U$ ,  $A \cup \bar{A} = S$ .

Sind  $A_k, k = 1, \dots, m$ , Ereignisse aus  $\mathfrak{A}$ , so bezeichne  $\bigcup_{k=1}^m A_k$  das Ereignis, das genau dann eintritt, wenn mindestens eines der Ereignisse  $A_k$  eintritt, und  $\bigcap_{k=1}^m A_k$  das Ereignis, das genau dann eintritt, wenn alle  $A_k$  eintreten.

Analog definiert man zu jeder Folge  $A_k, k \geq 1$ , aus  $\mathfrak{A}$  die Ereignisse  $\bigcup_{k=1}^{\infty} A_k$  und  $\bigcap_{k=1}^{\infty} A_k$ .

Folgerung: Die Menge  $\mathfrak{A}$  aller mit einem zufalligen Versuch verbundenen Ereignisse hat also die Eigenschaften:

- 1)  $U, S \in \mathfrak{A}$ ,

- 2) Für jedes  $A \in \mathfrak{A}$  ist auch  $\bar{A} \in \mathfrak{A}$ ,
- 3) Für jedes  $n \geq 2$  und alle  $A_1, A_2, \dots, A_n \in \mathfrak{A}$  gilt  $\bigcup_{k=1}^n A_k \in \mathfrak{A}$ .

Auf Grund dieser Eigenschaften und der Definition bzw. der Eigenschaften von  $U$  und  $S$  heißt  $\mathfrak{A}$  eine *Algebra* (bez. der Operationen  $\bigcup, \bar{\phantom{x}}$ ) mit Nullelement  $U$  und Einselement  $S$ .

Da außerdem

- 4) Für alle  $A_1, A_2, \dots, A_n, \dots \in \mathfrak{A}$  gilt  $\bigcup_{k=1}^{\infty} A_k \in \mathfrak{A}$

erfüllt ist, nennt man  $\mathfrak{A}$  auch eine  $\sigma$ -*Algebra*.

Literatur: Bauer (1991), Krengel (2002)

## 2.4 Ereignisse und Mengen

Es sei  $(\Omega, \mathfrak{A})$  ein zufälliger Versuch im Sinne der Schlußbemerkungen von Abschnitt 1.2.

Jedes mit diesem Versuch verbundene Ereignis  $A$ , d.h. jedes  $A$  aus  $\mathfrak{A}$ , wird durch eine Teilmenge  $A'$  von  $\Omega$  charakterisiert:

$$A \longleftrightarrow A' = \{\omega \in \Omega : \text{erscheint der Versuchsausgang } \omega, \text{ so tritt } A \text{ ein} \}$$

Wenn  $A$  eintritt, so ist ein Versuchsausgang  $\omega$  eingetreten, der zu  $A'$  gehört. Wenn ein  $\omega \in A'$  als Versuchsausgang auftritt, so tritt nach Definition von  $A'$  auch  $A$  ein.

$A$  und "Es erscheint ein  $\omega \in A'$  als Versuchsergebnis" sind somit im Sinne von Definition 2.7 einander gleiche Ereignisse, d. h. entweder treten sie beide ein oder beide nicht.

Insofern charakterisiert die Teilmenge  $A'$  von  $\Omega$  das Ereignis  $A$ . Identifiziert man  $A$  mit seiner zugehörigen Menge  $A'$ , so können wir feststellen:

Feststellung:

*Zufällige Ereignisse, die mit einem zufälligen Versuch  $(\Omega, \mathfrak{A})$  verbunden sind,*

*kann man identifizieren mit Teilmengen von  $\Omega$ , m. a. W., die  $\sigma$ -Algebra  $\mathfrak{A}$  ist ein System von Teilmengen von  $\Omega$ . Bei dieser Entsprechung wird das sichere Ereignis mit  $\Omega$ , das unmögliche Ereignis mit der leeren Menge  $\emptyset$  identifiziert. Die Korrespondenz  $A \longleftrightarrow A'$  ist bezüglich der Operationen Vereinigungs-, Durchschnitts-, Differenz- und Komplementbildung für Ereignisse bzw. für Mengen ein Isomorphismus.*

Das Paar  $(\Omega, \mathfrak{A})$  dient nun vorläufig als mathematisches Modell eines zufälligen Versuches, dessen mögliche Versuchsausgänge zu  $\Omega$  gehören, wobei die mit dem Versuch zusammenhängenden Ereignisse, also die Menge  $\mathfrak{A}$ , eine  $\sigma$ -Algebra von Teilmengen von  $\Omega$  ist.

In den Beispielen a) bis f) und j) ist  $\mathfrak{A} = \mathfrak{P}(\Omega)$  (Potenzmenge von  $\Omega$ ) in den Beispielen g) bis i) wählt man i. a.  $\mathfrak{A} \subsetneq \mathfrak{P}(\Omega)$ , die Begründung werden wir kennen lernen.

## 2.5 Beispiel: Münzenwurf

Wir formulieren zum Ende dieses Abschnittes noch ein mathematisches Modell, auf das wir später mehrfach zurückkommen werden. Eine Münze werde  $n$ -mal geworfen. Die möglichen Ausgänge  $\omega$  dieser Wurfserie sind die  $n$ -Tupel  $\omega = (x_1, x_2, \dots, x_n)$  mit  $x_k \in \{-1, 1\}$ ,  $k = 1, 2, \dots, n$ , wobei  $x_k = +1 (= -1)$  gesetzt wird, falls beim  $k$ -ten Wurf die Zahl (bzw. das Wappen) oben liegt. Die Menge  $\Omega$  aller möglichen Ausgänge  $\omega$  der Wurfserie besteht aus  $2^n$  Elementen.

Wir setzen  $\mathfrak{A} = \mathfrak{P}(\Omega)$ , denn für jede Teilmenge  $A'$  von  $\Omega$  ist  $A :=$  "Der zufällige Versuch endet mit einem  $\omega \in A'$ " ein im Sinne von Definition 2.8 mit dem  $n$ -maligen Werfen der Münze verbundenes Ereignis.

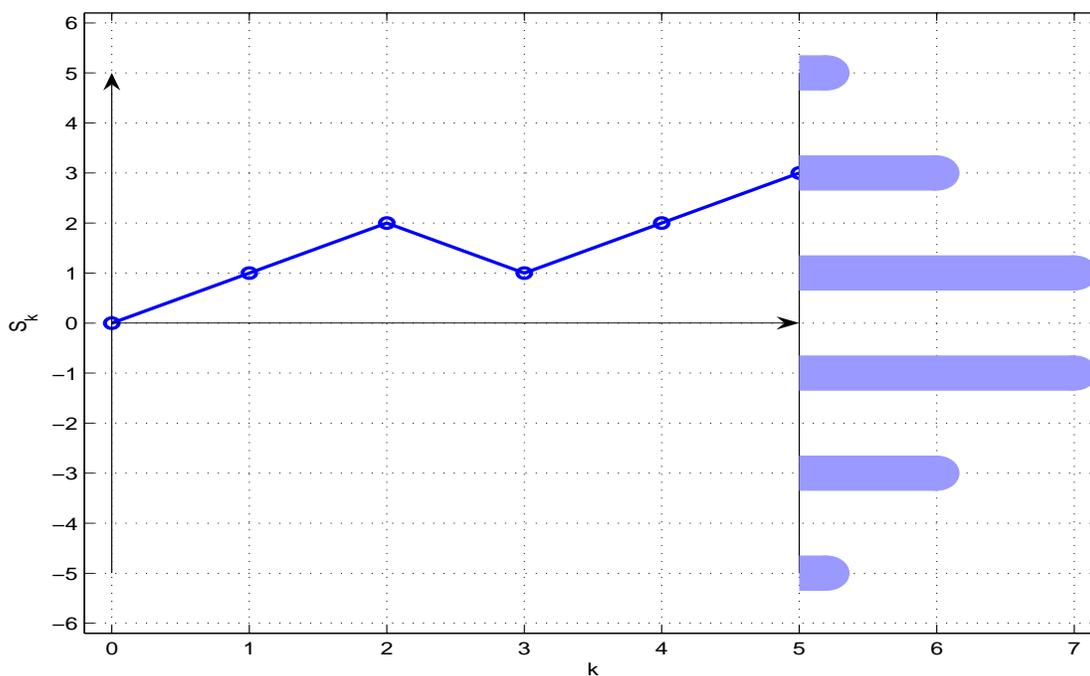


Abbildung 2.1: Ein Pfad der Länge fünf

Für jedes  $\omega = (x_1, x_2, \dots, x_n) \in \Omega$  definieren wir

$$s_0 = 0, s_k = \sum_{l=1}^k x_l, k = 1, 2, \dots, n$$

und nennen die Folge

$$s = ((k, s_k), k = 0, 1, \dots, n)$$

den zu  $\omega$  gehörenden *Pfad*. Wir veranschaulichen jeden Pfad  $s$  durch die Punkte  $(k, s_k)$  in der Ebene und verbinden die benachbarten Punkte  $(k, s_k)$  und  $(k+1, s_{k+1})$  linear. Diese Pfade  $s$  haben die Eigenschaft

$$s_0 = 0 \text{ und } |s_k - s_{k-1}| = 1, k = 1, 2, \dots, n$$

und entsprechen den Versuchsausgängen  $\omega$  eineindeutig.

Vereinbart man ein Spiel, in dem der Spieler A von einer Bank den Betrag +1 erhält, falls im Ergebnis eines Münzwurfes die Zahl erscheint, und er den

Betrag 1 zu zahlen hat, wenn Wappen oben liegt, so ist sein Gewinn nach  $k$  Würfeln gleich  $s_k$ .

Das Ereignis

$C$ : = "Spieler A hat zum Schluss einen positiven Betrag gewonnen"

tritt genau dann ein, wenn der Versuchsausgang  $\omega = (x_1, x_2, \dots, x_n)$  zur Menge

$$C = \{\omega \in \Omega \mid s_n = \sum_{l=1}^n x_l > 0\}$$

gehört. Zu  $C$  gehören also alle  $s(\omega)$  mit Pfaden, die nach  $n$  Schritten im Positiven enden.

Das Ereignis

$D$ : = "Das Guthaben des Spielers A sinkt im Verlauf des Spieles niemals unter Null"

tritt genau dann ein, wenn ein Versuchsausgang  $\omega \in \Omega$  mit

$$\min_{k=1,2,\dots,n} s_k \geq 0$$

auftritt, d. h., wenn der zugehörige Pfad niemals die  $-1$  berührt.

Zur Vorbereitung allgemeinerer Definitionen führen wir folgende Funktionen  $X_k$  und  $S_m$  auf  $\Omega$  ein:

$$X_k(\omega) := x_k, k = 1, 2, \dots, n,$$

$$S_m(\omega) := \sum_{k=1}^m X_k(\omega) = \sum_{k=1}^m x_k, m = 1, 2, \dots, n.$$

Die "Zufallsgröße"  $X_k$  gibt das Ergebnis des  $k$ -ten Wurfes an, die "Zufallsgröße"  $S_m(\omega)$  ist der Gewinn des Spielers A nach  $m$  Würfeln,  $m = 1, 2, \dots, n$ .

Für die oben eingeführten Ereignisse  $C$  und  $D$ , gilt dann

$C = \{\omega \in \Omega \mid S_n(\omega) > 0\}$ , kurz geschrieben  $C = \{S_n > 0\}$ , und

$D = \{\omega \in \Omega \mid \min_{k=1, \dots, n} S_k(\omega) \geq 0\}$  oder kurz  $D = \{\min_{k=1, \dots, n} S_k \geq 0\}$ .

Kontrollfragen:

Man gebe im Modell des  $n$ -maligen Münzenwurfes die zum Ereignis  $B :=$  "Der Spieler gewinnt nach  $n$  Würfeln mindestens einen Betrag der Höhe  $+1$ " gehörende Teilmenge von  $\Omega$  an.

Welche Zufallsgröße  $Z$ , d. h. welche Funktionen  $Z$  auf  $\Omega$  gibt an, bei welchem Wurf der Spieler A zum ersten Mal eine Zahl wirft?



# Kapitel 3

## Wahrscheinlichkeiten und Zufallsgrößen

Zufällige Ereignisse unterscheiden sich im Grad der Gewissheit ihres Eintretens, d. h., in der *Wahrscheinlichkeit* ihres Eintretens.

Es ist eine Erfahrungssache, dass sich die relative Häufigkeit, mit der ein Ereignis in einer langen Reihe von Versuchen, die immer wieder neu unter im Wesentlichen gleichartigen Bedingungen ausgeführt werden, um einen festen Wert stabilisiert. Diesen Wert könnte man als Grad der Gewissheit des Eintretens des Ereignisses in einem einzelnen Versuch ansehen. Ausgehend von dieser Vorstellung formulieren wir einige plausible Eigenschaften von Wahrscheinlichkeiten, die sich dann auch wieder finden im *Axiomensystem der Wahrscheinlichkeitstheorie*.

Zu den mathematisch übersichtlichsten zufälligen Versuchen gehören die *Laplace-Versuche*. Sie besitzen nur endlich viele und dabei gleichwahrscheinliche Ausgänge. Die Bestimmung von Wahrscheinlichkeiten der mit ihnen zusammenhängenden Ereignisse läuft auf das Abzählen gewisser Fälle, häufig unter Verwendung *kombinatorischer Formeln*, hinaus.

Der Begriff der *Zufallsgröße* gehört ebenfalls zum Grundbestand der Wahrscheinlichkeitstheorie. Zufallsgrößen vermitteln stets Teilaspekte eines zufälligen Versuchs und fungieren als beobachtbare (bzw. interessierende) Größen, wenn der Ausgang des Versuches selbst nicht beobachtbar ist (bzw. nicht von Interesse ist). Von Wichtigkeit sind die von ihnen induzierten *Wahrscheinlich-*

keitsverteilungen.

### 3.1 Axiomensystem und erste Folgerungen

Wir betrachten einen zufälligen Versuch und nehmen an,  $\Omega$  sei die Menge seiner möglichen Versuchsausgänge  $\omega$  und  $\mathfrak{A}$  die  $\sigma$ -Algebra der mit diesem Versuch verbundenen Ereignisse, d. h. gemäß Abschnitt 2.4 eine  $\sigma$ -Algebra von Teilmengen von  $\Omega$ .

Es sei  $A$  irgend ein Ereignis aus  $\mathfrak{A}$ . Wir wissen, dass  $A$  bei der Versuchsausführung eintreten kann, aber nicht eintreten muss. Sein Eintreten ist nicht gewiss. Unterschiedliche Ereignisse können sich allerdings im *Grad der Gewissheit* ihres Eintretens unterscheiden.

#### Beispiel 3.1 (Werfen einer Zündholzschachtel)

Mögliche Versuchsausgänge sind die der drei Seiten, auf denen die Schachtel zu liegen kommen kann: Große Seite (Ober- bzw. Unterseite) / Mittlere Seite (Seiten mit Reibflächen) / Kleine Seite (Stirn- bzw. Hinterseite): Als  $\Omega$  wählen wir die Menge  $\Omega = \{G, M, K\}$ .

Wir bemessen den Grad der Gewissheit des Eintreffens jeder der möglichen Fälle aus der Erfahrung oft wiederholter Versuche. Dieser Grad wird umso höher eingeschätzt, je häufiger bei längerer Versuchsreihe die Schachtel auf der entsprechenden Seite zu liegen kommt.

Betrachten wir die Situation von Beispiel 3.1 etwas allgemeiner. Es sei  $A$  ein Ereignis, das mit einem zufälligen Versuch verbunden ist. Der Versuch werde  $n$  mal durchgeführt, jedes Mal unter im Wesentlichen gleichartigen Bedingungen und unabhängig voneinander.

In  $n(A)$  Fällen trete  $A$  ein. Dann zeigt die *relative Häufigkeit* des Eintretens von  $A$  in  $n$  Versuchen, nämlich  $\frac{n(A)}{n}$ , mit wachsendem  $n$  eine bemerkenswerte Stabilität:  $\frac{n(A)}{n}$  verändert sich immer weniger, sie scheint gegen einen Grenzwert zu konvergieren, wenn  $n$  unbegrenzt wächst.

Wir nennen diese Erscheinung das *empirische Gesetz der großen Zahlen*.

Das folgende Beispiel verdeutlicht die Stabilisierung der relativen Häufigkeiten.

### Beispiel 3.2 (Werfen eines Kronenverschlusses)

Relative Häufigkeit dafür, dass die offene Seite nach oben zeigt:

Zahl der Versuche

100	200	300	400	500	600	700
0,7300	0,7750	0,7767	0,7750	0,7800	0,7900	0,7943
800	900	1000				
0,8012	0,7967	0,7910				

(Nach Nawrotzki, K., Lehrmaterial zur Ausbildung von Diplomlehrern Mathematik, Jena 1984)

Mit dem nächsten Beispiel wird deutlich, dass das empirische Gesetz der großen Zahlen häufig auch unbewusst angewandt wird.

### Beispiel 3.3 (Skatspiel)

Die relativen Häufigkeiten bestimmter Konstellationen prägen sich beim Spieler ein. Zwei Buben im Skat sind z. B. relativ selten. Daraus wird geschlossen, dass auch im nächsten Spiel nur mit geringer Chance zwei Buben im Skat zu finden sein werden. Es ergibt sich eine Verhaltensgrundlage: "Auf den Skat reizt man nicht".

Aus dem genannten Gesetz der großen Zahlen leitet man die Überzeugung ab, dass es zu jedem zufälligen Ereignis  $A$  eine Zahl  $P(A)$  gibt, die *Wahrscheinlichkeit* von  $A$ , die den Grad der Gewissheit des Eintretens von  $A$  (in einem einzelnen Versuch) ausdrückt.

Für lange Versuchsreihen sollte das eine Zahl sein, um die sich  $\frac{n(A)}{n}$  stabilisiert:

$$\frac{n(A)}{n} \approx P(A).$$

Daraus ergeben sich plausible Eigenschaften für  $P(A)$ :

$$0 \leq P(A) \leq 1,$$

$$P(\Omega) = 1, \quad P(\emptyset) = 0,$$

$$P(A \cup B) = P(A) + P(B), \text{ falls } A \cap B = \emptyset.$$

Aus diesen Vorstellungen hat sich ein Axiomensystem entwickelt, das 1933 A.N. Kolmogorov in einer berühmten Arbeit eingeführt hat. (Kolmogorov, Grundlagen der Wahrscheinlichkeitstheorie, Springer, Berlin 1933).

Dieses Axiomensystem ist heute die Grundlage der Wahrscheinlichkeitstheorie und Mathematischen Statistik und lautet wie folgt.

### Axiomensystem der Wahrscheinlichkeitstheorie

Wir stellen uns wieder  $(\Omega, \mathfrak{A})$  als einen zufälligen Versuch vor, die Elemente  $\omega$  von  $\Omega$  bilden die möglichen Versuchsausgänge,  $\mathfrak{A}$  sei die  $\sigma$ -Algebra der mit dem Versuch verbundenen Ereignisse, also eine  $\sigma$ -Algebra von Teilmengen von  $\Omega$  (siehe Abschnitt 2.4).

Als *Wahrscheinlichkeitsverteilung*  $P(\cdot)$  auf der  $\sigma$ -Algebra  $\mathfrak{A}$  von Teilmengen einer nichtleeren Menge  $\Omega$  bezeichnet man jede Abbildung  $P$  von  $\mathfrak{A}$  in  $[0, 1]$  mit

A1.  $P(\Omega) = 1$  und  $P(\emptyset) = 0$ ,

A2. Für jedes  $n \geq 2$  und jede Folge  $(A_k, k = 1, \dots, n)$  aus  $\mathfrak{A}$  mit

$$A_k \cap A_l = \emptyset, k \neq l \text{ (paarweise Unvereinbarkeit) gilt}$$

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k)$$

(Endliche Additivität der Wahrscheinlichkeitsverteilung  $P$ )

A2.' Für jede abzählbar unendliche Folge  $(A_k, k \geq 1)$  aus  $\mathfrak{A}$  mit

$A_k \cap A_l = \emptyset$  ,  $k \neq l$  (paarweise Unvereinbarkeit) gilt

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$$

( $\sigma$ -Additivität der Wahrscheinlichkeitsverteilung  $P$ )

$(\Omega, \mathfrak{A}, P)$  ist mit dieser Definition ein (normierter) Maßraum,  $P$  heißt ein *Wahrscheinlichkeitsmaß* auf  $\mathfrak{A}$ . Statt *Wahrscheinlichkeitsverteilung*  $P(\cdot)$  auf  $\mathfrak{A}$ , sprechen wir einfach auch von einer *Verteilung*  $P(\cdot)$  auf  $\mathfrak{A}$ .

**Definition 3.4** Sind  $\Omega$  eine nichtleere Menge,  $\mathfrak{A}$  eine  $\sigma$ -Algebra von Teilmengen von  $\Omega$  und  $P$  eine Abbildung von  $\mathfrak{A}$  in  $[0, 1]$  mit den Eigenschaften A1., A2. und A2'., so heißt das Tripel  $(\Omega, \mathfrak{A}, P)$  ein *Wahrscheinlichkeitsraum*.

### Bemerkung 3.5

Jeder Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  ist das mathematische Modell eines zufälligen Versuches.  $\Omega$  enthält dabei die Menge der möglichen Versuchsergebnisse,  $\mathfrak{A}$  entspricht der Menge der mit dem Versuch verbundenen Ereignisse,  $P$  ist die Wahrscheinlichkeitsverteilung des Versuches.  $P$  legt fest, mit welcher Wahrscheinlichkeit  $P(A)$  jedes mit dem Versuch verbundene Ereignis  $A \in \mathfrak{A}$  bei der Versuchsdurchführung eintritt.

**Folgerungen 3.6**

1. Für jedes  $A \in \mathfrak{A}$  ergibt sich aus A1. und A2. wegen  $A \cup \bar{A} = \Omega$  und  $A \cap \bar{A} = \emptyset$  die Gleichung

$$P(\bar{A}) = 1 - P(A). \quad (3.1)$$

In den zwei folgenden Punkten seien  $A$  und  $B$  irgend zwei Ereignisse aus  $\mathfrak{A}$ .

2. Stets gilt

$$P(B) = P(B \cap A) + P(B \setminus A) \text{ und} \quad (3.2)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (3.3)$$

3. Ist  $A \subseteq B$ , so folgt aus (3.2)

$$P(B \setminus A) = P(B) - P(A) \quad (3.4)$$

und somit

$$P(A) \leq P(B) \text{ (Monotonie der Verteilung } P)$$

4. Für alle  $A_1, A_2, \dots, \in \mathfrak{A}$  gilt

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (3.5)$$

5. Für alle  $A_1, A_2, \dots, \in \mathfrak{A}$  gilt

$$P\left(\bigcup_1^n A_k\right) \leq \sum_1^n P(A_k) \quad (\text{endliche Subadditivität}) \quad (3.6)$$

Das ergibt sich aus (3.3) mittels vollständiger Induktion.

Das Axiom A2.' ermöglicht es, die Wahrscheinlichkeiten von Ereignissen zu bestimmen, die im Zusammenhang mit unendlichen Folgen von Ereignissen stehen.

Das nächste Lemma und seine Folgerungen stellen zu A2. äquivalente Eigenschaften bereit.

**Lemma 3.7** ( *$\sigma$ -Stetigkeit von Wahrscheinlichkeitsverteilungen*)

Wenn für die Abbildung  $P$  von  $\mathfrak{A}$  in  $[0, 1]$  die Axiome A1. und A2. gelten, so ist A2.' äquivalent mit jeder der folgenden Eigenschaften:

a) Für jede monoton fallende Folge  $(A_n, n \geq 1)$  aus  $\mathfrak{A}$  mit

$$\bigcap_{n=1}^{\infty} A_n = \emptyset \quad \text{gilt} \quad \lim_{n \rightarrow \infty} P(A_n) = 0$$

b) Für jede monoton wachsende Folge  $(A_n, n \geq 1)$  aus  $\mathfrak{A}$  mit

$$\bigcup_{n=1}^{\infty} A_n = \Omega \quad \text{gilt} \quad \lim_{n \rightarrow \infty} P(A_n) = 1$$

Beweis:

A2.'  $\implies$  a): Mit  $B_n = A_n \setminus A_{n+1}$ ,  $n \geq 1$  ist  $(B_n, n \geq 1)$  eine Folge paarweise disjunkter Ereignisse mit  $\bigcup_{n=m}^{\infty} B_n = A_m, m \geq 1$ . Folglich gilt mit Axiom A2' die Gleichung

$$P(A_m) = \sum_{n=m}^{\infty} P(B_n), \quad m \geq 1.$$

Somit haben wir wegen  $\sum_{n=1}^{\infty} P(B_n) = P(A_1) \leq 1$  die Beziehung  $\lim_{m \rightarrow \infty} P(A_m) = 0$ , also gilt a).

a)  $\implies$  b):  $(\bar{A}_n, n \geq 1)$  ist monoton fallend mit  $\bigcap_{n=1}^{\infty} \bar{A}_n = \emptyset$ , somit gilt  $\lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} (1 - P(\bar{A}_n)) = 1 - 0 = 1$ . Somit haben wir b) gezeigt.

b)  $\implies$  A2.': Ist  $(C_n, n \geq 1)$  eine Folge paarweise disjunkter Ereignisse, so definieren wir  $C'_n = \bigcup_{m=1}^n C_m, C'_\infty = \bigcup_{m=1}^{\infty} C_m, A_n = C'_n \cup \bar{C}'_\infty, n \geq 1$ . Damit folgt  $A_n \subseteq A_{n+1}, n \geq 1$  und  $\bigcup_{n=1}^{\infty} A_n = \Omega$ , und deshalb nach Voraussetzung

$\lim_{n \rightarrow \infty} P(A_n) = 1$ . Wegen Axiom A2. ergibt sich  
 $1 = \lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} P(C'_n) + P(\bar{C}'_\infty)$ , also  $P(C'_\infty) = 1 - P(\bar{C}'_\infty) = \lim_{n \rightarrow \infty} P(C'_n) = \lim_{n \rightarrow \infty} \sum_{m=1}^n P(C_m) = \sum_{m=1}^{\infty} P(C_m)$ . Damit ist A2' nachgewiesen.  $\square$

### Folgerungen 3.8

1. Ist  $(A_n)$  eine monoton fallende (monoton wachsende) Folge aus  $\mathfrak{A}$ , so gilt

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right) \text{ bzw. } \lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right).$$

Beweis: Man wende das Lemma 3.7 auf  $\left(A_n \setminus \left(\bigcap_{k=1}^{\infty} A_k\right)\right)$

bzw. auf  $\left(A_n \cup \left(\overline{\bigcup_{k=1}^{\infty} A_k}\right)\right)$  an.  $\square$

2. Für jede Folge  $(A_n)$  aus  $\mathfrak{A}$  gilt:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{k=1}^{\infty} P(A_k) \quad (\text{abzählbare Subadditivität}) \quad (3.7)$$

Beweis:  $B_n = \bigcup_1^n A_k \uparrow B = \bigcup_1^{\infty} A_k$ , beachte die Ungleichung (3.5).  $\square$

Mittels der eben bewiesenen Folgerungen 1. und 2. ergibt sich das

**Lemma 3.9** (*Erstes Borel-Cantelli-Lemma*)

Falls  $A_n \in \mathfrak{A}$ ,  $n \geq 1$ , und  $\sum_{n=1}^{\infty} P(A_n) < \infty$ , so gilt

$$P\left(\limsup_{n \rightarrow \infty} A_n\right) = 0$$

Beweis:  $P(\limsup_{n \rightarrow \infty} A_n) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{m \geq n} A_m\right)$

$$\leq \lim_{n \rightarrow \infty} \sum_{m=n}^{\infty} P(A_m) = 0 \quad \text{wegen Ungleichung (3.6).}$$

$\square$

In Worten kann man dieses Lemma wie folgt fassen: Gilt  $\sum_{n=1}^{\infty} P(A_n) < \infty$ , so ist die Wahrscheinlichkeit dafür, dass unendlich viele der Ereignisse  $A_n$  eintreten, gleich Null. Anders ausgedrückt, mit Wahrscheinlichkeit Eins treten höchstens endlich viele der Ereignisse  $A_n$  ein.

Wir geben noch eine nützliche Formel zur Berechnung von  $P\left(\bigcup_{k=1}^{\infty} A_k\right)$  an, bei der die  $(A_k, k = 1, \dots, n)$  nicht paarweise disjunkt sein müssen. Sie ist eine Verallgemeinerung der Formel 3.3.

**Aussage 3.10 (Ein- und Ausschlussformel)**

Für alle  $n \geq 2$  und alle  $A_1, A_2, \dots, A_n \in \mathfrak{A}$  gilt

$$\begin{aligned}
P\left(\bigcup_{k=1}^n A_k\right) &= \sum_{k=1}^n P(A_k) - \sum_{1 \leq i < k \leq n} P(A_i \cap A_k) \\
&\quad + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \dots (-1)^{n+1} P(A_1 \cap \dots \cap A_n) \\
&= \sum_{r=1}^n (-1)^{r+1} \sum_{\substack{J_r \subseteq \{1, \dots, n\} \\ \text{card } J_r = r}} P(A_{k_1} \cap \dots \cap A_{k_r}) \tag{3.8}
\end{aligned}$$

wobei  $\mathfrak{J}_r = \{k_1, k_2, \dots, k_r\}$  alle  $r$ -elementigen Teilmengen von  $\{1, \dots, n\}$  durchläuft.

Beweis mittels vollst. Induktion, siehe z. B. Henze (2006), Kap. 11.

Die Ein- und Ausschlussformel vereinfacht sich wesentlich, falls die Wahrscheinlichkeiten  $P(A_{k_1} \cap \dots \cap A_{k_r})$  nur von  $r$  und nicht von der Wahl des Tupels  $(k_1, \dots, k_r)$  abhängen. Wir definieren:

**Definition 3.11** *Es seien  $(\Omega, \mathfrak{A}, P)$  ein  $W$ -Raum und  $A_1, \dots, A_n$  Ereignisse aus  $\mathfrak{A}$ . Diese Ereignisse heißen (untereinander) austauschbar, falls  $P(A_{k_1} \cap \dots \cap A_{k_r}) = P(A_1 \cap \dots \cap A_r)$  gilt für alle  $r$  mit  $1 \leq r \leq n$  und alle  $r$ -elementigen Teilmengen  $\{k_1, \dots, k_r\}$  von  $\{1, \dots, n\}$  gilt.*

**Aussage 3.12** *Sind  $A_1, \dots, A_n$  austauschbar, so gilt*

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{r=1}^n (-1)^{r+1} \binom{n}{r} P(A_1 \cap \dots \cap A_r). \tag{3.9}$$

Beweis: Es gibt  $\binom{n}{r}$  Teilmengen  $\mathcal{J}_r$  von  $\{1, \dots, n\}$  mit  $r$  Elementen. □

**Aussage 3.13 (Bonferroni-Ungleichungen)**

Falls  $A_i \in \mathfrak{A}, i = 1, \dots, n$ , dann gilt

$$P\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j) \quad (3.10)$$

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j) \\ + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) \quad (3.11)$$

Beweis: mittels vollständiger Induktion.

Als Ergänzung erwähnen wir schließlich folgende Formel.

**Aussage 3.14** Für alle  $n \geq 2$  und alle  $A_1, A_2, \dots, A_n$  aus  $\mathfrak{A}$  und

$A_{n,m} :=$  "Es treten genau  $m$  der  $n$  Ereignisse  $A_1, A_2, \dots, A_n$  ein"  $1 \leq m \leq n$ , gilt:

$$P(A_{n,m}) = \sum_{r=m}^n (-1)^{r-m} \binom{r}{m} \sum_{\substack{J_r \subseteq \{1, \dots, n\} \\ \text{card } J_r = r}} P(A_{k_1} \cap \dots \cap A_{k_r}) \quad (3.12)$$

Der Beweis erfolgt ebenfalls mittels vollständiger Induktion.

## 3.2 Laplace-Experimente

In diesem Abschnitt werden wir erste konkrete mathematische Modelle zufälliger Versuche kennen lernen, die sogenannten *Laplace-Experimente*. Sie zeichnen sich durch besondere Einfachheit aus und sind dennoch in Anwendungen häufig anzutreffen.

**Definition 3.15** Als Laplace-Experiment (kurz: L-Experiment) bezeichnet man einen zufälligen Versuch mit:

1. der Versuch hat nur endlich viele ( $= N$ ) mögliche Ausgänge.
2. Alle Ausgänge haben die gleiche Wahrscheinlichkeit.

Als mathematisches Modell eines L-Experimentes wählt man einen Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  mit:

1.  $\Omega$  ist endlich:  $\Omega = \{1, 2, \dots, N\}$  (allgemeiner, aber äquivalent dazu:

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_N\} \text{ mit } \omega_i \neq \omega_j, i \neq j),$$

2.  $\mathfrak{A} = \mathfrak{P}(\Omega)$  und alle Versuchsausgänge sind gleichwahrscheinlich:

$$P(\{\omega\}) \equiv: p, \omega \in \Omega. \quad (3.13)$$

$N$  heißt der Parameter des L-Experimentes.

Auf Grund (3.11) gilt wegen

$$1 = P(\Omega) = P\left(\bigcup_{k=1}^N \{\omega_k\}\right) = \sum_{k=1}^N P(\{\omega_k\}) = N \cdot p$$

die Beziehung  $P(\{\omega\}) \equiv p = \frac{1}{N}$  und für jede Teilmenge  $A$  von  $\Omega$  ergibt sich mit  $N(A) =$  Anzahl der Elemente von  $A$ :

$$P(A) = P\left(\bigcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} P(\{\omega\}) = p \cdot N(A) = \frac{N(A)}{N} = \quad (3.14)$$

$$= \frac{\text{Anzahl der (für } A) \text{ günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}. \quad (3.15)$$

Die Gesamtwahrscheinlichkeit 1 (d.h., die Wahrscheinlichkeit des sicheren Ereignisses "Irgendein Versuchsausgang tritt ein") ist bei einem Laplace-Experiment gleichmäßig auf die Versuchsausgänge  $\omega$  verteilt: Man nennt  $P$  auch die *gleichmäßige Wahrscheinlichkeitsverteilung* auf  $\{1, 2, \dots, N\}$  oder einfach die *Gleichverteilung* auf  $\{1, 2, \dots, N\}$  bzw. auf  $\Omega = \{\omega_1, \dots, \omega_N\}$ .

Bei Laplace-Experimenten spricht man auch davon, dass das Versuchsergebnis "auf gut Glück" oder "rein zufällig" ausgewählt wird, um die Gleichwahrscheinlichkeit aller möglichen Versuchsausgänge hervorzuheben.

**Beispiel 3.16** Der zufällige Versuch bestehe im Werfen zweier regulärer Würfel und im Registrieren, welche Augenzahl der erste und welche der zweite Würfel zeigt. Wir setzen also  $\Omega = \{(i, j) : i, j \in \{1, 2, \dots, 6\}\}$  mit  $i =$  Augenzahl des ersten Würfels und  $j =$  Augenzahl des zweiten Würfels.

Alle Ergebnisse sind aus Symmetriegründen (reguläre Würfel) gleichwahrscheinlich, also gilt:

$$P((i, j) \text{ tritt auf}) = \frac{1}{36} \quad \text{für alle } i, j \in \{1, \dots, 6\}.$$

Mit  $A :=$  "Die Augensumme ist gleich 6" haben wir

$$P(A) = \frac{N(A)}{N} = \frac{5}{36},$$

und für  $B :=$  "Die Augenzahlen sind verschieden" erhalten wir

$$P(B) = \frac{N(B)}{N} = \frac{30}{36} = \frac{5}{6}.$$

In Anwendungsbeispielen mit endlichem  $\Omega$  muss man genau prüfen, ob es sich tatsächlich um ein Laplace-Experiment handelt.

Beobachtet man zum Beispiel im obigen Beispiel nur die Augensumme, so ist dies ein neuer zufälliger Versuch. Man wählt  $\Omega = \{2, \dots, 12\}$ . Jetzt sind aber nicht alle Ausgänge gleichberechtigt, d.h. gleichwahrscheinlich:

"Augensumme = 2" hat eine kleinere Wahrscheinlichkeit ( $= \frac{1}{36}$ ) als "Augensumme = 4" ( $= \frac{1}{12}$ ), denn das erste Ereignis tritt nur beim Versuchsausgang (1,1), das zweite dagegen bei jedem der Ausgänge (1,3), (2,2), (3,1) ein.

Wir kehren zurück zum Modell des  $n$ -maligen Münzenwurfes aus Abschnitt 2.5.

### 3.3 Münzenwurf, zum Zweiten

Wir setzen hier das Studium des zufälligen Versuches "n-maliges Werfen einer Münze" aus Abschnitt 2.5 fort. Die Münze, die wir für das Spiel verwenden, sei regulär, d. h. symmetrisch. Das bedeute, beide Seiten erscheinen bei einem Wurf mit gleicher Chance. Dann ist das  $n$ -malige Werfen ein Laplace-Experiment mit  $2^n$  gleichwahrscheinlichen Ausgängen. Der entsprechende Wahrscheinlichkeitsraum ist  $(\Omega, \mathfrak{P}(\Omega), P)$  mit  $P(\{\omega\}) = 2^{-n}, \omega \in \Omega$  und

$$P(A) = \sum_{\omega \in A} 2^{-n} = \frac{N(A)}{2^n}, A \subseteq \Omega. \quad (3.16)$$

Wir berechnen die Wahrscheinlichkeit dafür, dass nach  $n$  Würfeln der zugehörige Pfad  $(k, S_k(\omega)), k \geq 0$ , mit  $S_0(\omega) := 0$  bei  $r$  endet:  $P(\{\omega \in \Omega | S_n(\omega) = r\})$ .

**Aussage 3.17** Für  $P(S_n = r)$  gelten die Formeln

a) Ist  $n$  gerade, also  $n = 2m$  für ein  $m \geq 1$ , so gilt

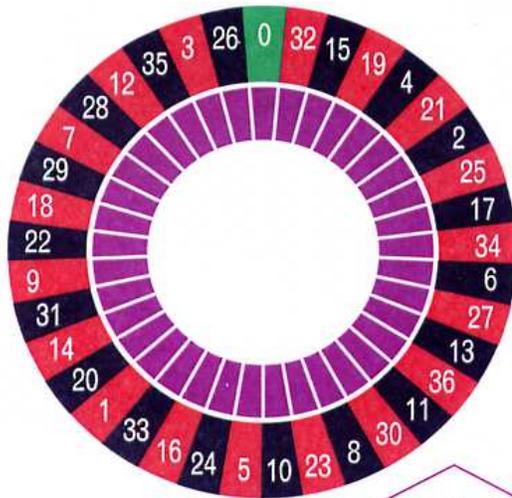
$$\begin{aligned} P(S_{2m} = 2l) &= \binom{2m}{m+l} 2^{-2m} \quad \text{falls } |l| \leq m, \\ P(S_{2m} = r) &= 0 \quad \text{für alle anderen ganzzahligen } r. \end{aligned}$$

b) Ist  $n$  ungerade, also  $n = 2m + 1$  für ein  $m \geq 1$ , so ist

$$\begin{aligned} P(S_{2m+1} = 2l + 1) &= \binom{2m+1}{m+l+1} 2^{-2m-1} \quad \text{falls } -m - 1 \leq l \leq m, \\ P(S_{2m+1} = r) &= 0 \quad \text{für alle anderen ganzzahligen } r. \end{aligned}$$

Bemerkung: In beiden Fällen handelt es sich um eine um den Nullpunkt symmetrische Verteilung mit Null bzw.  $\pm 1$  als Punkte maximaler Wahrscheinlichkeit, vgl. Abschnitt 4.

Die Folge  $(S_0, S_1, \dots, S_n)$  heißt auch eine symmetrische Irrfahrt.



- A. One Number or 'Straight Up'
- B. Two Numbers or 'Split'
- C. Three Numbers or 'Street'
- D. Four Numbers or 'Corner'
- E. Six Numbers or 'Six Line'
- F. Column
- G. Dozen
- H. Low (1-18)
- I. High (19-36)
- J. Even
- K. Odd
- L. Red
- M. Black

			0
	D		C
	1	2	3
	D	4	5
H	1-18	6	7
J	EVEN	8	9
G	1 <sup>st</sup> DOZEN	10	11
		12	13
L	♦	14	15
G	2 <sup>nd</sup> DOZEN	16	17
		18	19
M	♦	20	21
		22	23
K	ODD	24	25
		26	27
I	19-36	28	29
G	3 <sup>rd</sup> DOZEN	30	31
		32	33
	E	B	
	A	35	36
	F	F	F

Abbildung 3.1: Quantile der Standardnormalverteilung

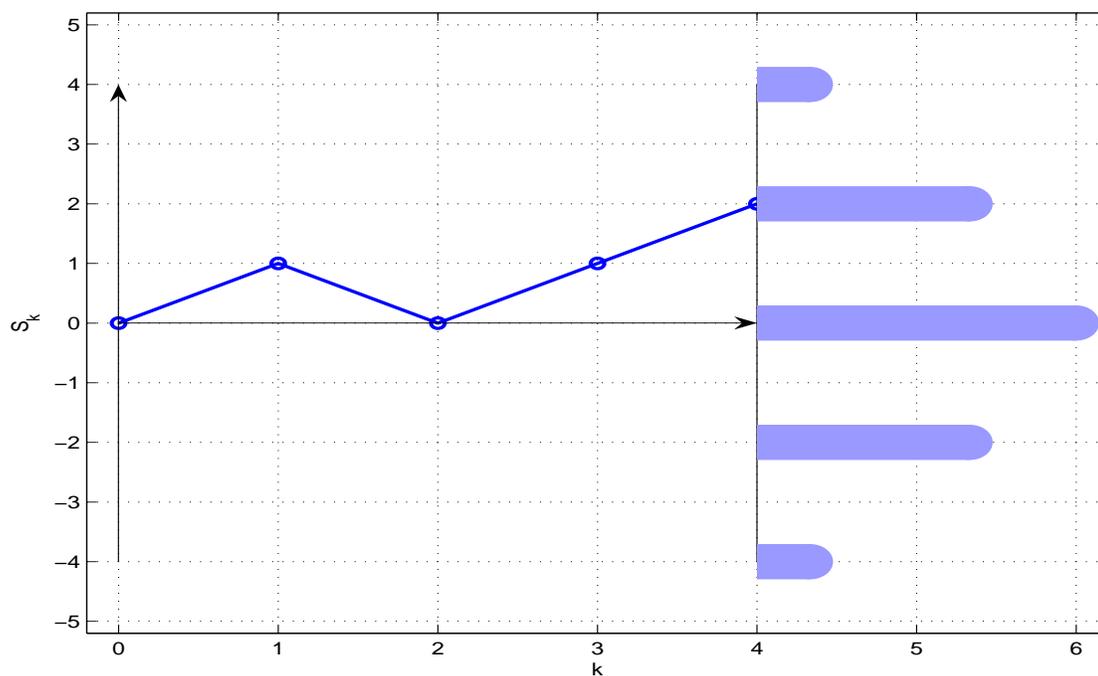


Abbildung 3.2: Beispiel des Pfades der Länge vier und der Orte der Pfadenden

Beweis:

- a)  $\{S_{2m} = 2l\}$  tritt genau dann ein, wenn in  $\omega = (x_1, \dots, x_{2m})$  genau  $l + m$  mal die Eins enthalten ist.
- b)  $\{S_{2m+1} = 2l + 1\}$  tritt genau dann ein, wenn  $\omega = (x_1, \dots, x_{2m+1})$  genau  $m + l + 1$  Einsen enthält.

Für den Spieler  $A$  (siehe Abschnitt 2.5) ist es von Interesse, mit welcher Wahrscheinlichkeit er wann zum ersten Mal ein negatives Guthaben hat.

**Aussage 3.18** *Die Wahrscheinlichkeit dafür, dass Spieler  $A$  zur Zeit  $n$  zum ersten Mal ein negatives Gutachten hat, ist für ungerades  $n = 2m + 1, m \geq 1$ , gleich dem Wert*

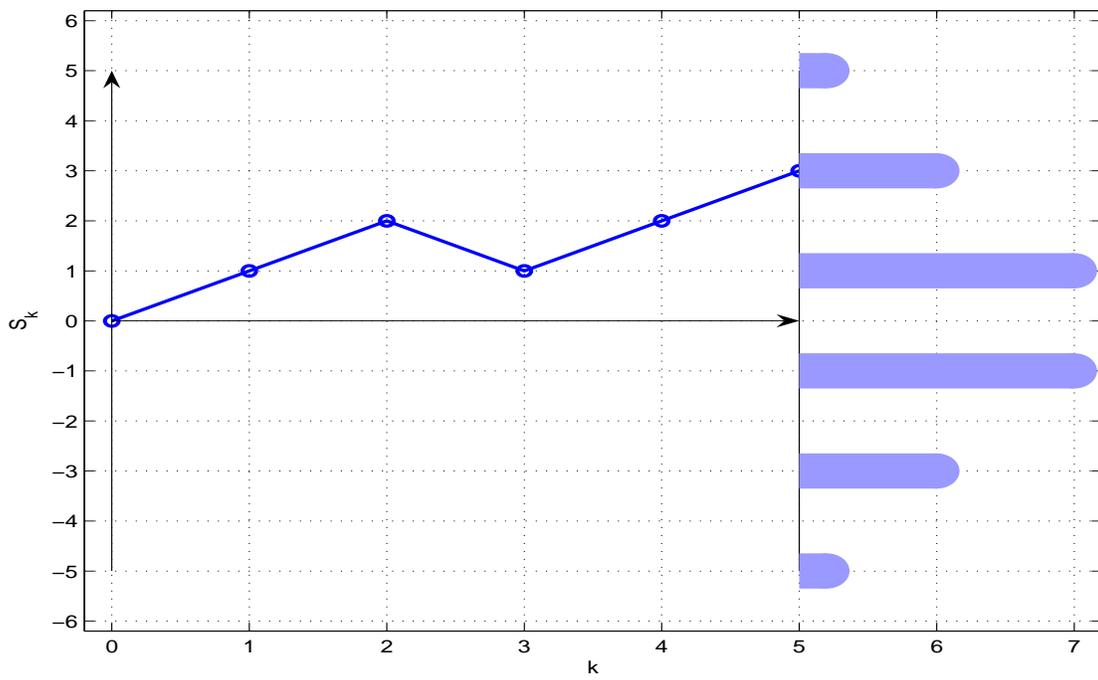


Abbildung 3.3: Beispiel des Pfades der Länge fünf und der Orte der Pfadenden

$$\begin{aligned}
& P(\{\omega : s_1 \geq 0, s_2 \geq 0, \dots, s_{n-1} \geq 0, s_n = -1\}) \\
&= P(S_1 \geq 0, S_2 \geq 0, \dots, S_{n-1} \geq 0, S_n = -1) = \frac{(2m-1)!!}{(m+1)!2^{m+1}} \quad (3.17)
\end{aligned}$$

mit  $(2m-1)!! = (2m-1)(2m-3)\cdots 3 \cdot 1$  und für gerades  $n$  gleich Null.

**Beweis:** Für jeden der für das betrachtete Ereignis günstigen Pfade  $s = (s_0, s_1, \dots, s_n)$  gilt  $s_{n-1} = 0$ . Wir bestimmen deshalb die Zahl aller Pfade von  $(0, 0)$  nach  $(2m, 0)$  die  $-1$  niemals berühren und beachten dabei die Eigenschaft (2.1) jedes Pfades.

Es gibt insgesamt  $\binom{2m}{m}$  Pfade von  $(0, 0)$  nach  $(2m, 0)$ . Zur Berechnung der gesuchten Zahl der Pfade bedienen wir uns des sogenannten *Spiegelungsprinzips*. Jedem durch ein  $\omega$  erzeugten Pfad  $s = (s_0, s_1, \dots, s_n)$ , der die Zahl  $-1$  jemals vor  $n$  erreicht, wird der Pfad  $s'$  zugeordnet, der bei  $-2$  startet und bis zur Zeit  $T_{-1}(\omega) = \min\{k \geq 1 | s_k = -1\}$  spiegelbildlich bezüglich der Horizontalen der Höhe  $-1$  zu  $(s_1, s_2, \dots, s_n)$  verläuft, sowie danach mit  $(s_1, s_2, \dots, s_n)$  übereinstimmt.

Die Zuordnung ist eineindeutig. Folglich ist die Zahl der Pfade, die  $-1$  vor der Zeit  $n$  berühren und zur Zeit  $n-1 = 2m$  bei Null sind gleich der Zahl der Pfade, die bei  $-2$  starten und zur Zeit  $n-1 = 2m$  in Null enden.

Davon gibt es  $\binom{2m}{m+1}$  Exemplare. Somit ist die gesuchte Anzahl gleich

$$\binom{2m}{m} - \binom{2m}{m+1} = \binom{2m}{m} \frac{1}{m+1}.$$

Die gesuchte Wahrscheinlichkeit ist also  $2^{-2m-1} \binom{2m}{m} \frac{1}{m+1}$ . Eine einfache Umformung liefert 3.16. Zu geraden Zeiten  $n$  kann  $S_n$  nicht zum ersten Mal negativ sein.  $\square$

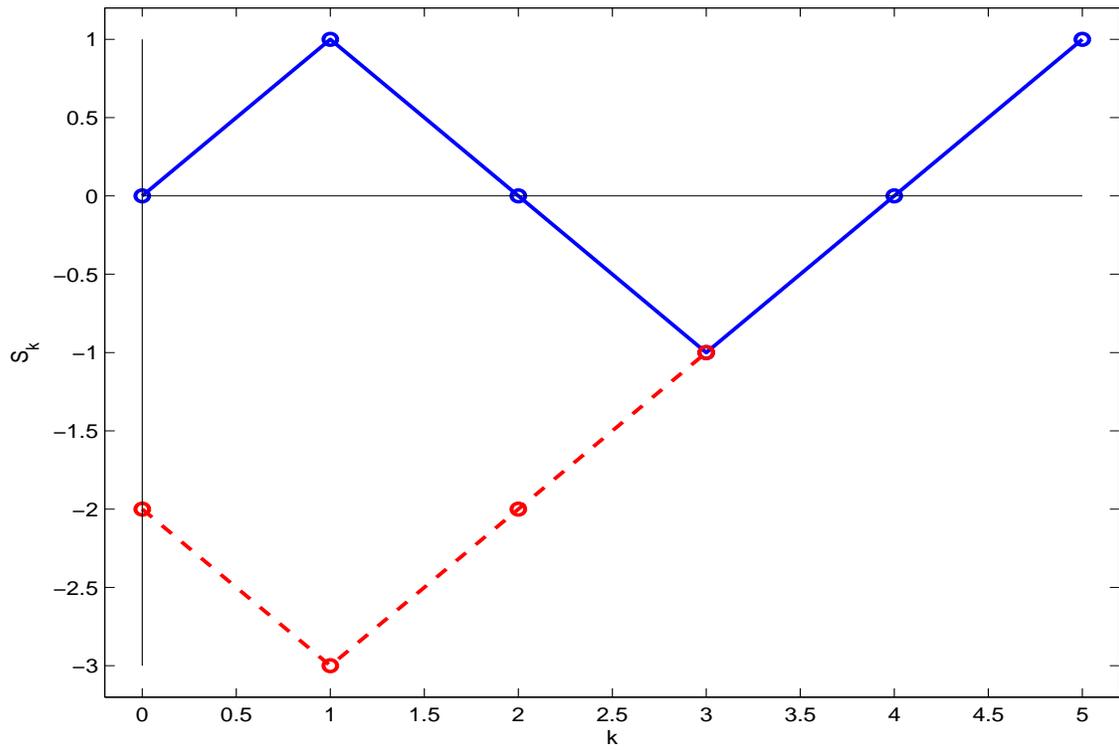


Abbildung 3.4: Gespiegelter Pfad

### 3.4 Was sagen uns Wahrscheinlichkeiten?

Welche anschauliche Bedeutung hat die Wahrscheinlichkeit  $P(A)$  eines zufälligen Ereignisses  $A$ ? Es gibt zwei grundlegende Erfahrungen im Umgang mit dem Zufall.

- a) Empirisches Gesetz der großen Zahlen: In einer langen Reihe gleichartiger, voneinander unabhängiger Versuche ist die relative Häufigkeit  $\frac{n(A)}{n}$  des Eintretens von  $A$  etwa gleich  $P(A)$ . Wenn  $P(A) > \frac{1}{2}$  gilt, so kann man auf das Eintreten von  $A$  Wetten abschließen und wird bei fortlaufenden Wetten dieser Art schließlich im Vorteil sein.
- b) Es ist neben dem empirischen Gesetz der großen Zahlen eine zweite Erfahrungstatsache, dass zufällige Ereignisse mit sehr kleinen Wahrscheinlichkeiten bei einmaliger Versuchsdurchführung praktisch nicht eintreten.

Genauer gesagt: Man muss bei einem zufälligen Versuch, den man einmal durchführt, mit dem Eintreten von  $A$  nicht rechnen, falls  $P(A)$  sehr klein ist. Diese Erfahrung hat jeder Mensch verinnerlicht.

Beispiel für a):

- 1) Werfen Sie einen regulären Spielwürfel mehrere Mal und beobachten Sie das Verhalten der relativen Häufigkeit des Auftretens einer "Sechs" im Verlaufe der Würfe. Sie tendiert zu  $\frac{1}{6}$ .

Beispiele für b):

- 1) Man erhält keinen Kredit von der Bank, wenn man als Sicherheit anbietet, dass man auf seinen wöchentlichen Tippschein im Lotto "6 aus 49" innerhalb eines Jahres einen "Sechser" erzielt. Die Wahrscheinlichkeit dieses Ereignisses ist so gering, dass man mit seinem Eintreten nicht wirklich rechnet.

- 2) Man rechnet nicht damit, dass jemand durch maximal dreimaliges zufälliges Raten der PIN bei einer EC-Karte die richtige PIN errät.

$$P(A_1 \cup A_2 \cup A_3) \approx \frac{3}{10^4} = 0,0003$$

- 3) Wenn ein vorgegebenes Ereignis, das eine sehr kleine Wahrscheinlichkeit hat, tatsächlich eintritt, so zweifelt man mitunter daran, dass der "reine Zufall" zu diesem Ereignis geführt hat. Man stellt eher die Richtigkeit der zugrunde liegenden Annahmen in Frage und prüft sie sorgfältig: Den Ausspruch "Das kann kein Zufall sein!" hat jeder schon mal gehört.

Beispiel seltener Ziehungen beim Lotto "6 aus 49" wie (1, 2, 3, 4, 5, 6) führen regelmäßig zur Aufmerksamkeit der Medien und der Frage, ob hier nicht der Zufall außer Kraft gesetzt sei. Diese Zahlenkombination hat aber die gleiche (geringe) Wahrscheinlichkeit wie jede andere.

- 4) Aus Dorothy L. Sayers "Keines natürlichen Todes", rororo, 1991:
- a) S. 58<sub>8</sub>: Ein merkwürdiger Zufall, sagte er (der Chef von Scotland Yard) geduldig, und ich kann verstehen, dass Sie sich darüber aufregen.
  - b) S. 62<sub>14</sub>: Schon wieder ein Reifall sagte Winsey: "Aber ein sonderbarer Zufall ist das schon."
- 5) Wenn jemand beim Skatspiel dreimal hintereinander alle vier Buben erhält, glaubt man nicht mehr an reinen Zufall, obwohl dieses Ereignis eine positive, wenn auch sehr kleine Wahrscheinlichkeit hat.

### 3.5 Elemente der Kombinatorik\*

Bei der Abzählung der "günstigen" Fälle bei  $L$ -Experimenten erweisen sich Formeln der Kombinatorik häufig als günstig.

Wir geben hier vier Grundaufgaben der Kombinatorik an, sie werden häufig

auch in Form sogenannter Urnenprobleme formuliert.

Wir beginnen mit einer elementaren aber wichtigen Feststellung.

**Aussage 3.19** *Es seien  $M_1, \dots, M_m$   $m$  Mengen mit  $m_1, \dots, m_m$  Elementen. Dann hat die Menge  $M$  aller  $m$ -Tupel  $(i_1, \dots, i_m)$  mit  $i_k \in M_k$  für  $k = 1, \dots, m$  genau  $m_1, m_2, \dots, m_m$  Elemente.*

Beweis: Mittels vollständiger Induktion

Als Nächstes kommen wir zu den vier angekündigten Aufgaben der Kombinatorik.

Aus einer Menge  $M = \{a_1, a_2, \dots, a_m\}$  von  $m$  Elementen ( $m \geq 1$ ) werden  $r$  Elemente ausgewählt,  $r \geq 1$ . Man spricht von einer Stichprobe vom Umfang  $r$  aus der Menge  $M$ . Die Entnahme von Stichproben kann auf unterschiedliche Weise erfolgen:

Mit Wiederholung oder ohne Wiederholung

Mit Berücksichtigung der Reihenfolge oder ohne Berücksichtigung der Reihenfolge.

(d. h. geordnet oder ungeordnet)

Dementsprechend unterscheiden wir vier Fälle.

Anzahl von möglichen Stichproben des Umfanges  $r$  aus einer Menge  $M$  vom Umfang  $m$

(In den Fällen ohne Wiederholung ist  $r \leq m$  vorauszusetzen.)

	mit Wiederholung	ohne Wiederholung
geordnet	$r$ -Tupel $(a_1, \dots, a_r)$ mit $a_i \in M, i = 1, \dots, r$  $r$ -Permutation mit $W$ , $m^r$ <b>A1</b>	$r$ -Tupel $(a_1, \dots, a_r)$ mit $a_i \in M, i \in 1, \dots, r$ , paarw. verschieden $r$ -Permutation ohne $W$ . $m(m-1) \dots (m-r+1) =: (m)_r$ <b>A2</b>
ungeordnet	$[a_1, a_2, \dots, a_r]$ , Anordnung von $r$ Elementen $\{a_1, \dots, a_r\} \subset M$ $a_i \in M, i = 1, \dots, r$ $r$ -Kombination mit $W$ . $\binom{m+r-1}{r}$ <b>A3</b>	Teilmenge vom Umfang $r$ $r$ -Kombination ohne $W$ . $\binom{m}{r} = \frac{(m)_r}{r!}$ <b>A4</b>

Die Fälle A1, A2 und A4 sind leicht zu beweisen.

Der Fall A3:

Jede ungeordnete Stichprobe vom Umfang  $r$  mit Wiederholung aus der Menge  $M$  ist eindeutig charakterisiert durch eine Folge  $(i_1, i_2, \dots, i_m)$  natürlicher Zahlen  $i_k \geq 0$  mit  $\sum_{k=1}^m i_k = r$ , wobei  $i_k$  angibt, wie oft das Element  $a_k$  aus  $M$  in der Stichprobe vorkommt.

Diese Vektoren  $(i_1, \dots, i_m)$  lassen sich eineindeutig auf die Menge aller Anordnungen der Form  $\bullet \bullet \bullet | \bullet \bullet || \bullet \bullet \bullet$  von  $r$  Punkten und  $(m-1)$  Strichen abbilden, wobei vor dem ersten Strich  $i_1$  Punkte stehen, zwischen dem  $k$ -ten und  $(k+1)$ -ten Strich  $i_{k+1}$  Punkte stehen, und nach dem  $(m-1)$ -ten Strich  $i_m$  Punkte platziert sind. Insgesamt gibt es  $\binom{m+r-1}{r}$  solcher Anordnungen.

Zu jedem der vier Fälle der Entnahme von Stichproben vom Umfang  $r$  aus einer Menge vom Umfang  $m$  gibt es ein sogenanntes "duales Problem" der Verteilung von  $r$  Kugeln auf  $m$  Urnen.

$A_1$ : Duales Problem:  $r$  unterscheidbare Kugeln werden auf  $m$  Urnen verteilt, wobei in jeder Urne beliebig viele Kugeln liegen dürfen. Auf wie viel verschiedene Weisen ist dies möglich? Jede Verteilung der Kugeln ist charakterisiert durch die  $r$  Nummern  $(a_1, a_2, \dots, a_r)$  der Urnen, in die die erste zweite, ...,  $r$ -te Kugel zu liegen kommt. Für jede dieser Nummern gibt es  $m$  Möglichkeiten. Im Ergebnis entsteht wieder eine Stichprobe  $(a_1, a_2, \dots, a_r)$  vom Umfang  $r$  mit Wiederholung aus einer Menge vom Umfang  $m$ .

$A_2$ : Duales Problem:  $r$  unterscheidbare Kugeln werden auf  $m$  Urnen verteilt, wobei in jeder Urne höchstens eine Kugel liegen darf. Auf wie viel verschiedene Weisen ist dies möglich? Jede Verteilung der Kugel ist charakterisiert durch die  $r$  Nummern  $(a_1, a_2, \dots, a_r)$  der Urnen, in die die erste, zweite  $k$ -te,  $r$ -te Kugel zu liegen kommt. Für die erste Kugel gibt es  $m$ , für die zweite  $m - 1$ , für die  $r$ -te Kugel  $m - r + 1$  Möglichkeiten. Im Ergebnis entsteht wieder eine Stichprobe  $(a_1, a_2, \dots, a_r)$  vom Umfang  $r$  ohne Wiederholung aus einer Menge vom Umfang  $m$ .

$A_3$ : Duales Problem:  $r$  ununterscheidbare Kugeln werden auf  $m$  Urnen aufgeteilt, wobei jede Urne auch mehrfach besetzt werden kann. Jede Aufteilung ist charakterisiert durch die Anzahl  $i_k$  der Kugeln, die in die  $k$ -te Urne fallen,  $k = 1, \dots, m$ ,  $\sum_{k=1}^m i_k = r$ .

$A_4$ : Duales Problem:  $r$  ununterscheidbare Kugeln sind auf  $m$  Urnen so aufzuteilen, dass in jeder Urne höchstens eine Kugel zu liegen kommt. Die Aufteilung ist charakterisiert durch die Menge  $\{a_1, a_2, \dots, a_r\}$  der Urnen, die durch eine Kugel besetzt werden, also durch eine  $r$ -elementige Teilmenge von  $M$ .

Anzahl der Aufteilungen von  $r$  Kugeln auf  $m$  Urnen

	mit Mehrfach- -besetzung	ohne Mehrfach- -besetzung
unterscheidbare Kugeln	$m^r$ <b>A1</b>	$m(m-1) \dots (m-r+1)$ <b>A2</b>
ununterscheidbare Kugeln	$\binom{m+r-1}{r}$ <b>A3</b>	$\binom{m}{r}$ <b>A4</b>

Beispiele:

1. Wie groß ist die Wahrscheinlichkeit, beim Werfen von vier Würfeln mindestens eine "Sechs" zu erzielen?

Der zufällige Versuch "Werfen von vier Würfeln" ist ein Laplace-Experiment mit  $6^4$  möglichen Ausgängen. Es bezeichne  $A$  das Ereignis "Es erscheint mindestens eine Sechs". Dann gibt es  $5^4$  günstige Ausgänge für das komplementäre Ereignis  $\bar{A}$  = "Es erscheint keine Sechs". Also gilt

$$P(A) = 1 - \left(\frac{5}{6}\right)^4 = 0,52.$$

2. Es werden  $k$  Kugeln auf  $n$  Urnen aufgeteilt,  $k \leq n$ . Jede Kugel habe die gleiche Wahrscheinlichkeit in jede Urne zu gelangen. Wie groß ist die Wahrscheinlichkeit für das Ereignis  $A$ , dass es nach der Aufteilung Urnen gibt, in der mehr als eine Kugel liegt?

Lösung: Es gibt  $n^k$  Möglichkeiten der geschilderten Aufteilung und  $\binom{n}{k}$  Möglichkeiten, die günstig sind für das komplementäre Ereignis  $\bar{A}$  = "In keiner Urne liegt mehr als eine Kugel". Daraus folgt

$$P(A) = 1 - \frac{\binom{n}{k}}{n^k}.$$

3. In einem Raum mögen sich  $k$  Personen befinden. Wie groß ist die Wahrscheinlichkeit für das Ereignis  $A$ , dass mindestens zwei dieser Personen am gleichen Tag Geburtstag haben? (Jeder Tag des Jahres komme bei

jeder Person mit gleicher Wahrscheinlichkeit als Geburtstag in Frage, Schaltjahre bleiben unberücksichtigt.)

Lösung:  $\bar{A}$  = "Alle  $k$  Personen haben an verschiedenen Tagen Geburtstag". Es gibt  $N = (365)^k$  möglich Fälle für Geburtstage und  $N(\bar{A}) = (365)_k$  für  $\bar{A}$  günstige Fälle, somit ist  $P(A) = 1 - \frac{(365)_k}{365^k}$ .

Diese Wahrscheinlichkeit wächst mit  $k$  und ist gleich 0,507 für  $k = 23$ .

4. Koinzidenzproblem:

$n$  Briefe werden auf rein zufällige Weise in  $n$  adressierte Umschläge gesteckt. Wie groß ist die Wahrscheinlichkeit, dass mindestens ein Brief in den richtigen Umschlag kommt?

Lösung: Mögliche Versuchsausgänge  $\omega = (a_1, \dots, a_n)$  sind die Permutation von  $(1, \dots, n)$  mit  $a_k$  gleich der Nummer des Umschlages, in den der  $k$ -te Brief kommt. Wir setzen

$$A_k := \{\omega \mid a_k = k\}, k = 1, \dots, n.$$

Das interessierende Ereignis  $A$  ist gleich  $\bigcup_{k=1}^n A_k$ . Zur Anwendung der Ein- und Ausschlussformel berechnen wir

$$P(A_{k_1} \cap \dots \cap A_{k_r}) = P(\{\omega = (a_1, \dots, a_n) \mid a_{k_1} = k_1, \dots, a_{k_r} = k_r\}) =$$

$$\frac{\text{card}\{\omega : a_{k_1} = k_1, \dots, a_{k_r} = k_r\}}{n!} = \frac{(n-r)!}{n!}.$$

Die Ein- und Ausschlussformel liefert

$$P(A) = \sum_{r=1}^n (-1)^{r+1} \binom{n}{r} \frac{(n-r)!}{n!} = \sum_{r=1}^n (-1)^{r+1} \frac{1}{r!} \xrightarrow{n \rightarrow \infty} 1 - \frac{1}{e} = 0,632$$

Für  $n \geq 7$  ist die Näherung auf drei Stellen genau. (Weitere Bemerkungen zum Koinzidenzproblem und anderen kombinatorischen Aufgaben findet man in Henze, Kap. 9-11.)

5. Am Eröffnungstag eines großen Kongresses mit 500 Teilnehmern soll jeder teilnehmenden Person, die an diesem Tag Geburtstag hat, ein Blumenstrauß überreicht werden. Wie viele Sträuße braucht man mindestens, wenn man mit Sicherheit ausschließen will, dass man zu wenige Sträuße hat?

Wie groß muss die Zahl der Sträuße mindestens sein, wenn man mit der Wahrscheinlichkeit von 0,95 diesen blamablen Fall vermeiden will?

Wir nehmen näherungsweise an, dass für jede Person die Wahrscheinlichkeit, an einem bestimmten Tag Geburtstag zu haben, gleich ist für alle Tage des Jahres. Schaltjahre werden nicht berücksichtigt. Dann ist die Feststellung der Geburtstage aller Teilnehmer ein Laplace-Experiment mit den möglichen Ausgängen  $\omega = (i_1, \dots, i_{500})$ , wobei  $i_k$  die Nummer des Tages angibt, an denen der  $k$ -te Teilnehmer Geburtstag hat. Die Menge aller möglichen Versuchsausgänge hat den Umfang  $N = 365^{500}$ . Es gibt nämlich  $N = 365^{500}$  Möglichkeiten der Verteilung der Geburtstage der 500 Personen auf das Jahr.

Für das Ereignis  $A_k :=$  "Genau  $k$  Personen haben am Eröffnungstag Geburtstag" gibt es  $N(A_k) = \binom{500}{k} \cdot 364^{500-k}$  "günstige" Versuchsausgänge.

Es gilt  $P(A_k) = \frac{N(A_k)}{N}$ , und folglich ergibt sich

$k$	0	1	2	3	4	5
$P(A_k)$	0,2532	0,3478	0,2384	0,1087	0,0371	0,0101

Deshalb ist

$$P\left(\bigcup_{k=1}^5 A_k\right) = \sum_{k=1}^5 P(A_k) = 0,9953.$$

Das heißt, mit an Eins grenzender Wahrscheinlichkeit haben höchstens fünf der Personen am Eröffnungstag Geburtstag.

(Zur Berechnung wurde die Näherung

$$P(A_k) \approx \frac{\lambda^k}{k!} e^{-\lambda} \text{ mit } \lambda = \frac{500}{365}$$

benutzt.)

### 3.6 Rein zufällige Wahl eines Punktes aus $[0, 1)$

Das Laplace-Experiment lässt sich nicht unmittelbar auf das in der Überschrift genannte Problem anwenden, da  $[0, 1)$  unendlich viele Punkte enthält. Wir müssen hier den Begriff der "rein zufälligen Wahl" etwas modifizieren.

Rein zufällige Wahl soll bedeuten, dass für jedes Intervall  $[a, b) \subseteq [0, 1)$  die Wahrscheinlichkeit, dass der gewählte Punkt aus  $[a, b)$  stammt, unabhängig von der Lage des Intervalls sein soll. Das heißt

$$P([a, b)) = P([a + x, b + x)) \tag{3.18}$$

für alle  $x$  mit  $a + x \geq 0, b + x \leq 1$ .

Daraus folgt

$$P([a, b)) = b - a. \tag{3.19}$$

(Beweisen Sie (3.19).)

Es existiert allerdings kein Wahrscheinlichkeitsmaß  $P$  auf der Potenzmenge von  $[0, 1)$  mit den Eigenschaften (3.18) und (3.19), siehe Elstrodt, III. 2. Man kann aber zeigen, dass ein Wahrscheinlichkeitsmaß  $P$  mit (3.18) und (3.19) existiert auf der kleinsten  $\sigma$ -Algebra  $\mathcal{B}_{[0,1]}$  von Teilmengen von  $[0, 1)$ , die alle Intervalle der Form  $[a, b)$  mit  $0 \leq a < b \leq 1$  enthält ( $\sigma$ -Algebra der Borelmengen aus  $[0, 1)$ ). Dieses Maß ist eindeutig bestimmt und heißt Lebesgue-Borel-Maß auf  $([0, 1), \mathcal{B}_{[0,1]})$  oder einfach Lebesguemaß. Wir werden es mit  $\lambda_{[0,1]}$  bezeichnen.

Die Tatsache, dass man  $\lambda_{[0,1]}$  unter Beibehaltung von (3.18) nicht auf  $\mathfrak{P}([0, 1))$  erweitern kann, führt zu der auf den ersten Blick eigenartigen Situation, dass man nicht jede Teilmenge  $C$  von  $[0, 1)$  als zufälliges Ereignis bei der rein zufälligen Wahl eines Punktes aus  $[0, 1)$  ansehen kann.

Der Wahrscheinlichkeitsraum  $([0, 1), \mathcal{B}_{[0,1]}, \lambda_{[0,1]})$  ist das mathematische Modell des zufälligen Versuches, einen Punkt aus dem Intervall "rein zufällig" oder "auf gut Glück" auszuwählen.

Die Wahrscheinlichkeitsverteilung  $\lambda_{[0,1]}$  "verteilt" die Gesamtwahrscheinlichkeit Eins "gleichmäßig" auf das Intervall  $[0, 1)$ . Sie heißt *gleichmäßige Verteilung* auf  $[0, 1)$ . Wir werden sie mit  $U([0, 1))$  bezeichnen. Insbesondere hat dann auch jeder Punkt  $x \in [0, 1)$  als Ereignis  $\{x\}$  die gleiche Wahrscheinlichkeit, die folglich gleich Null sein muss. Das folgt auch aus  $\{x\} = \bigcap_{k=1}^{\infty} \left[ x, x + \frac{1}{k} \right)$  und der  $\sigma$ -Stetigkeit von  $\lambda_{[0,1]}$ .

In dem eben eingeführten Wahrscheinlichkeitsraum gibt es Ereignisse, die nicht unmöglich (bzw. nicht sicher) sind, aber dennoch die Wahrscheinlichkeit Null (bzw. Eins) haben. Das führt uns auf folgende Definition.

**Definition 3.20** *Es sei  $(\Omega, \mathfrak{A}, P)$  ein Wahrscheinlichkeitsraum. Jedes Ereignis  $A \in \mathfrak{A}$  mit  $P(A) = 1$  ( $P(A) = 0$ ) heißt fast sicheres Ereignis (bzw. fast unmögliches Ereignis).*

Bei der rein zufälligen Wahl eines Punktes aus  $[0, 1)$  ist das Ereignis  $A :=$  "Es wird ein irrationaler Punkt gewählt" ein fast sicheres Ereignis und  $\bar{A} =$  "Es

wir ein rationaler Punkt gewählt" ein fast unmögliches Ereignis.

### 3.7 Zufallsgrößen

Unter einer Zufallsgröße versteht man umgangssprachlich eine Größe, die im Rahmen gewisser zufälliger Erscheinungen einen Wert annimmt, der nicht von vornherein feststeht, sondern vom Zufall abhängt. Beispiele findet man überall. In der Natur (Wetter), der Wirtschaft (Aktienkurse), der Technik (Ausfallzeitpunkte von Konsumgütern). Ihre mathematische Erfassung und Untersuchung ist ein zentraler Punkt der Wahrscheinlichkeitstheorie.

Im Allgemeinen sind zufällige Erscheinungen von sehr komplexer Natur. Man denke nur an das Wetter oder das Geschehen an einer Aktienbörse. Durch die Konzentration auf Zufallsgrößen, wie Tageshöchsttemperatur, monatliche Niederschlagsmenge bzw. Aktientagesschlusskurse oder wöchentliche Rendite bestimmter Unternehmen werden Teilaspekte der zugrunde liegenden zufälligen Prozesse herausgestellt, für die man sich besonders interessiert.

Es seien  $(\Omega, \mathfrak{A}, P)$  ein Wahrscheinlichkeitsraum und  $X$  eine Abbildung von  $\Omega$  in eine Menge  $E$ , z.B.  $E = R_1$ ,  $E = R_n$  oder  $E = N_0 = \{0, 1, 2, \dots, n, \dots\}$ . Indem man nicht den Versuchsausgang  $\omega \in \Omega$  zur Kenntnis nimmt, sondern nur den Wert  $X(\omega)$  beobachtet, den die Funktion  $X$  in Abhängigkeit von  $\omega$  annimmt, ist ein neues zufälliges Experiment definiert mit möglichen Ausgängen  $x = X(\omega)$ , die aus  $E$  stammen. Die mit diesem neuen Experiment verbundenen Ereignisse sind nunmehr Teilmengen von  $E$ . Sie bilden eine  $\sigma$ -Algebra  $\mathfrak{E}$  von Teilmengen von  $E$ . Das Ereignis  $B$  aus  $\mathfrak{E}$  tritt für den neuen Versuch offenbar genau dann ein, wenn der ursprüngliche Versuch zu einem  $\omega$  führt, für das  $X(\omega) \in B$  gilt.

**Beispiel 3.21** Wir betrachten das Laplace-Experiment des gleichzeitigen Werfens zweier regulärer Würfel und wählen

$$\Omega = \{\omega = (i, j) : i, j \in \{1, 2, \dots, 6\}\}, \quad \mathfrak{A} = \mathfrak{P}(\Omega),$$

$$P(\{\omega\}) = 36^{-1}, \quad \omega \in \Omega,$$

$$X(\omega) = i + j \quad , \quad \omega = (i, j) \in \Omega.$$

Hier wählt man den Bildraum  $E$  als die Menge  $\{2, 3, \dots, 12\}$  und für  $\mathfrak{E}$  die Potenzmenge  $\mathfrak{P}(E)$ .

Die Funktion  $X$  gibt also die Augensumme der zwei geworfenen Würfel an.

Das Ereignis "Augensumme ist gleich 4" entspricht der Menge  $\{4\}$  aus  $E$  und tritt genau dann ein, wenn ein  $\omega = (i, j)$  mit  $i + j = 4$  Ergebnis des Würfels ist, also wenn  $(1, 3)$ ,  $(2, 2)$  oder  $(3, 1)$  gewürfelt wurde.

Wir kehren zurück zum allgemeinen Fall und wollen auf  $(E, \mathfrak{E})$  eine Wahrscheinlichkeitsverteilung  $P^X$  einführen, die den Ereignissen des neuen Versuches ihre Wahrscheinlichkeiten zuordnet. Das geschieht durch

$$P^X(B) := P(\{\omega \in \Omega | X(\omega) \in B\}) = P(X^{-1}(B)), \quad B \in \mathfrak{E} \quad (3.20)$$

Üblicherweise ist  $\mathfrak{E}$  zusammen mit  $E$  von vornherein festgelegt. Damit die Definition von  $P^X$  dann sinnvoll ist, müssen wir eine Forderung an  $X$  stellen, die wir in der nächsten Definition formulieren.

**Definition 3.22** *Die Abbildung  $X$  von  $(\Omega, \mathfrak{A})$  in  $(E, \mathfrak{E})$  heißt eine Zufallsgröße über  $(\Omega, \mathfrak{A}, P)$  mit Werten in  $(E, \mathfrak{E})$ , falls gilt*

$$X^{-1}(B) := \{\omega \in \Omega | X(\omega) \in B\} \in \mathfrak{A}, \text{ für alle } B \text{ aus } \mathfrak{E}, \quad (3.21)$$

*m.a.W., falls die Abbildung  $X$  (in der Sprache der Maßtheorie) eine  $\mathfrak{A} - \mathfrak{E}$ -messbare Abbildung ist.*

Die Eigenschaft (3.21) kann man kurz schreiben als

$$X^{-1}(\mathfrak{E}) \subseteq \mathfrak{A}. \quad (3.22)$$

**Notation 3.23** Für  $\{\omega \in \Omega | X(\omega) \in B\}$  schreiben wir häufig kürzer  $\{X \in B\}$ , und statt  $P^X(B)$  bzw.  $P(X^{-1}(B))$  verwenden wir die Schreibweise  $P(X \in B)$ .

**Aussage 3.24** *Durch*

$$P^X(B) := P(X \in B), \quad B \in \mathfrak{E} \quad (3.23)$$

ist auf  $\mathfrak{E}$  eine Wahrscheinlichkeitsverteilung  $P^X$  gegeben. Die Verteilung  $P^X$  nennt man die Wahrscheinlichkeitsverteilung der Zufallsgröße  $X$  oder die durch  $X$  induzierte Wahrscheinlichkeitsverteilung.

Zum Beweis prüft man die Axiome A1.-A2.' nach.

Das zufällige Experiment "Beobachtung der Zufallsgröße  $X$ " wird also mathematisch modelliert durch den Wahrscheinlichkeitsraum  $(E, \mathfrak{E}, P^X)$ .

**Beispiel 3.25 (Fortsetzung von Beispiel 3.21)**

Für die Wahrscheinlichkeitsverteilung  $P^X$  der Zufallsgröße  $X$  gelten mit

$$E = \{2, \dots, 12\}, \quad \mathfrak{E} = \mathfrak{P}(E)$$

die Gleichungen

$$P^X(\{k\}) = P(\{\omega \in \Omega | X(\omega) = k\}) = \frac{\#\{\omega = (i, j) | i + j = k\}}{36} = \frac{6 - |7 - k|}{36}, \quad k \in E$$

und

$$P^X(B) = \sum_{k \in B} P^X(\{k\}), \quad B \in \mathfrak{E}. \quad (3.24)$$

Die Forderung  $X^{-1}(\mathfrak{E}) \subseteq \mathfrak{A}$  ist in diesem Fall natürlich erfüllt, da  $\mathfrak{A} = \mathfrak{P}(\{1, 2, \dots, 6\}^2)$ .

Aus vorgegebenen Zufallsgrößen kann man durch eine Vielzahl von Operationen neue Zufallsgrößen bilden. Exemplarisch erwähnen wir hier einige Fälle. Sie sind in den beiden folgenden Aussagen enthalten.

**Aussage 3.26** Ist  $(E, \mathfrak{E}) = (R_n, \mathfrak{B}_n)$ , so sind mit Zufallsgrößen  $X, X_1, X_2, \dots$  über  $(\Omega, \mathfrak{A}, P)$  auch die Vielfachen  $aX, a \in R_1$ , die Summen  $\sum_{k=1}^n X_k$ , die Funktionen  $\max(X_1, X_2, \dots, X_n)$   $\min(X_1, X_2, \dots, X_n)$ , der Grenzwert  $\lim_{n \rightarrow \infty} X_n$  (sofern er existiert) wieder Zufallsgrößen. (Die Operationen verstehen sich dabei punktweise, also  $\omega$ -weise.)

Diese Tatsache ergibt sich sofort aus den entsprechenden Eigenschaften messbarer Funktionen, die in der Maßtheorie bewiesen werden.

**Aussage 3.27** Es seien  $(\Omega, \mathfrak{A}, P)$  ein Wahrscheinlichkeitsraum,  $(E, \mathfrak{E})$  und  $(F, \mathfrak{F})$  messbare Räume.

Ist  $X$  eine Zufallsgröße über  $(\Omega, \mathfrak{A}, P)$  mit Werten in  $(E, \mathfrak{E})$  und ist  $Y$  eine Zufallsgröße über  $(E, \mathfrak{E}, P^X)$  mit Werten in  $(F, \mathfrak{F})$ , so ist  $Z = Y \circ X$  eine Zufallsgröße über  $(\Omega, \mathfrak{A}, P)$  mit Werten in  $(F, \mathfrak{F})$  und ihre Wahrscheinlichkeitsverteilung  $P^Z$  auf  $\mathfrak{F}$  ist gegeben durch

$$\begin{aligned} P^Z(C) &= P(Z \in C) = P(X \in Y^{-1}(C)) = \\ &P^X(Y^{-1}(C)) = P(X^{-1}(Y^{-1}(C))), C \in \mathfrak{F}. \end{aligned} \quad (3.25)$$

Beweis: Nach Definition gilt  $Y^{-1}(\mathfrak{F}) \subseteq \mathfrak{E}$  und  $X^{-1}(\mathfrak{E}) \subseteq \mathfrak{A}$ , folglich ist  $Z^{-1}(\mathfrak{F}) = X^{-1}(Y^{-1}(\mathfrak{F})) \subseteq \mathfrak{A}$ . Also (siehe (3.22)) ist  $Z$  eine Zufallsgröße über  $(\Omega, \mathfrak{A}, P)$  mit Werten in  $(F, \mathfrak{F})$ . Die Formel (3.25) ergibt sich unmittelbar aus der Definition der Verteilung von  $Z$ . Man beachte die Notation 3.23.  $\square$

**Definition 3.28** Jede Zufallsgröße über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  mit Werten in  $(R_1, \mathfrak{B}_1)((R_n, \mathfrak{B}_n))$  heißt eine reellwertige Zufallsgröße (bzw. ein  $n$ -dimensionaler zufälliger Vektor).

$\mathfrak{B}_1$  bzw.  $\mathfrak{B}_n$  bezeichnen dabei die  $\sigma$ -Algebren der Borelmengen aus  $R_1$  bzw.  $R_n$ .

**Beispiel 3.29** In Abschnitt 3.3 haben wir die symmetrische Irrfahrt  $(S_0, S_1, \dots, S_n)$  kennen gelernt. Definiert man die reellwertige Zufallsgröße  $T_{-1}$  durch

$$T_{-1}(\omega) = \min\{K \leq n : S_k(\omega) = -1\}, \omega \in \Omega = \{-1, 1\}^n$$

mit  $\min \emptyset := \infty$ , so hat  $T_{1-}$  die möglichen Werte  $1, 3, 5, \dots, 2\left[\frac{n+1}{2}\right] - 1, \infty$  und es gilt (siehe (3.17))

$$\left. \begin{aligned} P(T_{1-} = 2m + 1) &= q_m, & m = 1, 2, \dots, \left[\frac{n+1}{2}\right] - 1 \\ \text{mit } q_m &= \frac{(2m-1)!!}{(m+1)!2^{m+1}}, \quad m \geq 1, \text{ und} \\ P(T_{1-} = 1) &= \frac{1}{2}, \quad P(T_{1-} = \infty) = \sum_{m \geq \left[\frac{n+1}{2}\right]} q_m. \end{aligned} \right\} \quad (3.26)$$

Die Zufallsgröße  $T_{-1}$  gibt den Zeitpunkt an, zu dem der Spieler  $A$  zum ersten Mal einen negativen Gewinn verbucht, also den Zeitpunkt seines Ruines, falls er kein zusätzliches Kapital besitzt. Das Ereignis  $\{T_{-1} = \infty\}$  tritt ein, falls er nach  $n$ -maligem Werfen der Münze noch nicht "ruiniert" ist.

Wenn er unbegrenzt lange spielt, ergibt sich

$$P(T_{-1} < \infty) = \sum_{m=0}^{\infty} P(T_{-1} = 2m + 1) = \sum_{m=0}^{\infty} q_m = 1.$$

(Den Beweis der letzten Gleichung führen wir später.)

Das bedeutet, bei unbegrenzter Fortführung des Münzwurfes wird der Spieler  $A$  mit Wahrscheinlichkeit Eins irgendwann "ruiniert", d. h. sein Guthaben wird irgendwann negativ.

Völlig analog kann man aber auch schlussfolgern, dass er mit Wahrscheinlichkeit Eins irgendwann mindestens einen Betrag der Größe Eins auf seinem Konto hat. (Wir setzen dabei voraus, dass er in der Zwischenzeit, wenn sein Guthaben im Negativen ist, immer genügend Finanzmittel besitzt, das Spiel fortzusetzen.) Wenn er die Strategie verfolgt, in dem Moment aufzuhören zu spielen, wenn er das erste Mal einen Gesamtgewinn der Höhe Eins hat, so gewinnt er bei dem vereinbarten durchaus fairen Spiel des Münzenwurfes ohne Zeitlimit mit Wahrscheinlichkeit Eins eine Geldeinheit. Ein Paradoxon.

### 3.8 Verteilungsfunktionen

#### Verteilungsfunktionen auf $R_1$

**Definition 3.30** Ist  $Q$  eine Wahrscheinlichkeitsverteilung auf  $R_1, \mathcal{L}_1$ , so bezeichnet man die durch

$$F(x) := Q((-\infty, x]), \quad x \in R_1 \quad (3.27)$$

auf  $R_1$  gegebene Funktion  $F$  als Verteilungsfunktion der Verteilung  $Q$ . Ist  $X$  eine reellwertige Zufallsgröße über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$ , so nennt man die zu  $P^X$  (siehe (3.19) und (3.22)) gehörende Verteilungsfunktion  $F$  die Verteilungsfunktion der Zufallsgröße  $X$ . Gegebenenfalls schreibt man  $F_Q$  bzw.  $F_X$  an Stelle  $F$ .

Es gilt

$$F(b) - F(a) = Q((a, b]), \quad a < b, \quad (3.28)$$

$$F_X(x) = P^X((-\infty, x]) = P(X \leq x) \quad (3.29)$$

Es sei  $Q$  eine Wahrscheinlichkeitsverteilung auf  $(R_1, \mathcal{L}_1)$ .

**Aussage 3.31** Die Verteilungsfunktion  $F$  der Verteilung  $Q$  hat folgende Eigenschaften:

1.  $F$  ist monoton nichtfallend:  $x \leq y \implies F(x) \leq F(y)$ ,
2.  $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow \infty} F(x) = 1$ ,
3.  $F$  ist an jeder Stelle  $x \in R_1$  von rechts stetig:

$$F(x+0) := \lim_{y \downarrow x} F(y) = F(x),$$

4. Für jedes  $x \in R_1$  gilt mit  $F(x-0) := \lim_{y \uparrow x} F(y)$

$$F(x) - F(x-0) = Q\{x\}.$$

Beweis: Unter Verwendung von (3.27), Lemma 3.6 und Folgerung 3.7 haben wir:

1.  $x \leq y \implies (-\infty, x] \subseteq (-\infty, y] \implies F(x) \leq F(y)$ ,
2.  $x_n \downarrow -\infty \implies (-\infty, x_n] \downarrow \emptyset \implies \lim_{n \rightarrow \infty} F(x_n) = 0$ ,  
  
 $x_n \uparrow \infty \implies (-\infty, x_n] \uparrow R_1 \implies \lim_{n \rightarrow \infty} F(x_n) = 1$ ,
3. Wegen  $(-\infty, x] = \bigcap_n (-\infty, x_n]$  für jede Folge  $(x_n)$  mit  $x_n \downarrow x$  folgt  $\implies F(x_n) \downarrow F(x)$ ,
4.  $(x_n, x] \downarrow \{x\}$  für jede Folge  $(x_n)$  mit  $x_n \uparrow x$ .  
 $\implies F(x) - F(x-0) := \lim_{n \rightarrow \infty} (F(x) - F(x_n))$   
 $= \lim_{n \rightarrow \infty} Q((x_n, x]) = Q(\{x\}).$  □

**Definition 3.32** Jede Funktion  $F$  und  $R_1$  mit den Eigenschaften 1. - 3. aus Aussage 3.31 heißt eine Verteilungsfunktion auf  $R_1$ .

Es sei  $F$  eine Verteilungsfunktion auf  $R_1$ , d. h. eine Funktion mit den Eigenschaften 1. - 3. aus Aussage 3.31.

**Aussage 3.33** Es gibt eine eindeutig bestimmte Wahrscheinlichkeitsverteilung  $Q$  auf  $(R_1, \mathcal{B}_1)$ , die  $F$  als Verteilungsfunktion besitzt.

Beweis: Wir setzen

$$Q((a, b]) := F(b) - F(a), \quad a < b.$$

Das Mengensystem  $\gamma = \{(a, b] \mid -\infty < a < b < \infty\}$  ist ein Semiring, und die dadurch definierte nichtnegative Mengenfunktion  $Q$  ist auf  $\gamma$   $\sigma$ -stetig (d. h., ist

$((a_n, b_n])$  eine monoton fallende Folge halboffener Intervalle mit  $\bigcap_{n \geq 1} (a_n, b_n] = \emptyset$ ,

so haben wir  $\lim_{n \rightarrow \infty} Q((a_n, b_n]) = 0$ .

(Der Beweis ist sehr technisch, siehe z. B. Siraeu, II § 3, Punkt 1). Außerdem gilt  $Q(R_1) = 1$ .

In der Maßtheorie wird gezeigt, dass man jedes  $Q$  mit diesen Eigenschaften auf eine und nur eine Weise zu einem Wahrscheinlichkeitsmaß, das wir ebenfalls mit  $Q$  bezeichnen, auf die  $\sigma$ -Algebra  $\mathfrak{B}_1 = \sigma(\gamma)$  der Borelmengen aus  $R_1$  erweitern kann. Wegen

$$Q((-\infty, x]) = \lim_{n \rightarrow \infty} Q((-n, x]) = \lim_{n \rightarrow \infty} (F(x) - F(-n)) = F(x)$$

folgt die Behauptung. □

### Beispiel 3.34

- a) Ist  $(\Omega, \mathfrak{A}, P)$  ein Laplace-Experiment mit  $\Omega = \{w_1, w_2, \dots, w_N\}$ ,  $\mathfrak{A} = \mathfrak{P}(\Omega)$ ,  $P(\{\omega\}) = \frac{1}{N}$ ,  $\omega \in \Omega$ , und ist  $X$  eine reellwertige Zufallsgröße über  $(\Omega, \mathfrak{A}, P)$  mit  $X(\omega_k) = x_k$ ,  $k = 1, \dots, N$ ,  $x_k \neq x_j$  für  $k \neq j$ , so lautet die Verteilungsfunktion  $F = F_X$  wie folgt:

$$F(x) = \frac{1}{N} \sum_{k: x_k \leq x} \mathbb{1} = \frac{1}{N} \sum_{k=1}^N \mathbb{1}_{(-\infty, x_k]}(x), x \in R_1$$

$F$  ist in diesem Fall eine stückweise konstante rechtsseitig stetige, nichtfallende Funktion auf  $R_1$  mit den Sprungstellen  $x_k$ , den Sprunghöhen  $\frac{1}{N}$  und  $F(x) = 0$  für  $x < \min_{k=1, \dots, N} x_k$  und  $F(x) = 1$  für  $x \geq \max_{k=1, \dots, N} x_k$ .

- b) Ist  $(\Omega, \mathfrak{A}, P) = ([0, 1], \mathfrak{B}_{[0,1]}, \lambda_{[0,1]})$  der Wahrscheinlichkeitsraum, der die rein zufällige Wahl eines Punktes  $\omega$  aus  $[0, 1)$  modelliert, und ist  $X(\omega) = \omega$ ,  $\omega \in \Omega$ , so gilt für  $F = F_X$

$$F(x) = (x \wedge 1) \vee 0 = \begin{cases} 0 & x \leq 0 \\ x & x \in [0, 1] \\ 1 & x \geq 1 \end{cases}$$

Eine Möglichkeit, sich von der Lage einer Wahrscheinlichkeitsverteilung auf  $R_1$  und ihre Ausbreitung eine Vorstellung zu verschaffen, besteht in der Berechnung ihrer Quantile.

**Definition 3.35** *Es seien  $Q$  eine Wahrscheinlichkeitsverteilung auf  $(R_1, \mathfrak{B}_1)$  und  $F$  ihre Verteilungsfunktion:*

$$F(x) = Q((-\infty, x]), \quad x \in R_1.$$

Weiterhin sei  $p$  irgendeine Zahl mit  $0 < p < 1$ .

Als  $p$ -Quantil der Verteilung  $Q$  bezeichnet man jede Zahl  $q_p \in R_1$  mit

$$F(q_p - 0) \leq p \leq F(q_p). \quad (3.30)$$

**Aussage 3.36** *Für jedes  $p \in (0, 1)$  ist die Menge aller  $p$ -Quantile von  $Q$  ist nichtleer und bildet ein beschränktes abgeschlossenes Intervall. Sie ist einelementig genau dann, wenn es keine zwei Zahlen  $x < y$  gibt mit  $F(x) = F(y) = p$ .*

**Definition 3.37** *Jedes Quartil  $q_{\frac{1}{2}}$  heißt Median. Jedes Quartil  $q_{\frac{1}{4}}$   $q_{\frac{3}{4}}$  heißt unteres (oberes) Quartil.*

Die Differenz  $q_{\frac{3}{4}} - q_{\frac{1}{4}}$  ist ein Maß für die "Ausbreitung" der Verteilung  $Q$ . Es gilt  $Q([q_{\frac{1}{4}}, q_{\frac{3}{4}}]) = Q((-\infty, q_{\frac{3}{4}}]) - Q((-\infty, q_{\frac{1}{4}}]) \geq \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$ .

Das heißt, zwischen  $q_{\frac{1}{4}}$  und  $q_{\frac{3}{4}}$  befindet sich mindestens die  $Q$ -Wahrscheinlichkeitsmaße"  $\frac{1}{2}$ .

Ein Median ist ein Wert, den man als zentrum der Wahrscheinlichkeitsverteilung  $Q$  bezeichnen kann.

**Aussage 3.38** Ist  $F$  eine streng wachsende (nicht notwendige stetige) Verteilungsfunktion auf  $R_1$ , so existiert zu jedem  $p \in (0, 1)$  das eindeutig bestimmte Quantil  $q_p$  und es gilt

$$q_p = F^{-1}(p) = (p \in (0, 1)) \quad (3.31)$$

wobei  $F^{-1}(p) := \inf\{x | F(x) > p\}$  (rechtsstetige Inverse) gesetzt wird.

## Verteilungsfunktionen auf $R_n$

Es seien  $n \geq 2$  und  $Q$  eine Wahrscheinlichkeitsverteilung auf  $(R_n, \mathcal{B}_n)$ .

**Definition 3.39** Mit der Bezeichnung  $x = (x_1, x_2, \dots, x_n)^T \in R_n$  und  $(-\infty, x] = \prod_{k=1}^n (-\infty, x_k]$  ist durch

$$F(x) = Q((-\infty, x]), \quad x \in R_n$$

eine Funktion auf  $R_n$  definiert, die Verteilungsfunktion der Wahrscheinlichkeitsverteilung  $Q$ .

Ist  $X = (X_1, X_2, \dots, X_n)$  ein zufälliger Vektor über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  mit Werten in  $R_n$ , so nennt man die Verteilungsfunktion  $F$  der Verteilung  $P^X$  auch Verteilungsfunktion von  $X$ . In diesem Fall gilt für alle  $x = (x_1, \dots, x_n)^T \in R_n$  die Beziehung

$$F(x) = P^X((-\infty, x]) = P(X \in (-\infty, x]) =$$

$$P(X_1 \in (-\infty, x_1], \dots, X_n \in (-\infty, x_n]) = P(X_1 \leq x_1, \dots, X_n \leq x_n) \quad (3.32)$$

**Aussage 3.40** Die Verteilungsfunktion  $F$  der Wahrscheinlichkeitsverteilung  $Q$  hat folgende Eigenschaften:

1.  $0 \leq F(x) \leq 1, x = (x_1, x_2, \dots, x_n)^T \in R_n,$
2.  $\lim_{x_k \downarrow -\infty} F(x_1, \dots, x_n) = 0$  für jedes  $k = 1, \dots, n,$

3.  $\lim_{x_1, \dots, x_n \uparrow \infty} F_X(x_1, \dots, x_n) = 1$
4.  $F$  ist an jeder Stelle  $x = (x_1, x_2, \dots, x_n)^T \in R_n$  von rechts stetig:  

$$\lim_{\substack{h_i \downarrow 0 \\ i=1, \dots, n}} F(x_1 + h_1, \dots, x_n + h_n) = F(x_1, \dots, x_n),$$
5. Mit der Definition  

$$\Delta_{h_i} F(x) = F(x_1, \dots, x_{i-1}, x_i + h_i, x_{i+1}, \dots, x_n) - F(x_1, \dots, x_n)$$
 gilt  

$$0 \leq \Delta_{h_1} \dots \Delta_{h_n} F(x) \in R_n, h_i \geq 0, i_1, \dots, n$$
  
 (Verallgemeinerung der Monotonie im Fall  $n = 1$ ).

**Bemerkung 3.41** Für  $n = 2$  lautet die Eigenschaft 5. wie folgt:

$$F_X(x_1 + h_1, x_2 + h_2) - F_X(x_1, x_2 + h_2) - F_X(x_1 + h_1, x_2) + F_X(x_1, x_2) \geq 0. \quad (3.33)$$

**Definition 3.42** Jede Funktion  $F$  auf  $R_n$  mit den Eigenschaften 1.-5. aus Aussage 3.40 nennen wir eine Verteilungsfunktion auf  $R_n$ .

Zu jeder Verteilungsfunktion  $F$  auf  $R_n$  definieren wir für alle  $a = (a_1, a_2, \dots, a_n)^T \in R_n, h_1, h_2, \dots, h_n > 0$  durch

$$Q\left(\prod_{k=1}^{n \otimes} (a_k, a_k + h_k]\right) = \Delta_{h_1} \Delta_{h_2} \dots \Delta_{h_n} F(a) \quad (3.34)$$

eine Mengenfunktion  $Q$  auf dem Semiring  $\gamma_n$  aller  $n$ -dimensionalen "Quader"  
 $\prod_{k=1}^{n \otimes} (a_k, b_k] \subseteq R_n$ .

**Aussage 3.43** Die durch (3.34) definierte Mengenfunktion  $Q$  ist auf  $\gamma_n$   $\sigma$ -additiv und lässt sich auf eine und nur eine Weise zu einer Wahrscheinlichkeitsmauverteiling auf  $\mathcal{B}_n$  fortsetzen, das wir wiederum mit  $Q$  bezeichnen. Die Verteilung  $Q$  besitzt  $F$  als Verteilungsfunktion.

Zum Beweis, der rein maßtheoretischer Natur ist, sei auch hier auf Siraev (...), II, §3, verwiesen.

**Beispiel 3.44**

a) Die Funktion  $F$ , definiert durch

$$F(x_1, x_2) = [(x_1 \wedge x_2) \wedge 1] \vee 0, \quad (x_1, x_2)^T \in R_2, \quad (3.35)$$

ist eine Verteilungsfunktion auf  $R_2$ .

Beweis: Die Eigenschaften 1.-4. aus Aussage 3.40 sind offensichtlich. Zum Nachweis von 5. bemerken wir zunächst, dass für jedes Rechteck  $(x_1, x_1 + h_1] \times (x_2, x_2 + h_2]$ , das disjunkt zu  $\{(x, x) : 0 < x \leq 1\}$  ist, gilt  $x_2 \geq x_1 + h_1$  oder  $x_1 \geq x_2 + h_2$ . Daraus folgt für diese Rechtecke

$$\Delta_{h_1} \Delta_{h_2} F(x_1, x_2) = 0.$$

Andererseits ist für jedes Rechteck  $(x, x + h] \times (x, x + h]$

$$\Delta_h \Delta_h F(x, x) = (x + h - x - x + x) = h > 0$$

Mit Hilfe der Additivität von  $Q$  auf  $\gamma_2$  ergibt sich 5. □

b) Sind  $F_k, k = 1, 2, \dots, n$  Verteilungsfunktionen auf  $R_1$ , so ist  $F$  definiert auf  $R_n$  durch

$$F(x) = \prod_{k=1}^n F_k(x_k), \quad x = (x_1, x_2, \dots, x_n)^T \in R_n$$

eine Verteilungsfunktion auf  $R_n$ .

Es seien  $Q$  eine Wahrscheinlichkeitsverteilung auf  $(R_n, \mathcal{B}_n)$  und  $F$  ihre Verteilungsfunktion.

**Definition 3.45** Für jede  $r$ -elementige Teilmenge  $\mathcal{J}_r = \{k_1, k_2, \dots, k_r\}$  von  $\{1, 2, \dots, n\}$  bezeichne  $\prod_{\mathcal{J}_r}$  den Projektionsoperator, definiert durch

$$\prod_{\mathcal{J}_r} x = (x_{k_1}, x_{k_2}, \dots, x_{k_r})^T \in R_r, \quad x = (x_1, \dots, x_n)^T \in R_n.$$

Offenbar ist  $\Pi_{\mathcal{J}_r}$  eine  $\mathcal{B}_n - \mathcal{B}_r$ -meßbare Abbildung.

Die durch

$$Q_{\mathcal{J}_r}(B) := Q\left(\prod_{\mathcal{J}_r}^{-1}(B)\right), \quad B \in \mathcal{L}_r \quad (3.36)$$

definierte Mengenfunktion  $Q_{\mathcal{J}_r}$  auf  $\mathcal{B}_r$  ist eine Wahrscheinlichkeitsverteilung und heißt die zu  $\mathcal{J}_r$  gehörende  $r$ -dimensionale Randverteilung von  $Q$ .

**Aussage 3.46** Die Verteilungsfunktion  $F_{\mathcal{J}_r}$  der Verteilung  $Q_{\mathcal{J}_r}$  hängt mit der Verteilungsfunktion  $F$  wie folgt zusammen:

$$F_{\mathcal{J}_r}(x_{k_1}, x_{k_2}, \dots, x_{k_r}) = F(\infty, \dots, \infty, x_{k_1}, \infty, \dots, x_{k_2}, \dots, \infty, x_{k_r}, \infty, \dots, \infty) \quad (3.37)$$

Beweis:

$$\begin{aligned} F_{\mathcal{J}_r}(x_{k_1}, x_{k_2}, \dots, x_{k_r}) &= Q_{\mathcal{J}_r}\left(\prod_{l=1}^r (-\infty, x_{k_l}]\right) = \\ &= Q\left(\prod_{\mathcal{J}_r}^{-1}\left(\prod_{l=1}^r (-\infty, x_{k_l}]\right)\right) = Q\left(\prod_{m=1}^n B_m\right) \text{ mit} \\ B_m &= (-\infty, x_{k_l}], \text{ falls } m = k_l \text{ für ein } l = 1, 2, \dots, r, \\ B_m &= (-\infty, \infty) \text{ falls } m \neq k_l \text{ für alle } l = 1, 2, \dots, r. \end{aligned}$$

□

Aus der Kenntnis der Randverteilungsfunktionen  $F_{\mathcal{J}_r}$  mit  $r < n$  kann die Verteilungsfunktion  $F$  selbst i.a. nicht rekonstruiert werden.

Zum Beispiel haben die beiden Verteilungsfunktionen

$$\begin{aligned} G(x_1, x_2) &= ((x_1 \wedge x_2) \wedge 1) \vee 0 \text{ und} \\ H(x_1, x_2) &= [(x_1 \wedge 1) \vee 0][(x_2 \wedge 1) \vee 0] \end{aligned}$$

(siehe Beispiel 3.42)

die gleichen Randverteilungsfunktionen:

$$G(x_1, \infty) = (x_1 \wedge 1) \vee 0 = H(x_1, \infty)$$

$$G(\infty, x_2) = (x_2 \wedge 1) \vee 0 = H(\infty, x_2).$$

Ist  $X = (X_1, \dots, X_n)^T$  ein  $n$ -dimensionaler zufälliger Vektor mit der Verteilungsfunktion  $F$ , so ist  $F_{\mathcal{J}_r}$  die Verteilungsfunktion des Vektors  $(X_{k_1}, \dots, X_{k_r})^T$  wobei  $\mathcal{J}_r = \{k_1, \dots, k_r\}$  gilt. Das ergibt sich einfach aus (3.37) und

$$F(\infty, \dots, \infty, x_{k_1}, \infty, \dots, x_{k_2}, \dots, x_{k_r}, \infty, \dots, \infty) =$$

$$P(X_1 < \infty, \dots, X_{k_1} \leq x_{k_1}, \dots, X_{k_r} \leq x_{k_r+1} < \infty, \dots, X_n < \infty) =$$

$$P(X_{k_1} \leq x_{k_1}, \dots, X_{k_n} \leq x_{k_n}).$$

### 3.9 Verteilungsdichten

Wir haben gesehen, dass sich Wahrscheinlichkeitsverteilungen auf  $(R_1, \mathfrak{B}_1)$  bzw.  $(R_n, \mathfrak{B}_n)$  durch ihre Verteilungsfunktionen charakterisieren lassen. Diese wiederum besitzen in Spezialfällen eine besonders einfache Struktur. Zum einen handelt es sich dabei um sogenannte *diskrete Verteilungen* bzw. *diskret verteilte Zufallsgrößen*. Diesen Verteilungen ist das Kapitel 4 gewidmet. Zum anderen heben sich Verteilungen mit einer *Verteilungsdichte*, man spricht auch einfach von *Dichten*, heraus. Diese Verteilungen werden wir in voller Allgemeinheit unter Verwendung des Begriffs des Lebesgueintegrals erst in Kapitel 7 behandeln. Vorab wollen wir jedoch einige wichtige Fälle, in denen man mit dem bereits bekannten Riemannintegral auskommt, vorstellen.

#### Verteilungsdichten auf $R_1$

Es sei  $F$  eine Verteilungsfunktion auf  $R_1$  (vgl. Definition 3.32).

**Definition 3.47** *Gibt es eine stückweise stetige Funktion  $f$  auf  $R_1$  mit*

$$1. \quad f(x) \geq 0, \quad x \in R_1, \quad (3.38)$$

$$2. \quad \int_a^b f(x)dx = F(b) - F(a), \quad a, b \in R_1, \quad (3.39)$$

so heißt  $f$  eine Dichte der Verteilungsfunktion  $F$ .

Ist  $X$  eine reellwertige Zufallsgröße und  $f_X$  eine Dichte ihrer Verteilungsfunktion  $F_X$ , so heißt  $f_X$  auch Dichte der Zufallsgröße  $X$ .

### Aussage 3.48

a) Für jede Dichte  $f$  gilt

$$\int_{-\infty}^x f(s)ds = F(x) \quad (3.40)$$

$$\int_{-\infty}^{\infty} f(x)dx = 1, \quad (3.41)$$

b) besitzt  $F$  eine Dichte  $f$ , so ist  $F$  stetig,

c) besitzt  $F$  eine Dichte  $f$ , die stetig in einer Umgebung von  $x$  ist, so ist  $F$  differenzierbar in diesem  $x_1$ , und es gilt

$$\frac{d}{dx}F(x) = f(x) \quad (3.42)$$

Beweis:

a) Folgt aus (3.39) für  $a \rightarrow -\infty, b = x$  bzw.  $a \rightarrow -\infty$  und  $b \rightarrow \infty$  sowie der Eigenschaft 2. aus Aussage 3.31.

$$b) \quad F(x+h) - F(x) = \int_x^{x+h} f(s)ds \xrightarrow{h \rightarrow 0} 0,$$

- c) nach Voraussetzung gilt es ein  $\delta > 0$ , so dass  $f$  stetig ist in  $(x - \delta, x + \delta)$ . Für jedes  $h$  mit  $|h| < \delta$  gilt

$$F(x+h) - F(x) = \int_x^{x+h} f(s) ds = h \cdot f(\xi)$$

für ein  $\xi$  zwischen  $x$  und  $x+h$ . Daraus und aus der Stetigkeit von  $f$  in  $(x - \delta, x + \delta)$  folgt c).  $\square$

### Beispiel 3.49

- a) Die Verteilungsfunktion  $F(x) = (x \wedge 1) \vee 0$ ,  $x \in R_1$ , (siehe Beispiel 3.33b)) besitzt die Dichte

$$f(x) = \mathbb{1}_{[0,1]}(x), \quad x \in R_1$$

Die Verteilungsfunktion  $F(x)$  aus Beispiel 3.33a) besitzt keine Dichte.

- b) Es sei  $\lambda > 0$  und  $f_\lambda$  die durch

$$f_\lambda(x) = \begin{cases} 0 & , x < 0 \\ \lambda e^{-\lambda x} & , x \geq 0 \end{cases} = \mathbb{1}_{[0,\infty)}(x) \lambda e^{-\lambda x}, \quad x \in R_1$$

definierte Funktion. Dann ist  $f_\lambda$  die Dichte einer Wahrscheinlichkeitsverteilung auf  $(R_1, \mathfrak{B}_1)$ , die man als *Exponentialverteilung*  $EXP(\lambda)$  mit dem Parameter  $\lambda$  bezeichnet. Ihre Verteilungsfunktion  $F_\lambda$  lautet

$$F_\lambda(x) = \begin{cases} 0 & , x \leq 0, \\ 1 - e^{-\lambda x} & , x > 0. \end{cases}$$

- c) Die *Normalverteilung*  $N(\mu, \sigma^2)$  ist für jedes  $\mu \in R_1$  und jedes  $\sigma^2 > 0$  definiert als die Verteilung mit der Dichte

$$\varphi_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2\sigma^2} (x - \mu)^2 \right], \quad x \in R_1. \quad (3.43)$$

Die zugehörige Verteilungsfunktion

$$\Phi_{\mu,\sigma^2}(x) = \int_{-\infty}^x \varphi_{\mu,\sigma^2}(u) du, \quad x \in \mathbb{R}_1,$$

ist nicht explizit durch elementare Funktionen ausdrückbar.

Die Verteilung  $N(0, 1)$  heißt *Standardnormalverteilung*. Die Werte ihrer Verteilungsfunktion  $\Phi_{0,1}$  sind vertafelt. Es bestehen folgende Beziehungen ( $\sigma = \sqrt{\sigma^2} > 0$ )

$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sigma} \varphi_{0,1}\left(\frac{x - \mu}{\sigma}\right), \quad \Phi_{\mu,\sigma^2}(x) = \Phi_{0,1}\left(\frac{x - \mu}{\sigma}\right) \quad (3.44)$$

Statt  $\varphi_{0,1}$  und  $\Phi_{0,1}$  schreiben wir auch einfach  $\varphi$  bzw.  $\Phi$ , falls keine Verwechslungen möglich sind.

**Aussage 3.50** *Die Dichte  $\varphi$  und die Verteilungsfunktion  $\Phi$  der  $N(0, 1)$ -Verteilung besitzen folgende Eigenschaften*

1.  $\varphi$  ist bezüglich Null symmetrisch:  $\varphi(-x) = \varphi(x), x \in \mathbb{R}_1$ ,
2.  $\varphi$  ist unimodal und hat ihr Maximum bei Null,
3.  $\varphi$  hat zwei Wendepunkte, und zwar bei  $+1$  und  $-1$ ,
4.  $1 - \Phi(x) = \Phi(-x), \quad x \in \mathbb{R}_1, \quad \Phi(0) = 0,5$ ,
5.  $1 - \Phi(x) \leq \frac{1}{x\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (x > 0)$ .

Beweis: Die Eigenschaften 1. - 4. sind offensichtlich, 5. folgt aus

$$1 - \Phi(x) \leq \frac{1}{x\sqrt{2\pi}} \int_x^{\infty} se^{-\frac{s^2}{2}} ds.$$

□

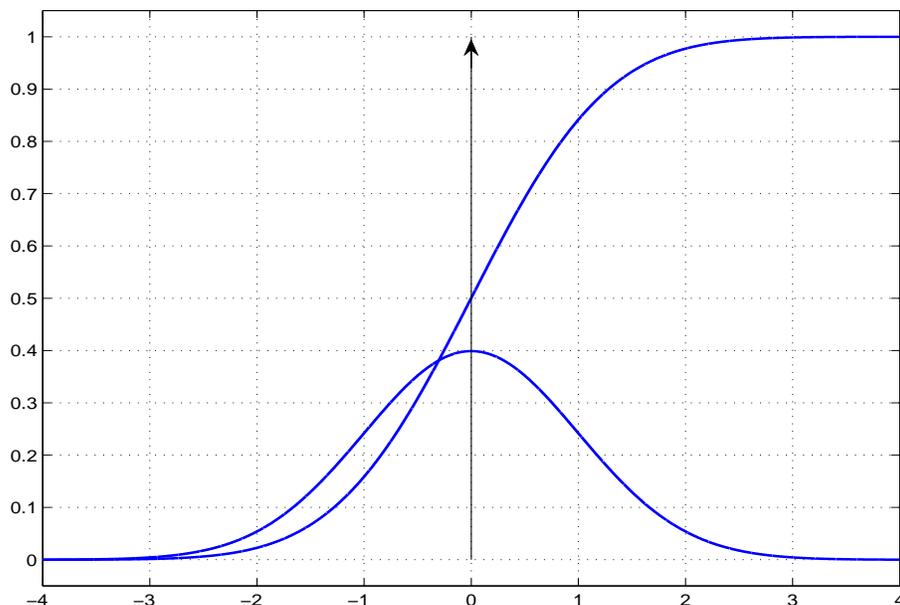


Abbildung 3.5: Verteilungsfunktion und Dichte der Standardnormalverteilung

Bei Anwendungen der Normalverteilung werden häufig ihre Quantile benötigt. Die wichtigsten sind in folgender Tafel zusammengefasst. Sie sind definiert als Lösung der Gleichung  $\Phi(q_p^{0,1}) = p$ :

$p$	0,5	0,9	0,95	0,975	0,99	0,995	(3.45)
$q_p^{(0,1)}$	0	1,282	1,645	1,960	2,326	2,576	

Tafel 1 Quantile der  $N(0, 1)$ -Verteilung

Die Quantile  $q_p^{\mu, \sigma^2}$  der  $N(\mu, \sigma^2)$ -Verteilung sind definiert durch  $\Phi_{\mu, \sigma^2}(q_p^{\mu, \sigma^2}) = p$  und berechnen sich aus den  $q_p^{0,1}$  wie folgt:

$$q_p^{\mu, \sigma^2} = \mu + \sigma q_p^{0,1}, q_{1-p}^{\mu, \sigma^2} = \mu - \sigma q_p^{0,1}. \tag{3.46}$$

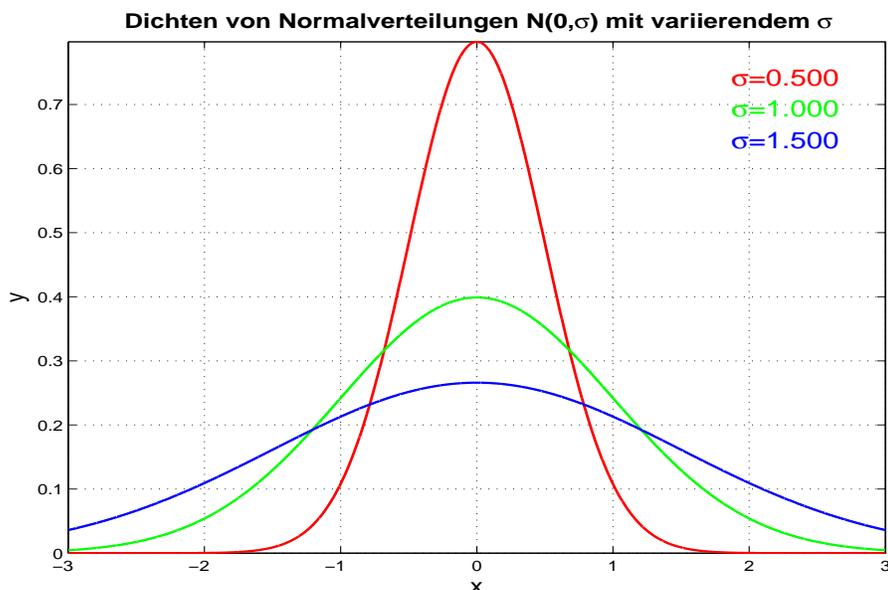


Abbildung 3.6: Normalverteilungsdichten mit verschiedenen Streuungen

**Aussage 3.51** *Es sei  $X$  eine Zufallsgröße mit einer Dichte  $f_X$ , für die  $\{x : f_X(x) > 0\}$  ein Intervall  $(a, b)$  mit  $-\infty \leq a < b \leq \infty$  bilde. Weiterhin sei  $\psi$  eine streng monotone stetig differenzierbare Funktion auf  $(a, b)$ . Dann besitzt  $Y = \psi(X)$  eine Dichte  $f_Y$ , die gegeben ist durch*

$$f_Y(y) = f_X(\psi^{-1}(y)) \cdot \left| \frac{d\psi^{-1}}{dy} \right|(y), \quad y \in Wb(\psi)$$

$$= 0 \quad y \notin Wb(\psi)$$

Beweis: Zunächst sei  $\psi$  streng wachsend. Nach Voraussetzung bildet  $\psi$  das Intervall  $(a, b)$  eineindeutig auf ein Intervall  $(c, d)$  ab. Für  $y \in (c, d)$  gilt

$$F_Y(y) = P(Y \leq y) = P(X \leq \psi^{-1}(y)) = F_X(\psi^{-1}(y)) = \int_a^{\psi^{-1}(y)} f_X(s) ds$$

und somit

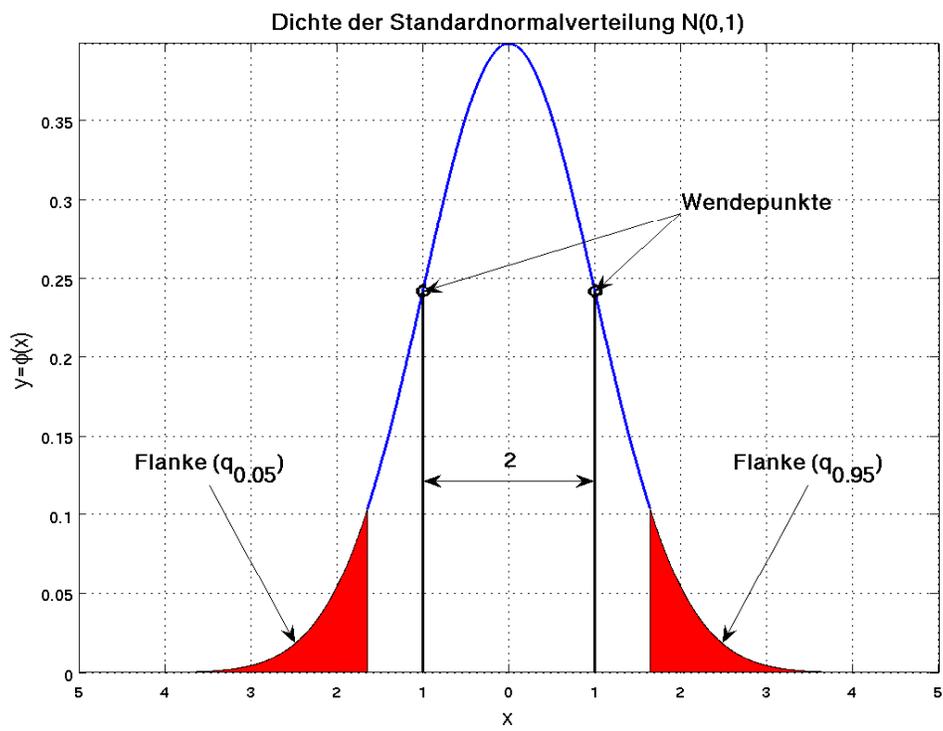


Abbildung 3.7: Quantile der Standardnormalverteilung

$$F_Y(y) = \int_c^y f_X(\psi^{-1}(t)) \frac{d\psi^{-1}}{dt}(t) dt.$$

Wir setzen

$$f_Y(y) = f_X(\psi^{-1}(y)) \frac{d\psi^{-1}}{dy}(y) \text{ für } y \in (c, d).$$

Ist  $y \notin (c, d)$ , so gilt

$$F_Y(y) = P(Y \leq y) \equiv 0 \text{ für } y < c \text{ bzw. für } y \geq d \text{ ist } F_Y(y) = P(Y \geq y) \equiv 1.$$

Somit haben wir  $f_Y(y) = 0$  für solche  $y$ . Insgesamt ergibt sich damit  $F_Y(y) = \int_{-\infty}^y f_Y(t) dt$ ,  $y \in R_1$ . Ist dagegen  $\psi$  streng fallend, so haben wir für  $y \in (c, d)$

$$\begin{aligned} F_Y(y) &= P(Y \geq y) = P(X \geq \psi^{-1}(y)) = 1 - P(X < \psi^{-1}(y)) = \\ &= 1 - \int_a^{\psi^{-1}(y)} f_X(s) ds = \int_{\psi^{-1}(y)}^b f_X(s) ds = - \int_c^y f_X(\psi^{-1}(t)) \frac{d\psi^{-1}}{dt} dt. \end{aligned}$$

Für  $y \notin (c, d)$  schließen wir wie im Fall wachsender Funktionen  $\psi$ . Analog wie oben setzen wir

$$f_Y(y) = f_X(\psi^{-1}(y)) \left| \frac{d\psi^{-1}}{dy}(y) \right|, \quad y \in (c, d), = 0, \quad y \notin (c, d).$$

Damit ist die Aussage bewiesen.  $\square$

**Beispiel 3.52**  $X$  sei gleichmäßig auf  $[0, 1)$  verteilt, es sei  $\lambda > 0$ , und es gelte  $\psi(x) = -\frac{1}{\lambda} \ln x$ ,  $x \in [0, 1)$ , dann hat  $Y = -\frac{1}{\lambda} \ln X$  die Dichte

$$f(y) = \mathbb{1}_{(0, \infty)}(y) \lambda e^{-\lambda y}, \quad y \in R_1$$

.

## Verteilungsdichten auf $R_n$

Es sei  $F$  eine Verteilungsfunktion auf  $R_n$  ( $n \geq 2$ ), siehe Definition 3.42 und Aussage 3.40.

**Definition 3.53** *Gibt es eine reellwertige Riemann-integrierbare Funktion  $f$  auf  $R_n$  mit*

$$1. \quad f(x) \geq 0, x \in R_n, \quad (3.47)$$

$$2. \quad F(x) = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f(s_1, s_2, \dots, s_n) ds_1 \dots ds_n, x = (x_1, \dots, x_n)^T, \quad (3.48)$$

so heißt  $f$  eine Dichte der Verteilungsfunktion  $F$ .

**Aussage 3.54** *Ist  $Q$  die durch  $F$  erzeugte Wahrscheinlichkeitsverteilung, so gilt*

$$Q\left(\prod_{k=1}^n \otimes (x_k, x_k + h_k]\right) = \Delta_{h_1} \dots, \Delta_{h_n} F(x_1, x_2, \dots, x_n) = \quad (3.49)$$

$$\int_{x_n}^{x_n+h_n} \dots \int_{x_1}^{x_1+h_1} f(s_1, \dots, s_n) ds_1 \dots ds_n$$

Beweis: Der Beweis folgt aus der Additivität des Integrals.  $\square$

**Beispiel 3.55** (Fortsetzung von Beispiel 3.44)

- a)  $F$  hat keine Dichte, die Verteilung  $Q_F$  ist auf  $\{(x, x) : 0 \leq x \leq 1\}$  konzentriert.
- b) Haben die Verteilungsfunktionen  $F_k$  die Dichten  $f_k, k = 1, \dots, n$  so besitzt  $F$  eine Dichte  $f$  mit

$$f(x_1, x_2, \dots, x_n) = \prod_{k=1}^n f_k(x_k), \quad x = (x_1, \dots, x_n)^T \in R_n. \quad (3.50)$$

**Aussage 3.56**

a) Für jede Dichte  $f$  von  $F$  gilt

$$F(x) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(s_1, \dots, s_n) ds_1 \dots ds_n, \quad (3.51)$$

$$x = (x_1, \dots, x_n)^T \in \mathbb{R}_n,$$

$$1 = F(\infty, \infty, \dots, \infty) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(s_1, \dots, s_n) ds_1 \dots ds_n \quad (3.52)$$

b) besitzt  $F$  eine Dichte, so ist  $F$  eine stetige Funktion,

c) Besitzt  $F$  eine Dichte  $f$ , die stetig in Umgebung von  $x \in \mathbb{R}_n$  ist, so ist  $F$   $n$ -mal differenzierbar in diesem  $x = (x_1, \dots, x_n)^T$ , und es gilt

$$\frac{\partial^n}{\partial x_1 \dots \partial x_n} F(x_1, \dots, x_n) = f(x_1, \dots, x_n). \quad (3.53)$$

Der Beweis erfolgt analog zum Beweis von Aussage 3.48.

**Aussage 3.57** Besitzt  $F$  eine Dichte  $f$ , so hat auch jede Randverteilungsfunktion  $F_{\mathcal{J}_r}$  mit  $\mathcal{J}_r = \{k_1, \dots, k_r\} \subseteq \{1, 2, \dots, n\}$  eine Dichte  $f_{\mathcal{J}_r}$ , die sich folgendermaßen berechnen lässt:

$$f_{\mathcal{J}_r}(x_{k_1}, \dots, x_{k_r}) = \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{(n-r)\text{-mal}} f(s_1, \dots, s_{k_1-1}, x_{k_1}, \dots, x_{k_r}, s_{k_1+1}, \dots, s_n) ds_1 \dots ds_n$$

Beweis: Der Beweis ergibt sich aus (3.37) und der Definition 3.53 der Dichte  $f$  durch Umordnung der Reihenfolge der entsprechenden  $n$ -fachen Integrale.  $\square$

**Beispiel 3.58**

a) Es sei

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \text{ mit } \sigma_1, \sigma_2 > 0, |\rho| < 1, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in R_2$$

Dann ist die Funktion  $\psi_{\mu, \Sigma}$  definiert durch

$$\psi_{\mu, \Sigma}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

$$\exp \left[ -\frac{1}{2(1-\rho^2)} \left( \left( \frac{x_1 - \mu_1}{\sigma_1} \right)^2 - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right) \right], (x_1, x_2)^T \in R_2$$

die Dichte einer Wahrscheinlichkeitsverteilung auf  $(R_1, \mathcal{B}_1)$  die als Normalverteilung  $N_2(\mu, \Sigma)$  bezeichnet wird.

Die Randverteilungsdichten der Verteilung  $N_2(\mu, \Sigma)$  sind eine  $N(\mu_1, \sigma_1^2)$ - bzw. eine  $N(\mu_2, \sigma_2^2)$ -Verteilung. Man beachte, dass in den Randverteilungen der Parameter  $\rho$  nicht mehr auftritt.

b) Es seien  $\mu \in R_n$  und  $\Sigma$  eine positiv definite symmetrische  $n \times n$ -Matrix. Dann ist die Funktion  $\varphi_{\mu, \Sigma}$ , definiert durch

$$\varphi_{\mu, \Sigma}(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \Sigma}} \exp \left[ -\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu) \right], x \in R_n,$$

die Dichte der sogenannten  $n$ -dimensionalen Normalverteilung  $N_n(\mu, \Sigma)$ .

Zu jeder Teilmenge  $\mathcal{J}_r$  von  $\{1, 2, \dots, n\}$  mit  $\mathcal{J}_r = (k_1, \dots, k_r)$  ist die zu  $\mathcal{J}_r$  gehörende Randverteilung ebenfalls eine Normalverteilung und zwar gleich  $N_r(\Pi_{\mathcal{J}_r} \mu, \Pi_{\mathcal{J}_r} \Sigma \Pi_{\mathcal{J}_r}^T)$  wobei  $\Pi_{\mathcal{J}_r}$  die Projektionsmatrix ist, die

$x = (x_1, x_2, \dots, x_n)^T$  auf  $\Pi_{\mathcal{J}_r} x_n = (x_{k_1}, \dots, x_{k_r})$  abbildet.

Der Vektor  $\Pi_{\mathcal{J}_r} \mu$  ergibt sich aus  $\mu$  durch Entfernen aller Komponenten  $x_l$  mit  $l \notin \mathcal{J}_r$  und die Matrix  $\Pi_{\mathcal{J}_r} \sum \Pi_{\mathcal{J}_r}^T$  ergibt sich aus  $\sum = (s_{ij})$  durch Entfernen aller Elemente  $s_{ij}$  mit  $i \notin \mathcal{J}_r$  oder  $j \notin \mathcal{J}_r$ .

**Aussage 3.59 (Transformationsformel für  $n$ -dimensionale Dichten)**

Es sei  $X = (X_1, \dots, X_n)^T$  ein zufälliger  $n$ -dimensionaler Vektor mit der Dichte  $f$ .

Weiterhin sei  $U$  eine offene Menge aus  $R_n$  mit  $P^X(U) = 1$  und  $h$  eine eindeutige stetig differenzierbare Funktion von  $U$  auf  $V \subseteq R_n$ , deren Jacobimatrix

$$\mathcal{J}_h(x) := \left( \frac{\delta h_i(x)}{\delta x_j} \right)_{i,j=1,\dots,n}$$

nirgends auf  $U$  singulär ist. Mit  $g$  werde die inverse Funktion  $h^{-1}$  bezeichnet.

Dann hat der zufällige Vektor  $Y := h(X)$  eine Dichte  $f_Y$  mit

$$f_Y(y) = \begin{cases} f_X(g(y)) |\det \mathcal{J}_g(y)|, & \text{falls } y \in V \\ 0, & \text{falls } y \in R_n \setminus V. \end{cases}$$

Bemerkung: Die soeben formulierte Aussage findet man in der Literatur in unterschiedlicher Form, je nachdem, welche Voraussetzungen man an  $g$  stellt. Siehe zum Beispiel Pfanzagl, 1991, Kap. 3.4.

**Beispiel 3.60** Es seien  $A$  eine reguläre  $n \times n$ -Matrix und  $b \in R_n$ . Wir definieren

$$g(x) = Ax + b, \quad x \in R_n.$$

$$Y := g(X).$$

Dann gilt  $g^{-1}(y) = A^{-1}(y - b)$ ,  $\mathfrak{J}_{g^{-1}}(y) = A^{-1}$  und  $Y$  hat die Dichte

$$f_Y(y) = f_X(A^{-1}(y - b)) |\det A^{-1}|, \quad y \in R_n.$$

# Kapitel 4

## Diskrete Verteilungen und Zufallsgrößen

Wahrscheinlichkeitsverteilungen und Zufallsgrößen haben wir in dem sehr allgemeinen Rahmen von Wahrscheinlichkeitsräumen  $(\Omega, \mathfrak{A}, P)$  eingeführt. In dieser Allgemeinheit, die den Vorteil der begrifflichen Klarheit, Übersichtlichkeit und der Spezialisierungsmöglichkeit hat, ist jedoch eine detaillierte Untersuchung bzw. Ausgestaltung der mit ihnen zusammenhängenden Begriffe anspruchsvoll und bedarf der Kenntnis der Maßtheorie. Für viele Anwendungen ist diese Allgemeinheit aber nicht notwendig. Wir stellen sie also zunächst zurück und schränken uns in diesem Kapitel auf den Spezialfall diskreter Wahrscheinlichkeitsverteilungen ein.

In diesem Fall tritt die Maßtheorie in den Hintergrund, da man es im Grunde stets mit höchstens abzählbar unendlich vielen Versuchsausgängen bzw. möglichen Werten (bei Zufallsgrößen) zu tun hat und deshalb der Verwendung der Potenzmenge als relevante  $\sigma$ -Algebra von Teilmengen nichts im Wege steht.

Diskrete Verteilungen sind, grob gesprochen, solche, bei denen die "Wahrscheinlichkeitsmasse" in höchstens abzählbar vielen Punkten konzentriert ist.

### 4.1 Definitionen und Beispiele

Es seien  $(\Omega, \mathfrak{A}, P)$  ein Wahrscheinlichkeitsraum und  $I$  eine Teilmenge der Menge  $N_0$  aller natürlichen Zahlen.

**Definition 4.1** Die Wahrscheinlichkeitsverteilung  $P$  heißt eine diskrete Ver-

teilung, falls es eine höchstens abzählbare Menge  $\Omega_P := \{\omega_i : i \in I\}$  aus  $\Omega$  gibt mit  $\{\omega_i\} \in \mathfrak{A}$ ,  $i \in I$ , und  $P(\Omega \setminus \Omega_P) = 0$ .

Insbesondere ist jede Wahrscheinlichkeitsverteilung  $P$  auf  $(\Omega, \mathfrak{A})$  diskret, falls  $\Omega$  selbst höchstens abzählbar unendlich ist.

**Folgerungen 4.2** Mit der Bezeichnung  $p_i := P(\{\omega_i\})$ ,  $i \in I$ , gilt

1.

$$\sum_{i \in I} p_i = 1 \quad (4.1)$$

2. Für alle  $A$  aus  $\mathfrak{A}$  ist

$$P(A) = P(A \cap \Omega_P) = P(\{\omega_i \in \Omega_P : \omega_i \in A\}) = \sum_{i: \omega_i \in A} p_i. \quad (4.2)$$

Das bedeutet, jede diskrete Wahrscheinlichkeitsverteilung  $P$  ist durch Angabe der Paare  $(\omega_i, p_i)_{i \in I}$  eindeutig bestimmt.

Aus diesem Grund wird häufig die Folge  $((\omega_i, p_i), i \in I)$  bereits als diskrete Wahrscheinlichkeitsverteilung auf  $\Omega_P = \{\omega_i : i \in I\}$  bezeichnet. Die Zahlen  $p_i$  heißen *Einzelwahrscheinlichkeiten* der Verteilung  $P$ .

3. o.B.d.A. kann man  $p_i > 0, i \in I$ , annehmen. Gilt nämlich  $p_i = 0$  für ein  $i \in I$ , so entfernt man dieses  $\omega_i$  aus  $\Omega_P$ . Die Menge  $\Omega_P^{\min} := \{\omega_i \mid i \in I, p_i > 0\}$  heißt *Träger* der diskreten Verteilung  $P$ .

Die Formel (4.2) kann man nutzen, um  $P$  für jede Teilmenge  $A$  von  $\Omega$  zu definieren, nicht nur für  $A \in \mathfrak{A}$ . Bei diskreten Verteilungen  $P$  ist also immer eine Erweiterung von  $\mathfrak{A}$  auf  $\mathfrak{P}(\Omega)$  möglich. Wir setzen in Zukunft deshalb bei diskreten Verteilungen stets voraus, dass  $\mathfrak{A} = \mathfrak{P}(\Omega)$  gilt.

### Beispiel 4.3

a) Gibt es Elemente  $\omega_1, \dots, \omega_N$  mit  $P(\{\omega_k\}) = p_k = \frac{1}{N}$ , so spricht man von der "Gleichmäßigen diskrete Verteilung auf  $\{\omega_1, \dots, \omega_N\}$ ."

b) Gibt es ein  $\omega_0 \in \Omega$  mit  $P(\{\omega_0\}) = 1$ , so heißt  $P$  die "ausgeartete Verteilung, konzentriert in  $\omega_0$ " oder die in  $\omega_0$  konzentrierte Einpunktverteilung.

c) Die Binomialverteilung

Es seien  $n \in N_1 = \{1, 2, \dots, m, \dots\}$  und  $p \in (0, 1)$ . Durch

$$b(n, p; k) := \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \dots, n\}$$

sind die Einzelwahrscheinlichkeiten einer Wahrscheinlichkeitsverteilung auf  $\{0, 1, \dots, n\}$  gegeben. Diese Verteilung heißt Binomialverteilung mit den Parametern  $n$  und  $p$ .

d) Die Poissonverteilung

Es sei  $\lambda > 0$ . Durch

$$p_k(\lambda) := \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \geq 0$$

sind die Einzelwahrscheinlichkeiten einer diskreten Verteilung auf  $N_0 = \{0, 1, 2, \dots, k, \dots\}$  gegeben.

Diese heißt Poissonverteilung mit dem Parameter  $\lambda$ .

e) Die geometrische Verteilung

Es sei  $p \in (0, 1)$ . Durch

$$g_k(p) := (1-p)^k p, \quad k \geq 0$$

sind die Einzelwahrscheinlichkeiten einer diskreten Verteilung auf  $N_0 \setminus \{0\} = \{1, 2, \dots, k, \dots\}$  gegeben. Diese Verteilung heißt geometrische Verteilung mit dem Parameter  $p$ .

f) Die hypergeometrische Verteilung

Es seien  $R, S$  positive ganze Zahlen,  $M := R + S$  und  $m$  eine ganze Zahl mit  $1 \leq m \leq M$ . Durch

$$h(M, R, m; k) := \frac{\binom{R}{k} \binom{S}{m-k}}{\binom{M}{m}}$$

sind die Einzelwahrscheinlichkeiten einer Wahrscheinlichkeitsverteilung auf  $\{0, 1, \dots, M\}$  gegeben. Diese Verteilung heißt hypergeometrische Verteilung mit den Parametern  $M, R, m$ .

Es gilt  $h(M, R, m; k) > 0$  genau dann, wenn  $[\max(0, m - S) \leq k \leq \min(m, R)]$ , wie man leicht an der Definition der Binomialkoeffizienten erkennt.

- g) Die negative Binomialverteilung  
Es seien  $p \in (0, 1)$  und  $v > 0$ . Durch

$$NB(p, v; k) := \binom{-v}{k} (-q)^k p^v \quad , \quad k \geq 0$$

mit  $q = 1 - p$  sind die Einzelwahrscheinlichkeiten einer diskreten Verteilung auf  $\{0, 1, 2, \dots, k, \dots\}$  gegeben. Diese Verteilung heißt negative Binomialverteilung mit den Parametern  $p$  und  $v$ .

Man beachte:

$$\binom{-v}{k} := \frac{(-v)(-v-1)\dots(-v-k+1)}{k!} = (-1)^k \binom{v+k-1}{k}$$

Die hier vorgestellten diskreten Verteilungen treten in Theorie und Anwendungen der Stochastik häufig auf. Sie sind Bestandteil gewisser Standardmodelle der Wahrscheinlichkeitstheorie und teilweise durch Grenzübergänge miteinander verbunden. Exemplarisch konstruieren wir als erstes ein Modell, bei dem die hypergeometrische Verteilung vorkommt und geben dann zwei Grenzwertaussagen an, die die hypergeometrische, die Binomial- und die Poissonverteilung miteinander verbinden. Zunächst erweitern wir jedoch den Begriff der diskreten Verteilung auf Zufallsgrößen.

**Definition 4.4** Ist  $X$  eine Zufallsgröße über  $(\Omega, \mathfrak{A}, P)$  mit Werten in  $(E, \mathfrak{E})$ , so heißt  $X$  eine diskret verteilte Zufallsgröße, kurz: diskrete Zufallsgröße, falls ihre Wahrscheinlichkeitsverteilung  $P^X$  auf  $(E, \mathfrak{E})$  diskret ist.

In diesem Fall gibt es nach Definition eine Folge  $(x_i, i \in I)$  mit  $I \subseteq N_0$  von Elementen aus  $E$  mit

$$\sum_{i \in I} P^X(\{x_i\}) = \sum_{i \in I} P(X = x_i) = 1 \text{ und} \quad (4.3)$$

$$P^X(B) = \sum_{i \in I: x_i \in B} P(X = x_i), \quad B \in \mathfrak{E}. \quad (4.4)$$

### Verteilungsfunktionen diskreter Verteilungen auf $R_1$

Es seien  $(x_i, i \in I)$  eine Folge reeller Zahlen und  $((x_i, p_i), i \in I)$  eine diskrete Wahrscheinlichkeitsverteilung. Das von ihr erzeugte Wahrscheinlichkeitsmaß  $P$  hat die Form

$$P(A) = \sum_{i: x_i \in A} p_i, \quad A \subseteq R_1$$

(siehe Formel (4.2)).

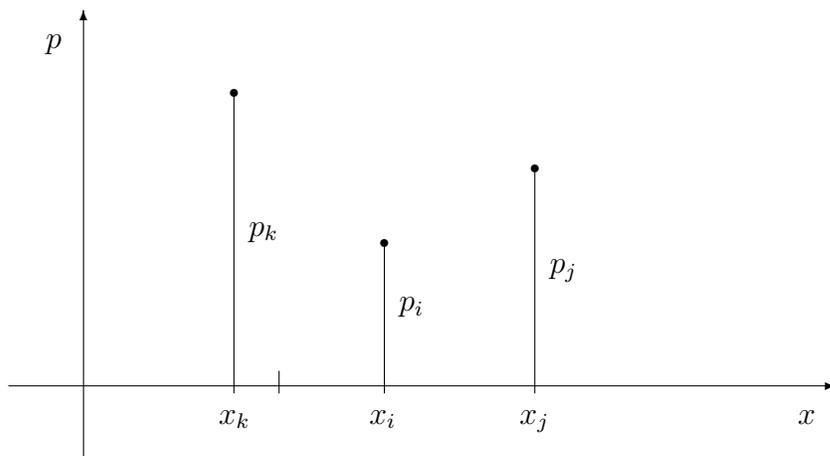


Bild 4.1

Die *Verteilungsfunktion*  $F$  der diskreten Verteilung  $((x_i, p_i), i \in I)$  ist definiert durch (siehe (3.27))

$$F(x) := P((-\infty, x]) = \sum_{i: x_i \leq x} p_i, \quad x \in \mathbb{R}_1. \quad (4.5)$$

F ur die Funktion  $F$  gilt die Aussage 3.31. Au erdem haben wir die

**Aussage 4.5** *Die Verteilungsfunktion  $F$  hat folgende Eigenschaften:*

- $\Delta F$  ist konstant auf jedem Intervall  $[a, b)$ , das keine der Zahlen  $x_i$  im Inneren enth alt.
- $F(x_i) - F(x_i - 0) = p_i$ ,  $i \in I$

Der Beweis folgt unmittelbar aus der Definition (4.3).

## Funktionen diskret verteilter Zufallsgr o en

Es sei  $X$  eine diskret verteilte Zufallsgr o e mit der Menge der m oglichen Werte  $E = \{x_i : i \in I\}$  und den zugeh origen Einzelwahrscheinlichkeiten  $(p_i^X, i \in I)$ . Ist  $\psi$  eine Funktion von  $E$  in eine abz ahlbare Menge  $F = \{f_j : j \in \mathcal{J}\}$ , so ist die Zufallsgr o e  $Y := \psi(X)$  ebenfalls diskret verteilt.

**Aussage 4.6** *Die Verteilung der Zufallsgr o e  $Y = \psi(X)$  ist diskret. Ihre m oglichen Werte sind die Elemente von  $F = \{\psi(x_i) : i \in I\} = \{f_j : j \in \mathcal{J}\}$  mit den Einzelwahrscheinlichkeiten*

$$p_j^Y = \sum_{\substack{i \in I: \\ \psi(x_i) = f_j}} p_i^X, \quad j \in \mathcal{J} \quad (4.6)$$

Beweis:  $p_j^Y = P^Y(\{f_j\}) = P^X(\psi^{-1}(\{f_j\})) = \sum_{\substack{i \in I: \\ \psi(x_i) = f_j}} p_i^X. \quad \square$

## 4.2 Die hypergeometrische Verteilung

Das folgende Modell steht für viele Situationen, in denen eine zufällige Auswahl von Elementen aus einer aus zwei Typen von Elementen bestehenden Menge (ohne Zurücklegen) vorgenommen wird (Lotto "6 aus 49", Qualitätskontrolle mit Hilfe einer Stichprobe usw.).

Gegeben sei eine Urne mit  $M$  Kugeln, davon  $R$  rote und  $S$  schwarze:

$$M = R + S.$$

Die Kugeln seien durchnummeriert von 1 bis  $M$ , dabei mögen die roten Kugeln die Nummern 1 bis  $R$  tragen. Auf gut Glück werden  $m$  Kugeln ausgewählt, nacheinander, ohne Zurücklegen. Der Einfachheit halber setzen wir  $m \leq \min(R, S)$  voraus.

Die möglichen Ausgänge  $\omega$  dieses Versuches sind, wenn die Reihenfolge der ausgewählten Kugeln keine Rolle spielt,  $m$ -elementige Teilmengen von  $\{1, 2, \dots, M\}$ :

$$\omega = \{i_1, \dots, i_m\}, \quad i_k \in \{1, 2, \dots, M\}, \quad k = 1, \dots, m.$$

Die Menge  $\Omega$  aller dieser  $\omega$  hat  $\binom{M}{m}$  Elemente. Es gibt also  $N = \binom{M}{m}$  mögliche Versuchsausgänge.

Weil die Auswahl auf gut Glück erfolgte, hat jedes  $\omega \in \Omega$  die gleiche Wahrscheinlichkeit aufzutreten. Folglich haben wir ein Laplace-Experiment mit dem Parameter  $N$ :

$$P(\{\omega\}) = \frac{1}{N} \binom{M}{m}^{-1}, \quad \omega \in \Omega.$$

Die Zufallsgröße  $X$ , definiert durch

$$X(\omega) = \sum_{k=1}^m \mathbb{1}_{\{1, \dots, R\}}(i_k), \quad \omega = \{i_1, i_2, \dots, i_m\} \in \Omega,$$

gibt an, wieviel rote Kugeln in der "Stichprobe"  $\omega$  enthalten sind. Sie hat die möglichen Werte  $0, 1, \dots, m$ , und für ihre Einzelwahrscheinlichkeiten gilt

$$P(X = j) = \frac{\#\{\omega : X(\omega) = j\}}{N} = \frac{\binom{R}{j} \binom{M-R}{m-j}}{\binom{M}{m}}, \quad j = 0, 1, \dots, m. \quad (4.7)$$

Es gilt somit

**Aussage 4.7** Werden aus einer Urne mit  $R$  roten und  $S$  schwarzen Kugeln  $m$  Kugeln nacheinander, ohne Zurücklegen und auf gut Glück ausgewählt, so hat die Zufallsgröße  $X$ , die die Anzahl der roten Kugeln in der ausgewählten Stichprobe angibt, eine hypergeometrische Verteilung mit den Parametern  $M = R + S$ ,  $R$  und  $m$ . Es gilt also (4.7).

**Bemerkung 4.8** Die Formel (4.7) bleibt auch gültig, falls  $m > \min(R, M - R)$  gilt.

**Beispiel 4.9** (Lotto "6 aus 49")

$$M = 49, m = 6, R = 6$$

(rote Kugeln  $\hat{=}$  Zahlen auf dem Tippschein,

schwarze Kugeln  $\hat{=}$  restliche der 49 Zahlen)

$X$  = Zahl der auf dem Tippschein richtig getippten Zahlen.

$$P(X = k) = \frac{\binom{6}{k} \binom{43}{6-k}}{\binom{49}{6}}, k = 0, 1, \dots, 6.$$

$k$	0	1	2	3
$P(X = k)$	0,43596498	0,41301945	0,13237803	0,0176504

$k$	4	5	6
$P(X = k)$	0,00096862	$1,845 \cdot 10^{-5}$	$7,15 \cdot 10^{-8}$

**Aussage 4.10** Mit der Bezeichnung

$$h(M, R, m; k) = \frac{\binom{R}{k}}{\binom{M-R}{m-k}}, k = 0, \dots, m, \quad (4.8)$$

gilt

$$\lim_{\substack{M, R \rightarrow \infty \\ R, M \rightarrow \infty}} h(M, R, m; k) = \binom{m}{k} p^k (1-p)^{m-k}, \quad (4.9)$$

wobei der Limes derart gebildet wird, dass für gegebenes  $p$  aus  $(0, 1)$  gilt  $M \rightarrow \infty$ ,  $R \rightarrow \infty$  mit  $R/M \rightarrow p$ ,  $m$  und  $k$  bleiben fest.

Im Grenzfall geht die hypergeometrische Verteilung also unter den genannten Bedingungen in eine Binomialverteilung mit den Parametern  $(m, p)$  über.

Beweis: Als Übungsaufgabe. (Man beachte, dass  $m$  und  $k$  beim Grenzübergang festgehalten werden.)

**Satz 4.11** (Poissonscher Grenzwertsatz)

Es gilt für jedes  $\lambda > 0$

$$\lim_{\substack{m \rightarrow \infty \\ m \cdot p_m \rightarrow \lambda}} \binom{m}{k} p_m^k (1 - p_m)^{m-k} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \geq 0 \quad (4.10)$$

Beweis: Wir schreiben  $\binom{m}{k} p_m^k (1 - p_m)^{m-k}$  in der Form  $\frac{1}{k!} \left( \prod_{j=0}^{k-1} (m-j)p_m \right) \cdot \left(1 - \frac{p_m \cdot m}{m}\right)^m \cdot \frac{1}{(1-p_m)^k}$ . Wegen  $\prod_{j=0}^{k-1} (m-j)p_m \rightarrow \lambda^k$ ,  $\left(1 - \frac{p_m \cdot m}{m}\right)^m \rightarrow e^{-\lambda}$  und  $(1-p_m)^k \rightarrow 1$  für  $m \rightarrow \infty$  mit  $m p_m \rightarrow \lambda$  folgt die Behauptung.  $\square$

### 4.3 Erwartungswert und Varianz

Erwartungswert und Varianz sind aufschlussreiche Kenngrößen einer Wahrscheinlichkeitsverteilung. Sie geben Anhaltspunkte dafür, um welchen "Schwerpunkt" sich die Trägerpunkte der Verteilung gruppieren bzw. wie stark sie um diesem Schwerpunkt "streuen".

#### Erwartungswert

Es sei  $((x_i, p_i), i \in I \subseteq N_0)$  eine diskrete Wahrscheinlichkeitsverteilung auf  $R_1$ . Ein zufälliger Versuch werde  $n$  mal (jedes Mal neu, unter im Wesentlichen gleichen Bedingungen) ausgeführt und zwar so, dass der Wert  $x_i$  mit der

Wahrscheinlichkeit  $p_i$  erscheint. Als Ergebnis der Versuchsreihe erhalten wir eine Folge  $(y_1, \dots, y_n)$  von Versuchsausgängen, wobei jedes  $y_j, j = 1, 2, \dots, n$ , gleich einem der  $x_i, i \in I$ , ist. Es sei  $n_i$  die (absolute) Häufigkeit, mit der  $x_i$  als Versuchsausgang unter den  $y_1, \dots, y_n$  auftritt, in einer Formel ausgedrückt, heißt das

$$n_i = \sum_{k=1}^n \mathbb{1}_{\{x_i\}}(y_k).$$

Offenbar gilt  $\sum_{i \in I} n_i = n$  und  $\sum_{i \in I} n_i x_i = \sum_{j=1}^n y_j$ .

Angenommen, wir erhalten nach jeder Versuchsdurchführung von einem Veranstalter so viele Euro, wie der Versuchsausgang  $x_i$  als Zahl angibt (negative Werte bedeuten Zahlungsverpflichtung für uns), dann haben wir insgesamt

$\sum_{j=1}^n y_j = \sum_{i \in I} n_i x_i$  Euro bekommen. Pro Versuch sind das also im Durchschnitt

$\frac{1}{n} \sum_{k=1}^n y_j = \sum_{i \in I} \frac{n_i}{n} x_i$ . Wir erinnern uns, bei großer Versuchsanzahl  $n$  ist die relative Häufigkeit  $\frac{n_i}{n}$  etwa gleich der Wahrscheinlichkeit  $p_i$  (Empirisches Gesetz der großen Zahlen).

Der Wert  $\mu := \sum_{i \in I} p_i x_i$  gibt also näherungsweise den Geldbetrag in Euro an, den wir in einer langen Reihe von Versuchen pro Versuch erhalten, wir sagen, den wir pro Versuch zu *erwarten* haben.

Dieser Wert wäre auch der *faire Preis*, den wir vor Durchführung jedes Versuches an den Veranstalter zu bezahlen hätten.

**Definition 4.12** Der Erwartungswert  $\mu$  einer diskreten Verteilung  $((x_i, p_i), i \in I)$  mit  $x_i \in \mathbb{R}_1, i \in I$ , existiert und ist definiert als

$$\mu = \sum_{i \in I} x_i p_i, \quad \text{falls} \quad \sum_{i \in I} x_i^+ p_i < \infty \quad \text{oder} \quad \sum_{i \in I} x_i^- p_i < \infty.$$

Anderenfalls sagt man,  $((x_i, p_i), i \in I)$  besitze keinen Erwartungswert.

Gilt  $\sum_{i \in I} |x_i| p_i < \infty$ , so ist  $|\mu| < \infty$ . In diesem Fall sagt man, die Verteilung hat einen endlichen Erwartungswert. (Dabei ist  $x^+ = \max(x, 0)$ ,  $x^- = \max(-x, 0)$ . Es gilt  $x = x^+ - x^-$ ,  $|x| = x^+ + x^-$ .)

Das empirische Gesetz der großen Zahlen kann man nach diesen Überlegungen also auch für arithmetische Mittel formulieren:

Wenn der Erwartungswert  $\mu$  existiert, so nähert sich das arithmetische Mittel  $\frac{1}{n} \sum_{k=1}^n y_j$  der Versuchsergebnisse immer mehr diesem Erwartungswert.

Fasst man die Verteilung  $((x_i, p_i), i \in I)$ , als eine Verteilung von Gewichten der Masse  $p_i$  im Punkt  $x_i, i \in I$ , auf, so ist der Erwartungswert  $\mu$  der physikalische Schwerpunkt dieser Massenverteilung. Um ihn gruppieren sich die möglichen Werte  $x_i$  der Verteilung. In erster Näherung liefert also  $\mu$  Informationen über die "Lage" dieser Verteilung. Man bezeichnet deshalb  $\mu$  auch als *Lageparameter*. Eine Verteilung heißt *zentriert*, falls ihr Erwartungswert  $\mu$  gleich Null ist.

Verschiebt man jeden Punkt  $x_i$  um einen Wert  $a$  in  $x_i + a$ , so verschiebt sich auch der Erwartungswert  $\mu$  um  $a$  in den neuen Erwartungswert  $\mu + a$ .

Setzt man  $a = -\mu$ , ergibt sich als neue Verteilung  $((x_i - \mu, p_i), i \in I)$ , und deren Erwartungswert ist gleich Null. Sie ist also *zentriert*.

**Beispiel 4.13** (Erste Fortsetzung des Beispiels 4.3):

$$\text{a) } \mu = \frac{1}{N} \sum_{k=1}^N \omega_k, \quad \text{falls } \Omega \subseteq R_1$$

$$\text{b) } \mu = \omega_0, \quad \text{falls } \Omega \subseteq R_1$$

$$\text{c) } \mu = np$$

$$\text{d) } \mu = \lambda$$

$$\text{e) } \mu = \frac{1-p}{p}$$

$$\text{f) } \mu = \frac{Rm}{M}$$

$$\text{g) } \mu = v \cdot \frac{1-p}{p}$$

**Definition 4.14** Ist  $X$  eine diskret verteilte reellwertige Zufallsgröße, so bezeichnet man als Erwartungswert von  $X$  den Erwartungswert ihrer Wahrscheinlichkeitsverteilung  $P^X$  und verwendet für ihn das Symbol  $EX$ :

$$EX = \sum_{i \in I} x_i P^X(\{x_i\}) = \sum_{i \in I} x_i P(X = x_i)$$

Dabei bilden die  $x_i, i \in I$ , die möglichen Werte von  $X$ .

Eine sehr einfache Zufallsgröße ist  $X(\omega) = \mathbb{1}_A(\omega)$  mit  $A \in \mathfrak{A}$ . Es gilt  $EX = E\mathbb{1}_A = P(A)$ .

**Aussage 4.15** (Erwartungswert der Funktion einer Zufallsgröße)

Es sei  $X$  eine diskret verteilte Zufallsgröße über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  mit Werten in irgendeiner abzählbaren Menge  $E = \{x_i : i \in I \subseteq \mathbb{N}_0\} \subseteq \mathbb{R}_1$  und mit den entsprechenden Einzelwahrscheinlichkeiten  $(p_i^X, i \in I)$ . Außerdem sei  $\psi$  eine reellwertige Funktion auf  $E$  mit Werten in  $F = \{f_j : j \in \mathfrak{J} \subseteq \mathbb{N}_0\}$ . Dann ist  $Y = \psi(X)$  eine reellwertige diskret verteilte Zufallsgröße, und es gilt (siehe (5)):

$$EY = E\psi(X)$$

$$\sum_{i \in I} \psi(x_i) P(X = x_i) p_i^X \tag{4.11}$$

wobei dieser Erwartungswert nach Definition nicht existiert, falls

$$\sum_{i \in I} (\psi(x_i))^+ P(X = x_i) \text{ und } \sum_{i \in I} (\psi(x_i))^- P(X = x_i) = \infty \text{ gilt.}$$

Beweis:

$$EY = \sum f_j p_j^Y = \sum_j f_j \sum_{\substack{i \in \mathcal{J}: \\ \psi(x_i) = f_j}} p_i^X = \sum_{j \in \mathcal{J}} \sum_{\substack{i \in \mathcal{J}: \\ \psi(x_i) = f_j}} f_j p_i^X = \sum_{i \in \mathcal{J}} \psi(x_i) p_i^X.$$

### Beispiel 4.16

- 1) Ist  $\psi(x) = ax + b$ ,  $x \in R_1$ ,  $a, b$  reellwertige Konstanten, so gilt, sofern  $EX$  existiert,

$$E(aX + b) = a(EX) + b$$

- 2) Für jede reellwertige diskrete Zufallsgröße  $X$  ist auch  $X^2$  eine Zufallsgröße, und es gilt

$$EX^2 = \sum_{i \in I} x_i^2 P(X = x_i).$$

## Momente diskreter Verteilungen auf $R_1$

Es sei  $((x_i, p_i), i \in I)$ , eine diskrete Wahrscheinlichkeitsverteilung auf  $R_1$ .

**Definition 4.17** Es sei  $k \geq 1$ . Als  $k$ -tes Moment der Wahrscheinlichkeitsverteilung  $((x_i, p_i), i \in I)$ , bezeichnet man die Größe

$$\mu_k := \sum_{i \in I} x_i^k p_i,$$

sofern  $\sum (x_i^+)^k p_i < \infty$  oder  $\sum (x_i^-)^k p_i < \infty$ . Anderenfalls sagt man, falls  $k$  ungerade ist, das  $k$ -te Moment existiert nicht. Sind beide Summen endlich, so konvergiert die Summe  $\sum_{i \in I} |x_i|^k p_i$  und das  $k$ -te Moment  $\mu_k = \sum_{i \in I} x_i^k p_i$  ist endlich.

Der Erwartungswert ist offensichtlich das erste Moment der Verteilung  $(x_i, p_i)$ :  $\mu = \mu_1$ . Gilt  $|\mu_k| < \infty$  für ein  $k > 1$ , so ist auch  $|\mu_l| < \infty$  für alle  $l$  mit  $1 \leq l < k$ . Das folgt sofort aus  $|\mu_l| \leq \sum_{i \in I} |x_i|^l p_i \leq \sum [\max(1, |x_i|)]^k p_i \leq 1 + \sum_{i \in I} |x_i|^k p_i$ .

**Definition 4.18** *Es sei  $k \geq 2$ . Als  $k$ -tes zentrales Moment einer Wahrscheinlichkeitsverteilung  $(x_i, p_i)$ ,  $i \in I$ , bezeichnet man das  $k$ -te Moment der zentrierten Verteilung  $(x_i - \mu, p_i)$ ,  $i \in I$ :*

$$m_k := \sum_{i \in I} (x_i - \mu)^k p_i,$$

sofern  $\sum ((x_i - \mu)^+)^k p_i < \infty$  oder  $\sum ((x_i - \mu)^-)^k p_i < \infty$  gilt. Anderenfalls sagt man, falls  $k$  ungerade ist, das  $k$ -te zentrale Moment existiert nicht.

Es gilt:  $|m_k| < \infty$  genau dann, wenn  $|\mu_k| < \infty$  ( $k \geq 2$ ). In diesem Fall ist

$$m_k = \sum_{\ell=0}^k \binom{k}{\ell} \mu_\ell (-\mu)^{k-\ell}, \quad k \geq 2 \quad (4.12)$$

mit  $\mu_0 := 1$ , insbesondere gilt:

$$m_2 = \mu_2 - \mu_1^2. \quad (4.13)$$

Umgekehrt haben wir

$$\mu_k := \sum_{i \in I} (x_i - \mu + \mu)^k p_i = \sum_{\ell=0}^k \binom{k}{\ell} m_\ell \cdot \mu^{k-\ell} \quad (4.14)$$

mit  $m_0 := 1, m_1 = 0$ .

Mit Hilfe der Momente einer Wahrscheinlichkeitsverteilung auf  $R_1$  kann man eine erste Vorstellung von der Lage und der Form der Wahrscheinlichkeitsverteilung auf  $R_1$  gewinnen.

**Definition 4.19** Als  $k$ -tes Moment einer diskreten reellwertigen Zufallsgröße  $X$  über  $(\Omega, \mathfrak{A}, P)$  bezeichnet man das  $k$ -te Moment  $\mu_k^X$  ihrer Wahrscheinlichkeitsverteilung  $P^X$ .

Es gilt:

$$\mu_k^X = \sum_{i \in I} x_i^k P^X(\{x_i\}) = \sum_{i \in I} x_i^k P(X = x_i) = E(X^k). \quad (4.15)$$

mit den gleichen Existenz- bzw. Nichtexistenzbedingungen wie beim  $k$ -ten Moment irgendeiner diskreten Verteilung auf  $R_1$ . Wir schreiben  $\mu^X = \mu_1^X$ . Schließlich führt man für  $k \geq 2$  das  $k$ -te zentrale Moment für  $X$  ein als

$$\begin{aligned} m_k^X &= \sum_{i \in I} (x_i - \mu^X)^k P^X(\{x_i\}) = \\ &= \sum_{i \in I} (x_i - \mu^X)^k P(X = x_i) = E(X - \mu^X)^k. \end{aligned} \quad (4.16)$$

## Varianz

Das erste Moment, der Erwartungswert  $\mu$ , kennzeichnet die Lage der Verteilung, das zweite zentrale Moment vermittelt eine Vorstellung, wie breit die Verteilung um den Erwartungswert platziert ist. Es hat einen eigenen Namen.

**Definition 4.20** Als Varianz oder Streuung der Wahrscheinlichkeitsverteilung  $((x_i, p_i), i \in I)$  bezeichnet man die Größe

$$\sigma^2 := \sum_{i \in I} (x_i - \mu)^2 p_i. \quad (4.17)$$

Die Wurzel aus der Varianz  $\sigma = (\sigma^2)^{\frac{1}{2}}$  nennt man Standardabweichung der zugrunde liegenden Verteilung.

Es gilt  $\sigma^2 \geq 0$ . Wir haben  $\sigma^2 = 0$  genau dann, wenn die Verteilung  $((x_i, p_i), i \in I)$  ausgeartet ist, also die Verteilung in nur einem Punkt  $x_{i_0}$  für ein  $i_0 \in I$  konzentriert ist, d. h. wenn gilt  $p_{i_0} = 1$ . In diesem Fall ist  $\mu = x_{i_0}$ .

**Beispiel 4.21** (Zweite Fortsetzung der Beispiele aus 4.3):

$$\text{a) } \sigma^2 = \frac{1}{N} \sum_{i \in I} (\omega_i - \mu)^2, \text{ falls } \Omega \subseteq R_1$$

$$\text{b) } \sigma^2 = 0 \quad , \text{ falls } \omega_0 \in R_1$$

$$\text{c) } \sigma^2 = np(1-p)$$

$$\text{d) } \sigma^2 = \lambda$$

$$\text{e) } \sigma^2 = \frac{1-p}{p^2}$$

$$\text{f) } \sigma^2 = \frac{Rm(M-R)(M-m)}{N^2(N-1)}$$

$$\text{g) } \sigma^2 = \left( \frac{v(1-p)}{p^2} \right)$$

**Definition 4.22** Die Varianz einer diskret verteilten reellwertigen Zufallsgröße  $X$  mit der Verteilung  $P^X$ , gegeben durch  $((x_i, p_i^X), i \in I)$ , ist definiert als

$$\sigma_X^2 := E(X - EX)^2 = \sum_{i \in I} (x_i - EX)^2 p_i^X.$$

Man schreibt auch  $\text{Var}(X)$  oder  $D^2X$  für  $\sigma_X^2$ . Die Standardabweichung  $\sigma_X$  der Zufallsgröße  $X$  ist definiert als der Wert  $(\sigma_X^2)^{\frac{1}{2}}$ .

Offenbar gilt die für Berechnungen nützliche Formel

$$D^2X = EX^2 - (EX)^2$$

**Aussage 4.23** (Tschebyschev'sche Ungleichung) Ist  $0 < D^2X < \infty$ , so gilt für jedes  $\varepsilon > 0$  die Ungleichung

$$P(|X - EX| > \varepsilon) \leq \frac{D^2X}{\varepsilon^2}.$$

Beweis:

$$\begin{aligned} P(|X - EX| > \varepsilon) &= P^X(\{x_i : |x_i - EX| > \varepsilon\}) = \\ &= \sum_{\substack{i \in I \\ |x_i - EX| > \varepsilon}} P^X(\{x_i\}) \leq \sum_{i \in I} \frac{|x_i - EX|^2}{\varepsilon^2} P^X(\{x_i\}) = \frac{D^2X}{\varepsilon^2}. \end{aligned}$$

□

Die Tschebyschev'sche Ungleichung besagt, dass, je kleiner die Varianz von  $X$  ist, umso unwahrscheinlicher ist es, dass die Zufallsgröße  $X$  bei einer Durchführung des zugrunde liegenden zufälligen Versuches um mehr als  $\varepsilon$  vom Erwartungswert  $EX$  abweicht.

Im Fall  $D^2X = 0$  gilt  $P(X = EX) = 1$ , es gibt also mit Wahrscheinlichkeit Eins keine Abweichung vom Erwartungswert, d.h. die Verteilung  $P^X$  ist ausgeartet und konzentriert in einem Punkt, der dann natürlich gleich  $EX$  ist.

## Diskret verteilte zweidimensionale zufällige Vektoren

In vielen Fällen interessiert man sich im Rahmen eines zufälligen Versuches nicht nur für einzelne Zufallsgrößen, sondern für mehrere verschiedene. Diese sind dann im Allgemeinen nicht ohne innere Zusammenhänge, was man nur durch die Untersuchung ihrer gemeinsamen Wahrscheinlichkeitsverteilung feststellen kann und nicht an den einzelnen Zufallsgrößen bzw. ihren Verteilungen. Man denke beispielsweise an Körpergröße und Gewicht einer zufällig gewählten Person. Wir geben hier eine Einführung in diese Fragestellung im Rahmen zweier diskret verteilter Zufallsgrößen, sie bilden, zusammengefasst,

einen zweidimensionalen zufälligen Vektor.

Es sei  $X = (U, V)^T$  ein zufälliger Vektor über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  mit Werten in einer Menge  $E := E_U \times E_V$ , wobei  $E_U$  und  $E_V$  höchstens abzählbar viele Elemente enthalten mögen:

$$E_U = \{u_i, i \in I\} \text{ und } E_V = \{v_j, j \in \mathfrak{J}\}.$$

Hier seien  $I$  und  $\mathfrak{J}$  Teilmengen von  $N_0$ .

Die möglichen Werte der Zufallsgröße  $U$  sind also die  $u_i \in E_U$ , die von  $V$  die  $v_j \in E_V$ .

Die möglichen Werte von  $X$  sind die Paare  $(u_i, v_j), (i, j) \in I \times \mathfrak{J}$ . Folglich besitzt  $X$  eine diskrete Wahrscheinlichkeitsverteilung  $P^X$ . Ihre Einzelwahrscheinlichkeiten seien gegeben durch

$$P^X((u_i, v_j)) = P(U = u_i, V = v_j) =: p_{ij} \quad , \quad i \in I, j \in \mathfrak{J}.$$

Nach Definition diskreter Verteilungen gilt dann für die Wahrscheinlichkeit  $P^X(B)$ , dass der zufällige Vektor  $X$  einen Wert aus  $B$  annimmt (siehe Notation...):

$$P^X(B) = P(X \in B) = P((U, V) \in B) = \sum_{\substack{(i,j): \\ (u_i, v_j) \in B}} p_{ij} \quad , \quad B \subseteq E. \quad (4.18)$$

**Definition 4.24** Die Verteilung  $P^X$  heißt gemeinsame Verteilung von  $U$  und  $V$  und ist gemäß Formel (4.20) eindeutig bestimmt durch ihre Einzelwahrscheinlichkeiten  $p_{ij}, i \in I, j \in \mathfrak{J}$ .

Die Verteilungen der einzelnen Zufallsgrößen  $U$  und  $V$  ergeben sich aus ihrer gemeinsamen Verteilung  $P^X$  durch

$$P^U(C) = P(U \in C) = P(U \in C, V \in E_V) = \sum_{\substack{i \in I: u_i \in C \\ j \in \mathfrak{J}}} p_{ij} \quad , \quad C \subseteq E_U \quad (4.19)$$

$$P^V(D) = P(V \in D) = P(U \in E_U, V \in D) = \sum_{\substack{j \in \mathfrak{J}: v_j \in D \\ i \in I}} p_{ij} \quad , \quad D \subseteq E_V \quad (4.20)$$

$P^U$  und  $P^V$  sind also die Randverteilungen von  $P^X$ . Ihre Eigenschaften ergeben sich wie folgt:

$$P^U(\{u_i\}) = \sum_{j \in \mathfrak{J}} p_{ij} =: p_{i\cdot} \quad i \in I, P^V(\{v_j\}) = \sum_{i \in I} p_{ij} =: p_{\cdot j}, \quad j \in \mathfrak{J} \quad (4.21)$$

Die Bezeichnung Randverteilung wird hier besonders verständlich, wenn man die Einzelwahrscheinlichkeiten  $p_{ij}$  in einem Schema wie folgt anordnet.

$i \setminus j$	1	2	3	...	$j$	...	
1	$p_{11}$	$p_{12}$	...	...	$p_{1j}$	...	$p_{1\cdot}$
2	$p_{21}$	$p_{22}$					$p_{2\cdot}$
3	.						
.	.						
.	.						
.	.						
$i$	$p_{i1}$	...			...	$p_{ij}$	...
.	.						$p_{i\cdot}$
	$p_{\cdot 1}$	$p_{\cdot 2}$				$p_{\cdot j}$	1

**Bemerkung 4.25** Die Verteilung  $(p_{ij})$  bestimmt die Randverteilungen  $(p_{i\cdot})$  und  $(p_{\cdot j})$  eindeutig. Die Randverteilungen bestimmen aber die gemeinsame Verteilung noch nicht eindeutig.

Das wird deutlich an dem nächsten Schema, das für jedes  $c \in [0, \frac{1}{4}]$  eine zwei-dimensionale diskrete Verteilung darstellt:

	0	1	
0	$\frac{1}{4} + c$	$\frac{1}{4} - c$	$\frac{1}{2}$
1	$\frac{1}{4} - c$	$\frac{1}{4} + c$	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	

**Beispiel 4.26** Aus einem Kartenspiel mit 32 Karten (Skatspiel) werden nacheinander auf gut Glück ohne Zurücklegen der ersten Karte zwei Karten gezogen. Es sei  $U = 1$  (bzw.  $V = 1$ ), falls die erste (bzw. zweite) Karte ein König ist. Anderenfalls setzen wir  $U = 0$  (bzw.  $V = 0$ ). Dann ergibt sich unter Verwendung des Modells für die hypergeometrische Verteilung für die Einzelwahrscheinlichkeiten  $p_{ij}$  der gemeinsamen Verteilung von  $U$  und  $V$  und die Randverteilungen (vgl. den Abschnitt über hypergeometrische Verteilungen)

$U \setminus V$	0	1	
0	$\frac{28}{32} \cdot \frac{27}{31}$	$\frac{28}{32} \cdot \frac{4}{31}$	$\frac{7}{8}$
1	$\frac{4}{32} \cdot \frac{28}{31}$	$\frac{4}{32} \cdot \frac{3}{31}$	$\frac{1}{8}$
	$\frac{7}{8}$	$\frac{1}{8}$	

## Funktionen diskret verteilter zufälliger Vektoren

**Aussage 4.27** Es sei  $\psi$  eine reellwertige Funktion auf  $E = E_U \times E_V$  mit Werten in einer höchstens abzählbar unendlichen Menge  $E_W = \{w_k, k \in K\}$ .

Dann ist

$$W(\omega) = \psi(U(\omega), V(\omega)), \quad \omega \in \Omega$$

eine diskret verteilte Zufallsgröße mit Werten in  $E_W$  und den Einzelwahrscheinlichkeiten

$$P^W(\{w_k\}) = \sum_{\substack{i,j: \\ \psi(u_i, v_j) = w_k}} p_{ij}, \quad k \in K. \quad (4.22)$$

Beweis:

$$P^W(\{w_k\}) = P(W = w_k) = P(\{\omega \in \Omega | W(\omega) = w_k\}) = P(W^{-1}(\{w_k\})) =$$

$$P(\{\omega : (U(\omega), V(\omega)) \in \psi^{-1}(\{w_k\})\}) = \sum_{\substack{i,j: \\ \psi(u_i, v_j) = w_k}} p_{ij}, \quad k \in K.$$

□

Wir benötigen im Weiteren den Erwartungswert reellwertiger Funktionen mehrere Zufallsgrößen und nutzen dafür die folgende

**Aussage 4.28** *Gilt  $E_W \subseteq$  und  $\sum_{(i,j) \in I \times \mathfrak{J}} |\psi(u_i, v_j)| p_{ij} < \infty$ , so hat  $W = \psi(U, V)$  einen endlichen Erwartungswert, und es gilt*

$$EW = E\psi(U, V) = \sum_{(i,j) \in I \times \mathfrak{J}} \psi(u_i, v_j) p_{ij} \quad (4.23)$$

Beweis:

$$\begin{aligned} EW &= \sum_{k \in K} w_k P(W = w_k) = \sum_{k \in K} w_k \sum_{\substack{(i,j): \\ \psi(u_i, v_j) = w_k}} p_{ij} = \\ &= \sum_{k \in K} \sum_{\substack{(i,j): \\ \psi(u_i, v_j) = w_k}} \psi(u_i, v_j) p_{ij} = \sum_{(i,j) \in I \times \mathfrak{J}} \psi(u_i, v_j) p_{ij}. \end{aligned}$$

□

**Folgerungen 4.29** Sind  $U$  und  $V$  reellwertige Zufallsgrößen mit endlichem Erwartungswert und  $a, b$  reelle Zahlen, so hat auch  $aU + bV$  einen endlichen

Erwartungswert, und es gilt

$$E(aU + bV) = aEU + bEV. \quad (4.24)$$

$$\text{Var}(aU + bV) = a^2\text{Var}(U) + b^2\text{Var}(V) + 2abE(u - EU)(V - EV) \quad (4.25)$$

Beweis: Wegen (4.25) gilt

$$\begin{aligned} E(aU + bV) &= \sum_{(i,j)} (au_i + bv_j)p_{ij} = a \sum_{i,j} u_i p_{ij} + b \sum_{i,j} v_j p_{ij} \\ &= a \sum_i u_i p_{i\cdot} + b \sum_j v_j p_{\cdot j} = aEU + bEV \end{aligned}$$

und

$$\begin{aligned} \text{Var}(aU + bV) &= E(aU + bV - E(aU + bV))^2 = E((aU - EaU) + (bV - EbV))^2 \\ &= a^2\text{Var}U + b^2\text{Var}V + 2abE(U - EU)(V - EV) \end{aligned}$$

□

**Bemerkung 4.30** Im Allgemeinen gilt nicht  $E(UV) = EUEV$ . Das sieht man am Beispiel  $\psi(U, V) = UV, P(U = i, V = j) = \frac{1}{4} + c \cdot (-1)^{i+j}, i, j \in \{0, 1\}$  für  $c \in (0, \frac{1}{4})$ .

## 4.4 Kovarianz und Korrelation

Es sei  $(U, V)$  ein diskret verteilter zufälliger Vektor über  $(\Omega, \mathfrak{A}, P)$  mit Werten  $(u_i, v_j)$  in  $R_2$ :

$$P(U = u_i, V = v_j) = p_{ij} \quad , \quad (i, j) \in I \times \mathfrak{J}$$

**Aussage 4.31** (Cauchy-Schwarz-Ungleichung) Gilt  $E(U^2) < \infty$  und  $E(V^2) < \infty$ , so ist  $E|U \cdot V| < \infty$  und

$$(E(UV))^2 \leq EU^2EV^2. (7) \quad (4.26)$$

Das Gleichheitszeichen gilt in (4.28) genau dann, wenn es eine reelle Zahl  $c$  gibt mit  $U = cV$   $P$ -f.s. oder mit  $V = cU$   $P$ -f.s.

(Eine Gleichung zwischen zwei Zufallsgrößen gilt  $P$ -fast sicher, kurz:  $P$ -f.s., falls die Menge aller  $\omega \in \Omega$ , für die sie nicht erfüllt ist, eine  $P$ -Nullmenge bildet.)

Beweis: O.B.d.A. sei  $EU^2 > 0$  und  $EV^2 > 0$ . Anderenfalls gilt  $U = 0$   $P$ -f.s. oder  $V = 0$   $P$ -f.s.. Das Gleichheitszeichen in (4.28) und der zweite Teil der Aussage sind dann richtig.

Für jedes  $\beta$  aus  $R_1$  ist  $E(U + \beta V)^2 < \infty$  und zwar wegen  $(a + b)^2 \leq 2(a^2 + b^2)$  ist  $E(U + \beta V)^2 \leq 2EU^2 + 2\beta^2EV^2$  und der Voraussetzung.

Setzt man zunächst

$$\beta = \left(\frac{EU^2}{EV^2}\right)^{\frac{1}{2}} \text{ und dann } \beta = -\left(\frac{EU^2}{EV^2}\right)^{\frac{1}{2}},$$

so erhält man wegen  $E(U + \beta V)^2 \geq$  die Ungleichungen

$$-(EU^2EV^2)^{\frac{1}{2}} \leq E(UV) \leq (EU^2EV^2)^{\frac{1}{2}},$$

woraus sich (4.28) ergibt.

Das Gleichheitszeichen in (4.28) gilt wegen  $EV^2 > 0$  genau dann, wenn  $E(U + \beta V)^2 = 0$  für ein  $\beta$  aus  $R_1$  richtig ist. In diesem Fall ist  $U = -\beta V$   $P$ -f.s. und notwendigerweise  $\beta^2 = \frac{EU^2}{EV^2}$ .

□

**Definition 4.32** Es sei  $E(U^2) < \infty$  und  $E(V^2) < \infty$ . Dann heißt die durch

$$Kov(U, V) := E((U - EU)(V - EV))$$

definierte Größe die Kovarianz zwischen  $U$  und  $V$ .

**Aussage 4.33** Die Kovarianz hat folgende Eigenschaften ( $\alpha, \beta$  seien zwei beliebige reelle Zahlen,  $W$  eine dritte Zufallsgröße):

1.  $Kov(U, V) = Kov(V, U)$
2.  $Kov(\alpha U, V) = \alpha Kov(U, V)$
3.  $Kov(U + W, V) = Kov(U, V) + Kov(W, V)$
4.  $Kov(U, V) = E(UV) - EUEV$
5.  $Kov(U, U) = D^2U$
6.  $Kov(U, \beta) = 0$
7.  $(Kov(U, V))^2 \leq D^2U \cdot D^2V$
8.  $(Kov(U, V))^2 = D^2UD^2V \iff \exists$  Es existieren  $a, b \in R_1 : V = aU + b P - f.s.$  oder es existieren  $c, d \in R_1 : U = cV + d P - f.s.$

Der Nachweis dieser Eigenschaften folgt für 1. - 6. unmittelbar aus der Definition der Kovarianz und für 7. und 8. mit Hilfe der Cauchy-Schwarz-Ungleichung.

**Definition 4.34** Es sei  $D^2U, D^2V \in (0, \infty)$ . Dann bezeichnet man die Zahl

$$Kor(U, V) := \frac{Kov(U, V)}{(D^2UD^2V)^{\frac{1}{2}}}$$

als den Korrelationskoeffizienten zwischen  $U$  und  $V$  oder einfach als Korrelation zwischen  $U$  und  $V$ .

Wegen der Cauchy-Schwarz-Ungleichung gilt  $|Kor(U, V)| \leq 1$ .

Wir haben  $|Kor(U, V)| = 1$  genau dann, wenn  $U$  und  $V$  linear (genauer: affin) abhängig sind, d. h., wenn es Zahlen  $a, b$  und  $c$  gibt mit  $aU + bV + c = 0$   $P$ -f.s. (Zum Beweis nutze man Eigenschaft 8 von Aussage 4.32. ) Im letzteren Fall gilt  $Kor(U, V) = 1$ , falls  $ab < 0$  und  $Kor(U, V) = -1$  falls  $ab > 0$ .

**Aussage 4.35** *Der Korrelationskoeffizient hat die Eigenschaften*

$$1.' \quad Kor(U, V) = Kor(V, U),$$

$$2.' \quad Kor(\alpha U, V) = Kor(U, V).$$

**Definition 4.36** *Gilt für zwei Zufallsgrößen  $U, V$  mit  $D^2U < \infty$  und  $D^2V < \infty$  die Beziehung  $Kor(U, V) = 0$ , so heißen  $U$  und  $V$  unkorreliert.*

Die Größe  $Kor(U, V)$  gibt den Grad der linearen Abhängigkeit zwischen den Zufallsgrößen  $U$  und  $V$  an. Für  $Kor(U, V) = 1$  und  $Kor(U, V) = -1$  liegt vollständige lineare Abhängigkeit vor.  $Kor(U, V) = 0$  deutet auf eine gewisse Unabhängigkeit in einem noch zu präzisierenden Sinn.

Man beachte, dass auf Grund der Definition der Eigenschaft 4. der Aussage 4.32 gilt

$$Kor(U, V) = 0 \iff Kov(U, V) = 0 \iff E(UV) = EU \cdot EV \quad (4.27)$$

## 4.5 Regressionsgerade

Wir beginnen mit einer Vorüberlegung über die beste Möglichkeit, den Wert einer Zufallsgröße, den sie bei dem ihr zugrunde liegenden zufälligen Versuch annehmen wird, vorherzusagen.

Es sei  $X$  eine reellwertige (diskret verteilte) Zufallsgröße über  $(\Omega, \mathfrak{A}, P)$  mit  $D^2X < \infty$ .

Wenn man vor Ausführung des zufälligen Versuches  $(\Omega, \mathfrak{A}, P)$  den Wert, den  $X$  annehmen wird, durch eine reelle Zahl  $c$  voraussagen soll, so ist das im Fall  $D^2X > 0$  zunächst einmal nicht mit Sicherheit möglich. Um es dennoch so gut wie möglich zu tun, muss man präzisieren, was man unter "so gut wie möglich" verstehen will. Eine Möglichkeit besteht darin, zu fordern, dass in einer langen Reihe von Realisierungen von  $X$ , nämlich  $(x_1, x_2, \dots, x_n)$ ,  $c$  so gewählt wird, dass  $\sum_{i=1}^n (x_i - c)^2$  minimal wird ("Minimierung der quadratischen Abweichung", "Methode der kleinsten Quadrate").

Das führt auf  $c = \frac{1}{n} \sum_{i=1}^n x_i$ . Das empirische Gesetz der großen Zahlen besagt, dass dieses arithmetische Mittel für die Zufallsgröße  $X$ , in der Nähe von  $EX$  liegt.

Wir machen uns von der Durchführung des Versuches unabhängig und verwenden als Vorhersage von  $X$  den Wert  $c = EX$ . Tatsächlich erreicht auch die Funktion  $c \rightarrow E(X - c)^2$  bei  $c = EX$  ein Minimum. Die "beste" Voraussage für  $X$  ist also  $EX$  (im Sinne der Minimierung des quadratischen Mittels).

Die Streuung  $D^2X = E(X - EX)^2$  ist gerade der Wert dieses Minimums und bildet ein Maß für die "Güte" der Voraussage von  $X$  durch  $EX$ . Je kleiner  $D^2X$  ist, umso genauer ("im quadratischen Mittel") wird diese Voraussage sein.

Wir wenden uns nun dem eigentlichen Anliegen dieses Abschnittes zu.

Es seien  $U$  und  $V$  zwei (diskret verteilte) reellwertige Zufallsgrößen über demselben Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  mit  $0 < EU^2 < \infty, 0 < EV^2 < \infty$ . Die Aufgabe bestehe darin, auf Grundlage der Kenntnis, welchen Wert  $U$  angenommen hat, den Wert von  $V$  möglichst gut vorherzusagen. Zur Illustration stelle man sich wieder den Fall vor, dass  $U$  die Körpergröße und  $V$  das Gewicht einer zufällig ausgewählten Person sind.

Im Allgemeinen gibt es keine deterministische Funktion  $\psi$ , so dass  $V = \psi(U)$  gilt. Um  $V$  mit Hilfe von  $U$  möglichst gut vorauszusagen, suchen wir Koeffizienten  $a, b \in R_1$ , die die mittlere quadratische Abweichung

$$(a, b) \longrightarrow E(V - aU - b)^2 =$$

$$EV^2 + a^2EU^2 + b^2 - 2aE(UV) - 2bEV + 2abEU$$

minimal werden lassen, d. h., wir suchen unter allen linearen Funktionen von  $U$  diejenige, die  $V$  am besten approximiert.

Das führt auf die Gleichungen

$$b = EV - aEU \text{ und}$$

$$aD^2U = Kov(U, V).$$

Also ist

$$\hat{V} := EV + Kor(U, V) \cdot \frac{\sigma_V}{\sigma_U}(U - EU)$$

die beste lineare Approximation von  $V$  durch  $U$ . **Definition:** Die Gerade  $v = g(u) = EV + a(u - EU)$ ,  $u \in R_1$

$$\text{mit } a = Kor(U, V) \left( \frac{\sigma_V}{\sigma_U} \right) = \frac{Kov(U, V)}{\sigma_U^2}$$

heißt *Regressionsgerade* für  $V$  bezüglich  $U$ . Die Zufallsgröße  $\hat{V} = g(U)$  ist die (im quadratischen Mittel) beste lineare Funktion von  $U$  für die Voraussage von  $V$  auf der Basis von  $U$  (= *Regressionsgerade* für  $V$  auf der Basis von  $U$ ).

Man wird mit der Vorhersage  $\hat{V}$  für  $V$  den tatsächlich eintretenden Wert von  $V$  i. Allg. nicht genau treffen. Im Mittel allerdings schon, denn es gilt  $E\hat{V} = EV$ . Die tatsächliche "Ungenauigkeit"  $\hat{V} - V$  hängt vom Zufall ab. Wir messen sie durch ihre Varianz  $E(\hat{V} - V)^2$ , für die sich nach einfacher Rechnung

$$E(V - \hat{V})^2 = \sigma_V^2(1 - (Kor(U, V))^2)$$

ergibt. Diese Zahl bezeichnet man als *Reststreuung*, die zwischen der Vorhersage  $\hat{V}$  und dem vorherzusagendem Wert  $V$  noch besteht, und die man auf Grundlage der Vorhersage von  $V$  durch eine lineare Funktion von  $U$  nicht beseitigen kann.

Hier wird noch einmal deutlich, dass  $Kor(U, V)$  ein Maß für den linearen Zusammenhang zwischen  $U$  und  $V$  ist.

Spezialfälle:

- a)  $Kor(U, V) = 0 \implies$  keine Reduzierung von  $\sigma_V^2$ , die beste lineare Funktion  $\hat{V}$  zur Vorhersage von  $V$  auf der Basis von  $U$  hängt gar nicht von  $U$  ab und ist gleich dem Wert  $EV$ .
- b)  $|Kor(U, V)| = 1 : \hat{V} = V$ , keine Reststreuung, exakte Voraussage von  $V$  durch eine lineare Funktion von  $U$  möglich

## 4.6 Erzeugende Funktionen

Für diskrete Verteilungen auf den natürlichen Zahlen stellen die sogenannten *erzeugenden Funktionen* ein wirkungsvolles analytisches Hilfsmittel dar, um zum Beispiel Momente der Verteilung zu bestimmen. Weitere Anwendungen werden wir später kennen lernen.

Es sei  $X$  eine Zufallsgröße über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$ , die nur Werte aus der Menge  $N_0$  der natürlichen Zahlen annehmen kann, und mit Einzelwahrscheinlichkeiten ihrer Verteilung

$$p_k = P(X = k), \quad k \geq 0.$$

**Definition 4.37** Als erzeugende Funktion  $g(s)$ ,  $s \in [-1, 1]$ , der Zufallsgröße  $X$  (genauer: ihrer Wahrscheinlichkeitsverteilung  $P^X$ ) bezeichnet man die Funktion

$$g(s) := E(s^X) = \sum_{k \geq 0} s^k p_k, \quad s \in [-1, 1].$$

Wegen  $p_k \geq 0$  und  $\sum_{k \geq 0} p_k = 1$  ist  $g(\cdot)$  eine Potenzreihe mit einem Konvergenzradius  $\rho \geq 1$ . Daraus ergeben sich sofort einige Eigenschaften, die wir in folgender Aussage zusammenfassen.

**Aussage 4.38** In der soeben eingeführten Terminologie gilt

(i)  $g(\cdot)$  ist in  $(-1, 1)$  unendlich oft differenzierbar mit

$$\frac{d^k}{ds^k} g(s) = \sum_{j \geq k} j(j-1) \cdots (j-k+1) s^{j-k} p_j =$$

$$E(X(X-1) \cdots (X-k+1) s^{X-k}),$$

es gilt

$$p_k = \frac{1}{k!} \frac{d^k}{ds^k} g(s) \Big|_{s=0}, \quad k \geq 0. \quad (4.28)$$

(ii) Im Fall  $EX^k < \infty$  haben wir die Gleichung

$$E(X(X-1)(X-2) \cdots (X-k+1)) = \lim_{s \uparrow 1} \frac{d^k}{ds^k} g(s) < \infty. \quad (4.29)$$

Gilt dagegen  $EX^k = \infty$ , so ist

$$E(X(X-1) \cdots (X-k+1)) = \lim_{s \uparrow 1} \frac{d^k}{ds^k} g(s) = \infty.$$

(iii) Sind  $g(\cdot)$  und  $h(\cdot)$  erzeugende Funktion zweier Zufallsgrößen  $X$  bzw.  $Y$  mit Werten in  $N_0$ , und gilt  $g(s) = h(s)$ ,  $s \in [0, \delta]$ , für ein  $\delta > 0$ , so sind die Verteilungen  $P^X$  und  $P^Y$  einander gleich:

$$P(X = k) = P(Y = k), \quad k \geq 0.$$

Beweis:

- (i) Es sei  $|s| < 1$  und  $\delta > 0$ , so dass  $(s - \delta, s + \delta) \subseteq (-1, 1)$ .  
Dann ist für alle  $h \in (-\delta, \delta)$

$$A(s, h) := |h^{-1}[g(s+h) - g(s)] - \sum_{k=1}^{\infty} k s^{k-1} p_k| =$$

$$\sum_{k \geq 1} [h^{-1}((s+h)^k - s^k) - k s^{k-1}] p_k$$

Weiterhin gibt es für jedes  $k \geq 2$  ein  $\xi_k$  mit  $|\xi_k| \leq h$ , so dass gilt

$$h^{-1}((s+h)^k - s^k) - k s^{k-1} = \frac{k(k-1)}{2} \cdot (s + \xi_k)^{k-2} \cdot h$$

(Mittelsatzwert). Wegen  $|s + \xi_k| \leq |s| + \delta < 1$  ergibt sich

$$|A(s, h)| \leq \sum_{k \geq 1} \frac{h(k-1)}{2} (|s| + \delta)^k p_k \cdot |h| = 0(h).$$

Für  $h \rightarrow 0$  folgt also

$$\frac{dg}{ds} = \sum_{k \geq 1} k s^{k-1} p_k, \text{ und es gilt } \frac{dg}{ds} \Big|_{s=0} = p_1.$$

Der Beweis für die höheren Ableitungen erfolgt analog.

- (ii) Mit  $EX^k < \infty$  gilt auch  $EX^l < \infty$  ( $1 \leq l < k$ ) und somit  $E(X(X-1) \dots (X-k+1)) < \infty$ . Für  $s \in (0, 1)$  ist  $\frac{d^k}{ds^k} g(s)$  eine nichtnegative monoton wachsende Funktion mit (siehe Teil (i) dieser Aussage)

$$\lim_{s \uparrow 1} \frac{d^k}{ds^k} g(s) \leq E(X(X-1) \dots (X-k+1)) < \infty. \quad (4.30)$$

Es sei  $\varepsilon$  irgendeine positive Zahl und  $j_0$  so groß, dass

$$\sum_{j=j_0+1}^{\infty} j(j-1) \dots (j-k+1) p_j < \frac{\varepsilon}{2}$$

gilt.

Weiterhin sei  $\delta < 0$  so gewählt, dass

$$\sum_{j=k}^{j_0} j(j-1)\dots(j-k+1)s^j p_j > \sum_{j=k}^{j_0} j(j-1)\dots(j-k+1)p_j - \frac{\varepsilon}{2}$$

für alle  $s$  mit  $s \in (1 - \delta, 1]$  richtig ist.

Dann gilt für  $s \in (1 - \delta, 1]$

$$\begin{aligned} \frac{d^k g}{s d^k}(s) &= \sum_{j=k}^{j_0} j(j-1)\dots(j-k+1)s^j p_j + \sum_{j=j_0+1}^{\infty} j(j-1)\dots(j-k+1)s p_j \\ &> \sum_{j=k}^{j_0} j(j-1)\dots(j-k+1)p_j - \frac{\varepsilon}{2} > \sum_{j=k}^{\infty} j(j-1)\dots(j-k+1)p_j - \varepsilon, \end{aligned}$$

und somit haben wir in (4...) das Gleichheitszeichen.

(iii) Nach Voraussetzung und wegen (i) gilt

$$\frac{d^k g}{ds^k}(s) = \frac{d^k h}{ds^k}(s), \quad k \geq 1, s \in (0, \delta).$$

Wegen der Stetigkeit aller Ableitungen von  $g$  und von  $h$  für  $|s| < 1$  folgt

$$\left. \frac{d^k g}{ds^k} \right|_{s=0} = \left. \frac{d^k h}{ds^k} \right|_{s=0}.$$

Aus (4...) ergibt sich nun (iii). □

**Definition 4.39** Die Größe  $f_k := EX(X-1)\dots(X-k+1)$  heißt faktorielles Moment  $k$ -ter Ordnung der Zufallsgröße  $X$ .

Formel (4...) kann man zur Berechnung anderer Momente der Zufallsgröße  $X$  nutzen. Zum Beispiel gilt

$$EX = f_1, D^2 X = EX^2 - (EX)^2 = f_2 + f_1 - f_1^2.$$

**Beispiel 4.40** (Fortsetzung der Beispiele aus 4.1.):

a) Im Fall  $\omega_k = k, k = 1, \dots, N$  ergibt sich

$$g(s) = \frac{1}{N} \sum_{k=1}^N s^k = \frac{1}{N} \cdot \frac{s - s^{N+1}}{1 - s}, \quad s \in [-1, 1)$$

und  $g(1) = 1$ ,

b)  $g(s) = s^{k_0}$  falls  $\omega_0 = k_0 \in N_0$ ,

$$c) g(s) = \sum_{k=0}^n \binom{n}{k} (ps)^k (1-p)^{n-k} = (1 - p(1-s))^n,$$

$$d) g(s) = \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} e^{-\lambda} = \exp(\lambda(s-1)),$$

$$e) g(s) = \sum_{k=0}^{\infty} (qs)^k p = \frac{p}{1-qs} \text{ mit } q = 1 - p,$$

$$f) g(s) = \sum_{k=0}^m \frac{\binom{R}{k} \binom{M-R}{m-k}}{\binom{M}{m}} s^k \text{ ist eine spezielle hypergeometrische Funktion,}$$

$$g) g(s) = \sum_{k=0}^{\infty} \binom{-v}{k} (qs)^k p = \left(\frac{p}{1-qs}\right)^v \text{ mit } q = 1 - p.$$

Der Beweis ist elementar.

## 4.7 Mehrstufige zufällige Versuche

Häufig läuft ein zufälliger Versuch in mehreren Schritten oder Stufen ab. Wir haben dafür bereits Beispiele kennen gelernt (mehrmaliges Werfen einer Münze). In diesem Abschnitt werden wir zunächst ein sehr allgemeines stochastisches Modell zusammengesetzter Versuche konstruieren. Danach konzentrieren wir uns auf den Fall abzählbar vieler Versuchsausgänge, in dem man

einige einfache Berechnungsformeln angeben kann.

Angenommen, der zufällige Versuch besteht aus  $n$  Einzelexperimenten, die nacheinander ausgeführt werden. Die möglichen Ergebnisse  $\omega_k$  des  $k$ -ten Experimentes mögen aus einer Menge  $\Omega_k$  stammen,  $k = 1, \dots, n$ . Das Ergebnis des Gesamtexperimentes wird dann beschrieben durch den Ausgang

$$\omega = (\omega_1, \dots, \omega_n) \in \Omega_1 \times \dots \times \Omega_n.$$

Da  $\omega$  aufgefasst wird als Ergebnis einer zeitlichen Abfolge von Experimenten, nennt man  $\omega$  auch einen "Pfad" oder eine "Trajektorie" des Gesamtversuches.

Wir setzen

$$\Omega := \Omega_1 \times \dots \times \Omega_n = \prod_{k=1}^{n \otimes} \Omega_k.$$

Die mit dem  $k$ -ten Experiment verbundenen Ereignisse bilden eine  $\sigma$ -Algebra  $\mathfrak{A}_k$  von Teilmengen von  $\Omega_k$ . Die  $\sigma$ -Algebra  $\mathfrak{A}$  aller mit dem Gesamtversuch verbundenen Ereignisse enthält natürlich alle Ereignisse der Form  $A := A_1 \times \dots \times A_n$  mit  $A_k \in \mathfrak{A}_k$ ,  $k = 1, \dots, n$ , da man nach Ablauf aller Teilexperimente entscheiden kann, ob ein solches  $A$  eingetreten ist oder nicht.

Wir definieren  $\mathfrak{A}$  als kleinste  $\sigma$ -Algebra von Teilmengen von  $\Omega$ , die alle Ereignisse dieser Form umfasst, also:

$$\mathfrak{A} := \sigma(A_1 \times \dots \times A_n | A_k \in \mathfrak{A}_k, k = 1, \dots, n).$$

**Definition 4.41**  $\mathfrak{A}$  heißt die Produkt- $\sigma$ -Algebra der  $\sigma$ -Algebren  $\mathfrak{A}_k$ ,  $k = 1, \dots, n$ , und wird auch mit  $\prod_{k=1}^{n \otimes} \mathfrak{A}_k$  oder  $\mathfrak{A}_1 \otimes \mathfrak{A}_2 \otimes \dots \otimes \mathfrak{A}_n$  bezeichnet.

Ist  $P$  ein Wahrscheinlichkeitsmaß auf  $\mathfrak{A}$ , so haben wir mit  $(\Omega, \mathfrak{A}, P)$  ein stochastisches Modell eines  $n$ -stufigen zufälligen Versuches.

Das System  $\gamma$  von Ereignissen aus  $\mathfrak{A}$ , definiert durch

$$\gamma := \{A_1 \times \dots \times A_n | A_k \in \mathfrak{A}_k, k = 1, \dots, n\}$$

ist eine Semialgebra mit  $\sigma(\gamma) = \mathfrak{A}$ . Folglich ist  $P$  durch die Angabe seiner Werte auf  $\gamma$  bereits eindeutig festgelegt (Maßtheorie). Das wird uns die Konstruktion des Maßes  $P$  aus einfacheren Größen ermöglichen.

Das  $k$ -te Einzelexperiment  $(\Omega_k, \mathfrak{A}_k, P_k)$  ist in dem Gesamtexperiment  $(\Omega, \mathfrak{A}, P)$  durch

$$\mathfrak{A}_k \ni A_k \longleftrightarrow (\Omega_1 \times \dots \times \Omega_{k-1} \times A_k \times \Omega_{k+1} \times \dots \times \Omega_n) =: A'_k \text{ und}$$

$$P_k(A_k) = P(A'_k), \quad A_k \in \mathfrak{A}_k$$

eingebettet. Die Verteilung  $P$  bestimmt also die "Randverteilungen"  $P_k$  auf  $\mathfrak{A}_k$ .

Aus den  $P_k, k = 1, \dots, n$ , dagegen ist  $P$  im Allgemeinen nicht reproduzierbar. Das gelingt nur in einem Fall, nämlich wenn gilt

$$P(A_1 \times \dots \times A_n) = \prod_{k=1}^n P(A_k), \quad A_k \in \mathfrak{A}_k, k = 1, \dots, n. \quad (4.31)$$

In diesem Fall bezeichnet man  $P$  als das von den  $P_k$  erzeugte *Produktmaß* auf der Produkt- $\sigma$ -Algebra  $\mathfrak{A}$  und schreibt  $P = \prod_{k=1}^n P_k = P_1 \otimes P_2 \otimes \dots \otimes P_n$ .

Im Allgemeinen ist jedoch  $P$  nicht gleich dem Produktionsmaß.

Wir wollen nun für den Fall, dass alle  $\Omega_k$  abzählbar sind, das Maß  $P$  aus einfacheren Kenngrößen konstruieren. Dazu beginnen wir mit einem einfachen Beispiel.

**Beispiel 4.42** In einer Urne mögen sich zwei rote und drei schwarze Kugeln befinden. Wir ziehen auf gut Glück eine der Kugeln und legen sie zusammen mit einer weiteren Kugel derselben Farbe wie die gezogene, in die Urne zurück. Danach wählen wir erneut auf gut Glück eine Kugel.

Das Experiment ist zweistufig mit  $\Omega_1 = \Omega_2 = \{r, s\}$ , seine möglichen Ausgänge sind die Elemente der Menge  $\Omega := \{(r, r), (r, s), (s, r), (s, s)\}$ . Für  $\mathfrak{A}$  wählen wir  $\mathfrak{P}(\Omega)$ . Die zu bestimmende Wahrscheinlichkeitsverteilung  $P$  ist diskret und durch ihre Einzelwahrscheinlichkeiten  $p((r, r)), p((r, s)), p((s, r)), p((s, s))$  eindeutig festgelegt.

$$P(A) = \sum_{\omega: \omega \in A} p(\omega), \quad A \subseteq \Omega. \quad (4.32)$$

Um eine Vorstellung zu bekommen, wie groß diese Einzelwahrscheinlichkeiten im betrachteten Fall sind, erinnern wir an das empirische Gesetz der großen Zahlen, dass bei wachsender Zahl von Versuchsdurchführungen die relative Häufigkeit  $\frac{n(A)}{n}$  eines Ereignisses  $A$  sich der Wahrscheinlichkeit  $P(A)$  immer mehr nähert. Wenn wir die geschilderten Ziehungen sehr oft wiederholen, so wird die relative Häufigkeit, beim jeweils ersten Zug eine rote Kugel zu erhalten, etwa gleich  $\frac{2}{5}$  sein, da in der Urne zwei der fünf Kugeln rot sind. Unter denjenigen Versuchsdurchführungen, bei denen man beim ersten Mal eine rote Kugel zieht, werden sich mit der relativen Häufigkeit von etwa  $\frac{3}{6} = \frac{1}{2}$  beim zweiten Ziehen eine schwarze Kugel ergeben, da sich vor dem zweiten Ziehen drei rote und drei schwarze Kugeln in der Urne befinden. Insgesamt wird also die relative Häufigkeit des Ergebnisses  $(r, s)$  etwa gleich  $\frac{2}{5} \cdot \frac{1}{2} = \frac{1}{5}$  sein. Wir setzen deshalb die Einzelwahrscheinlichkeit  $p((r, s))$  dafür, beim ersten Zug eine rote, beim zweiten Zug eine schwarze Kugel zu erhalten, gleich  $p((r, s)) = \frac{2}{5} \cdot \frac{1}{2} = \frac{1}{5}$ . Analog ergibt sich  $p((r, r)) = \frac{2}{5} \cdot \frac{3}{6} = \frac{1}{5}$ ,  $p((s, r)) = \frac{3}{5} \cdot \frac{2}{6} = \frac{1}{5}$ ,  $p((s, s)) = \frac{3}{5} \cdot \frac{4}{6} = \frac{2}{5}$ . Damit ist unter Beachtung von (4...) eine Verteilung auf  $\mathfrak{B}(\Omega)$  definiert.

Für die Randverteilungen  $P_1$  und  $P_2$  des ersten bzw. zweiten Zuges ergibt sich

$$P_1(\{r\}) = P(\{r\} \times \{r, s\}) = p((r, r)) + p((r, s)) = \frac{2}{5},$$

$$P_1(\{s\}) = 1 - P_1(\{r\}) = \frac{3}{5}$$

$$P_2(\{r\}) = P(\{r, s\} \times \{r\}) = p((r, r)) + p((s, r)) = \frac{2}{5}, \quad P_2(\{s\}) = 1 - P_2(\{r\}) = \frac{3}{5}.$$

## Erste Pfadregel

Im Folgenden seien alle  $\Omega_k, k = 1, \dots, n$ , höchstens abzählbar. Das erste der  $n$  Experimente ende mit der Wahrscheinlichkeit  $p^{(1)}(\omega_1)$  mit dem Ausgang  $\omega_1 \in \Omega_1$ . Die Wahrscheinlichkeitsverteilung für das zweite Experiment hängt

i.A. vom Ausgang des ersten ab, wir bezeichnen ihre Einzelwahrscheinlichkeiten mit  $p_{\omega_1}^{(2)}(\omega_2)$ . Analog hängt die Wahrscheinlichkeitsverteilung des  $k$ -ten Experiments vom bisherigen Verlauf ab:  $p_{\omega_1\omega_2\dots\omega_{k-1}}^{(k)}(\omega_k)$ . Als Wahrscheinlichkeit  $p(\omega)$  für den Ausgang des Gesamtexperimentes  $\omega = (\omega_1, \dots, \omega_n)$  definieren wir analog zum obigen Beispiel:

$$p(\omega) := p^{(1)}(\omega_1)p_{\omega_1}^{(2)}(\omega_2) \cdot \dots \cdot p_{\omega_1\dots\omega_{n-1}}^{(n)}(\omega_n). \quad (4.33)$$

Offenbar ist  $p(\omega) \geq 0$  und

$$\sum_{\omega \in \Omega} p(\omega) = \sum_{\omega_1 \in \Omega_1} p^{(1)}(\omega_1) \sum_{\omega_2 \in \Omega_2} p_{\omega_1}^{(2)}(\omega_2) \cdot \dots \cdot \sum_{\omega_n \in \Omega_n} p_{\omega_1\dots\omega_{n-1}}^{(n)}(\omega_n) = 1,$$

da nach Definition gilt:

$$p^{(1)}(\omega_1) \geq 0 \text{ und } \sum_{\omega_1 \in \Omega_1} p^{(1)}(\omega_1) = 1, \text{ sowie}$$

$$p_{\omega_1\omega_2\dots\omega_{k-1}}^{(k)}(\omega_k) \geq 0 \text{ und}$$

$$\sum_{\omega_k \in \Omega_k} p_{\omega_1\omega_2\dots\omega_{k-1}}^{(k)}(\omega_k) = 1, \quad k \geq 2.$$

Diese Regel (4....) zur Bestimmung der Einzelwahrscheinlichkeiten  $p(\omega)$  heißt auch "*Erste Pfadregel*".

Die Wahrscheinlichkeit  $p(\omega)$  des Pfades  $\omega = (\omega_1, \omega_2, \dots, \omega_n)$  wird also mittels der zumeist bekannten oder einfacher zu bestimmenden Größen  $p_{\omega_1\dots\omega_{k-1}}^{(k)}(\omega_k)$  berechnet, die die Wahrscheinlichkeiten angeben, dass beim  $k$ -ten Versuch der Ausgang  $\omega_k$  erscheint, wenn im bisherigen Verlauf der Versuche  $\omega_1, \omega_2, \dots, \omega_{k-1}$  aufgetreten sind.

## Zweite Pfadregel

Als "*Zweite Pfadregel*" bezeichnet man die Formel zur Bestimmung von  $P(A)$ , wie sie für jede diskrete Verteilung gilt:

$$P(A) = \sum_{\omega \in A} p(\omega), \quad A \subseteq \Omega. \quad (4.34)$$

Als Fazit halten wir fest, dass die Wahrscheinlichkeitsverteilung  $P$  in einem  $n$ -stufigen zufälligen Versuch sich aus den meist einfacher zu bestimmenden Wahrscheinlichkeiten

$$p^{(1)}(\omega_1), p_{\omega_1 \dots \omega_{k-1}}^{(k)}(\omega_k), k = 2, 3, \dots, n, \omega_i \in \Omega_i, i = 1, \dots, n$$

mit Hilfe der ersten und zweiten Pfadregel bestimmen lässt. Die Verteilung  $(p^{(1)}(\omega_1), \omega_1 \in \Omega_1)$  nennt man *Anfangsverteilung*, jede der Verteilungen  $p_{\omega_1 \dots \omega_{k-1}}^{(k)}(\omega_k), \omega_k \in \Omega_k$ , eine *Übergangsverteilung*.

#### Beispiel 4.43 (Polya'sches Urnenschema)

In einer Urne liegen  $R$  rote und  $S$  schwarze Kugeln,  $R + S = N$ . Auf rein zufällige Weise wird  $n$ -mal nacheinander eine Kugel entnommen, jedes Mal wieder zurückgelegt und  $c$  Kugeln derselben Farbe hinzu gefügt.

Die Anzahl der roten Kugeln in der Urne nach Abschluss den Ziehungen und des Zurücklegens ist eine Zufallsgröße. Gefragt ist ihre Wahrscheinlichkeitsverteilung. Der beschriebene Mechanismus wird als einfaches Modell der Ausbreitung infektiöser Krankheiten in Populationen von Individuen angesehen und heißt *Polya'sches Urnenschema*.

Wir beschreiben das Polya'sche Urnenschema als mehrstufigen zufälligen Versuch. Dazu setzen wir

$$\Omega_k = \{0, 1\}, \mathfrak{A}_k = \mathfrak{P}(\Omega_k), k = 1, \dots, n$$

und vereinbaren  $\omega_k = 0$ , falls beim  $k$ -ten Ziehen eine schwarze,  $\omega_k = 1$ , falls beim  $k$ -ten Ziehen eine rote Kugel erscheint.

Es sei  $\Omega := \prod_{k=1}^n \Omega_k$ ,  $\mathfrak{A} := \mathfrak{P}(\Omega)$ ,  $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ .

Die Anzahl der roten Kugeln, die bis zum (einschließlich)  $k$ -ten Ziehen gezogen wurden, werde mit  $R_k$  bezeichnet.

Die Zufallsgrößen  $R_k$  und  $S_k$  habe die möglichen Werte  $0, 1, \dots, k$ . Nach Definition gilt  $R_k(\omega) = \sum_{j=1}^k \omega_j$ ,  $\omega \in \Omega, k = 1, \dots, n$ .

Wir setzen  $R_0(\omega) \equiv 0$ .

Die Anzahl  $S_k$  der schwarzen Kugeln, die bis zum  $k$ -ten Ziehen erschienen sind, ist folglich gleich  $k - R_k(\omega) =: S_k(\omega)$ . Auch hier definieren wir  $S_0(\omega) \equiv 0$ .

Nach den beiden Pfadregeln für mehrstufige Versuche gilt:

$$(R_n = j) = \sum_{\substack{\omega \in \Omega \\ R_n(\omega) = j}} P(\{\omega\})$$

$$P(\{\omega\}) = p^{(1)}(\omega_1) p_{\omega_1}^{(2)}(\omega_2) \cdot \dots \cdot p_{\omega_1 \dots \omega_{n-1}}^{(n)}(\omega_n) \quad (9) \quad (4.35)$$

Die Übergangswahrscheinlichkeit  $p_{\omega_1 \dots \omega_{k-1}}^{(k)}(\omega_k)$  ist dabei die Wahrscheinlichkeit dafür, dass im  $k$ -ten Versuch eine rote ( $\omega_k = 1$ ) bzw. eine schwarze ( $\omega_k = 0$ ) Kugel gezogen wird, wobei  $(\omega_1, \dots, \omega_{k-1})$  den bisherigen Verlauf der Ziehungen darstellt.

**Lemma 4.44** *Es gilt*

$$p_{\omega_k \dots \omega_{k-1}}^{(k)}(\omega_k) = \frac{(R + R_{k-1}(\omega) \cdot c)^{\omega_k} (S + S_{k-1}(\omega) \cdot c)^{1-\omega_k}}{R + S + (k-1)c}, \quad k = 1, \dots, n. \quad (4.36)$$

Beweis: Vor dem  $k$ -ten Ziehen einer Kugel befinden sich  $R + S + (k-1)c$  Kugeln in der Urne, von denen  $R + R_{k-1}(\omega)c$  rot und  $S + S_{k-1}(\omega) \cdot c$  schwarz sind.  $\square$

Beim Einsetzen von (4.36) in (4.35) kann man das entstehende Produkt vereinfachen. Dazu nutzen wir das folgende

**Lemma 4.45** *Für alle  $\omega$  aus  $\Omega$  gilt*

$$\prod_{k=0}^{n-1} (R + R_k(\omega)c)^{\omega_{k+1}} = \prod_{k=1}^{R_n(\omega)} (R + (k-1)c) \quad \text{und} \quad (11) \quad (4.37)$$

$$\prod_{k=0}^{n-1} (S + S_k(\omega)c)^{1-\omega_{k+1}} = \prod_{k=1}^{S_n(\omega)} (S + (k-1)c), \quad (4.38)$$

wobei  $\prod_{k=1}^0 (\dots) = 1$  gesetzt wird.

Beweis: Wir zeigen nur, dass die erste Gleichung richtig ist, die zweite beweist man analog.

Für  $n = 1$  ist die Gleichheit in (11) offensichtlich. Die Beziehung (11) gelte für  $n = m$ . Dann haben wir für  $n = m + 1$ :

$$\begin{aligned} \prod_{k=0}^m (R + R_k(\omega)c)^{\omega_{k+1}} &= \prod_{k=0}^{m-1} (R + R_k(\omega)c)^{\omega_{k+1}} \cdot (R + R_m(\omega)c)^{\omega_{m+1}} \\ &= \prod_{k=1}^{R_m(\omega)} (R + (k-1)c) (R + R_m(\omega)c)^{\omega_{m+1}}. \end{aligned}$$

Der letzte Faktor ist gleich  $(R + (R_{m+1}(\omega) - 1)c)$  falls  $\omega_{m+1} = 1$  (wegen  $R_{m+1}(\omega) = R_m(\omega) + \omega_{m+1}$ ), und gleich Eins, falls  $\omega_{m+1} = 0$ . Wegen  $R_{m+1}(\omega) = R_m(\omega)$  im letzteren Fall ist somit (11) für  $n = m + 1$  und jedes  $\omega$  richtig.  $\square$

**Folgerungen 4.46** Für jedes  $\omega \in \Omega$  gilt

$$P(\{\omega\}) = \frac{\prod_{j=1}^{R_n(\omega)} (R + (j-1)c) \prod_{j=1}^{n-R_n(\omega)} (S + (j-1)c)}{\prod_{j=1}^n (R + S + (j-1)c)} \quad (4.39)$$

$$P(R_n = l) = \sum_{\omega: R_n(\omega)=l} P(\{\omega\}), l = 0, 1, \dots, n.$$

Alle  $\omega$  mit  $R_n(\omega) = l$  haben wegen (13) die gleiche Wahrscheinlichkeit  $P(\{\omega\})$ . Es gibt insgesamt  $\binom{n}{l}$  von ihnen (das ist die Anzahl aller möglichen Anordnungen von  $l$  Einsen und  $(n-l)$  Nullen in  $\omega = (\omega_1, \dots, \omega_n)$ ). Also folgt

$$P(R_n = l) = \frac{\binom{n}{l} \prod_{j=1}^l (R + (j-1)c) \prod_{j=1}^{n-l} (S + (j-1)c)}{\prod_{j=1}^n (R + S + (j-1)c)} \quad (4.40)$$

$$l = 0, 1, \dots, n.$$

Diese Wahrscheinlichkeitsverteilung von  $R_n$  heißt "Polya'sche Verteilung" auf  $\{0, 1, \dots, n\}$  mit den Parametern  $R, S$  und  $c$ .

Damit haben wir:

**Aussage 4.47** Die Anzahl der bis zum  $n$ -ten Zug gezogenen roten Kugeln im Polya'schen Urnenschema hat die durch (14) gegebene Polya'sche Verteilung mit den Parametern  $R, S$  und  $c$ .

Spezialfälle:

$c = 0$ : In diesem Fall wird nach jedem Ziehen nur die gezogene Kugel selbst zurückgelegt. Es ergibt sich eine Binomialverteilung mit den Parametern  $n$  und  $p = \frac{R}{R+S}$ ,

$c = -1$ : Jetzt wird nach jedem Ziehen die gezogene Kugel nicht zurückgelegt und auch keine weitere in die Urne gelegt. Wir erhalten eine hypergeometrische Verteilung mit den Parametern  $M = R + S, m = n$ . (Wir setzen in diesem Fall der Einfachheit halber  $m \leq \min(R, S)$  voraus.)

Es sei  $A_k =$  "bei  $k$ -ten Zug erscheint eine rote Kugel",  $1 \leq k \leq n$ . Wir wollen  $P(A_k)$  berechnen und beweisen zunächst das folgende Lemma.

**Lemma 4.48** Die Ereignisse  $A_1, A_2, \dots, A_n$  sind austauschbar im Sinne von:

Beweis: Wegen (4....) gilt für alle Tupel  $(i_1, \dots, i_l)$  mit  $1 \leq i_1 < \dots < i_l \leq n$  die Gleichung

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_l}) = \sum_{\substack{\omega \in \Omega \\ \omega_{i_j} = 1, j=1, \dots, l}} P(\{\omega\}) = \sum_{\substack{\omega \in \Omega: \\ \omega_{i_j} = 1, j=1, \dots, l}} H_n(R_n(\omega)) = \sum_{m=l}^n \sum_{\substack{\omega \in \Omega \\ \omega_{i_j} = 1, j=1, \dots, l \\ R_n(\omega) = m}} H_n(m) \quad (4.41)$$

mit

$$H_n(m) = \frac{\prod_{j=1}^m (R + (j-1)c) \prod_{j=1}^{n-m} (S + (j-1)c)}{\prod_{j=1}^n (R + S + (j-1)c)}, \quad 0 \leq m \leq n.$$

Die Summanden  $H_n(m)$  der rechten Seite von (4.41) sind für alle Tupel  $(i_1, \dots, i_l)$  mit derselben Länge  $l$  dieselben, und die Anzahl der  $\omega$  mit  $\omega_{i_k} = 1, k = 1, \dots, l$ , und  $R_n(\omega) = m$  ist bei  $m$  mit  $l \leq m \leq n$  gleich  $\binom{n-l}{m-l}$ , unabhängig von den Werten  $i_1, i_2, \dots, i_l \in \{1, 2, \dots, n\}$ . Somit gilt  $P(A_{i_1} \cap \cdots \cap A_{i_l}) = P(A_1 \cap \cdots \cap A_l)$ , die  $(A_k, k \leq n)$  sind also austauschbar im Sinne von .....

Mit Hilfe dieses Lemmas kommt man zu der zunächst überraschenden

$$P(A_k) = P(A_1) = \frac{R}{R+S}, \quad k = 1, 2, \dots, n$$

unabhängig von  $k$  und  $c$ .

**Aussage 4.49** Für alle  $k = 1, 2, \dots, n$  gilt

$$P(A_k) = P(A_1) = \frac{R}{R+S}$$

Beweis: Wegen der Austauschbarkeit der  $A_1, A_2, \dots, A_n$  folgt insbesondere  $P(A_k) \equiv P(A_1)$ , und  $P(A_1) = \frac{R}{R+S}$  ist offensichtlich, da das Ziehen der ersten Kugel auf gut Glück erfolgt.  $\square$



# Kapitel 5

## Bedingte Wahrscheinlichkeiten und Unabhängigkeit

Mitunter erhält man über das Ergebnis eines zufälligen Versuches Vorinformationen. Dann entsteht die Frage, wie sich für den Betrachter, den man als "Insider" bezeichnen könnte, die Wahrscheinlichkeiten der mit dem Versuch verbundenen Ereignisse ändern.

### 5.1 Definition und Eigenschaften

Es sei  $(\Omega, \mathfrak{A}, P)$  ein Wahrscheinlichkeitsraum. Wir verbinden mit  $(\Omega, \mathfrak{A}, P)$  wieder die Vorstellung, es handle sich um einen zufälligen Versuch mit möglichen Versuchsergebnissen  $\omega$  aus  $\Omega$ , mit den zufälligen Ereignissen  $A$  aus  $\mathfrak{A}$ , die mit dem Versuch verbunden sind, und mit der Wahrscheinlichkeitsverteilung  $P$  auf  $\mathfrak{A}$ , die jedem Ereignis  $A$  aus  $\mathfrak{A}$  eine Wahrscheinlichkeit  $P(A)$  zuordnet.

Es seien  $A$  und  $B$  zwei Ereignisse, d.h.  $A, B \in \mathfrak{A}$ .

Angenommen, vor Bekanntwerden des Ausganges dieses zufälligen Versuches erhält man als Vorinformation die Nachricht, dass das Ereignis  $B$  eingetreten ist. Wie ändert sich dadurch die Wahrscheinlichkeit  $P(A)$  für das Eintreten von  $A$ ? Um davon eine Vorstellung zu bekommen, nehmen wir wieder das empirische Gesetz der großen Zahlen zu Hilfe.

Wir stellen uns vor, der Versuch wird  $n$ -mal unter gleichartigen Bedingun-

gen und unabhängig voneinander ausgeführt. Dann gilt entsprechend diesem Erfahrungsgesetz

$$\frac{n(A)}{n} \approx P(A), \frac{n(B)}{n} \approx P(B).$$

Wir betrachten von dieser Versuchsreihe der Länge  $n$  jetzt nur diejenigen Versuche, bei denen  $B$  eingetreten ist. Davon gibt es  $n(B)$  an der Zahl. Wir können diese Reihe als eine neue Reihe von Versuchen ansehen, bei denen immer  $B$  eintritt, das Eintreten von  $B$  erscheint damit als eine zusätzliche Versuchsbedingung.

Will man die Wahrscheinlichkeit von  $A$  im Rahmen dieser neuen Versuche (unter der Bedingung, dass  $B$  eintritt) berechnen, so wäre ein Kandidat die Anzahl aller Versuche, bei denen  $A$  und  $B$  eintreten, geteilt durch die Anzahl aller Versuche, bei denen  $B$  eintritt:

$$\frac{n(A \cap B)}{n(B)} = \frac{n(A \cap B)}{n} : \frac{n(B)}{n}.$$

Das führt uns auf folgende Definition:

**Definition 5.1** *Es seien  $(\Omega, \mathfrak{A}, P)$  ein Wahrscheinlichkeitsraum,  $A, B \in \mathfrak{A}$  mit  $P(B) > 0$ . Dann heißt*

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \tag{5.1}$$

die bedingte Wahrscheinlichkeit von  $A$  unter der Bedingung, dass das Ereignis  $B$  eintritt (kurz: die bedingte Wahrscheinlichkeit von  $A$  unter der Bedingung  $B$ ).

### Beispiel 5.2

a) Beim Werfen zweier Würfel betrachten wir die Ereignisse

$A$ : = "Der zweite Würfel zeigt eine gerade Zahl" und

$B$ : = " Die Augensumme ist acht". Dann gilt

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{(6, 2), (4, 4), (2, 6)\})}{P(\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\})} = \frac{3}{5} \neq P(A) = \frac{1}{2}.$$

b) Bei 6800 rein zufällig aus der Bevölkerung ausgewählten Personen wurden Augen- und Haarfarbe registriert. Das Ergebnis ist in folgender Tabelle enthalten.

$k$ (Augen) \ / $l$ (Haar)	1 (hellblond)	2 (dunkelblond)	3 (schwarz)	4 (rot)	Zusammen $n_k$ .
1 (blau)	1768	807	189	47	2811
2 (grau oder grün)	946	1387	746	53	3132
3 (braun)	115	438	288	16	857
Zusammen $n_l$	2829	2632	1223	116	6800

(aus: Fisz, M., Wahrscheinlichkeitsrechnung und Mathematische Statistik, Deutscher Verlag der Wissenschaften, Berlin 1965.)

Die Tabelle erlaubt die Berechnung bedingter Wahrscheinlichkeiten, die die Zusammenhänge zwischen beiden Merkmalen Haar- und Augenfarbe deutlicher herausstellen als die Originaltabelle. Zum Beispiel gilt:

$A_3 :=$  "Eine rein zufällig ausgewählte Person hat braune Augen"

$B_3 :=$  "Eine rein zufällig ausgewählte Person hat schwarze Haare"

$$P(A_3|B_3) = \frac{P(A_3 \cap B_3)}{P(B_3)} = \frac{\frac{288}{6800}}{\frac{1223}{6800}} = \frac{288}{1223} = 0,234,$$

dagegen gilt die "absolute" Wahrscheinlichkeit

$$P(A_3) = 857/6800 = 0,126.$$

Entsprechend berechne man z.B.  $P(A_2|B_3)$ ,  $P(B_3|A_3)$  usw..

Es seien  $A, B, A_k \in \mathfrak{A}, k \geq 1$ , mit  $P(B) > 0$ . Dann bestehen folgende Eigenschaften, die sich unmittelbar aus der Definition ergeben:

**Aussage 5.3** (*Eigenschaften bedingter Wahrscheinlichkeiten*)

1.  $0 \leq P(A|B) \leq 1$
2.  $P(\Omega|B) = 1, \quad P(\emptyset|B) = 0$
3.  $P\left(\bigcup_{k=1}^{\infty} A_k|B\right) = \sum_{k=1}^{\infty} P(A_k|B),$

*falls  $A_k \cap A_l = \emptyset$  für alle  $k, l$  mit  $k \neq l$ .*

**Bemerkung 5.4** Aus 1. - 3. folgt, dass die Abbildung

$$A \rightarrow P(A|B), \quad A \in \mathfrak{A},$$

ebenfalls eine Wahrscheinlichkeitsverteilung auf  $\mathfrak{A}$  ist. Der Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P(\cdot|B))$  ist das mathematische Modell für den ursprünglichen Versuch mit der zusätzlichen Bedingung, dass  $B$  eintritt.

Unmittelbar aus der Definition der bedingten Wahrscheinlichkeit ergibt sich weiterhin

**Aussage 5.5** (*Multiplikationssatz*)

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A), \quad (5.2)$$

und allgemeiner:

$$P(A_1 \cap \dots \cap A_n) =$$

$$P(A_n|A_1 \cap \dots \cap A_{n-1})P(A_{n-1}|A_1 \cap \dots \cap A_{n-2}) \dots P(A_2|A_1) \cdot P(A_1) \quad (5.3)$$

*falls  $P(A_1 \cap \dots \cap A_{n-1}) > 0$ .*

Der Beweis von (5.3) erfolgt mittels vollständiger Induktion unter Verwendung von (5.2).

**Aussage 5.6** (*"Satz von der totalen Wahrscheinlichkeit"*)

Es sei  $(Z_i, i \in I)$  mit  $I \subseteq N_0$  eine Zerlegung von  $\Omega$  in Ereignisse  $Z_i$  aus  $\mathfrak{A}$ , die alle eine positive Wahrscheinlichkeit besitzen:

$$Z_i \cap Z_j = \emptyset \text{ falls } i \neq j; j \in I, \bigcup_{i \in I} Z_i = \Omega, P(Z_i) > 0, \quad i \in I.$$

(Bei jeder Versuchsdurchführung tritt also eines und nur eines der Ereignisse  $Z_i$  ein.)

Dann gilt für jedes Ereignis  $A \in \mathfrak{A}$  :

$$P(A) = \sum_{i \in I} P(A|Z_i)P(Z_i). \quad (5.4)$$

Beweis:

$$\begin{aligned} P(B) &= P(B \cap \Omega) = P(B \cap (\bigcup_{i \in I} Z_i)) = \\ &P(\bigcup_{i \in I} (B \cap Z_i)) = \sum_{i \in I} P(B \cap Z_i) = \sum_{i \in I} P(B|Z_i)P(Z_i). \end{aligned}$$

□

**Folgerungen 5.7** Wenn  $Z_1 = A, Z_2 = \bar{A}, 0 < P(A) < 1$  gilt, so ist

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}). \quad (5.5)$$

**Beispiel 5.8** Aus einer Urne mit  $R$  roten und  $S$  schwarzen Kugeln ( $R + S = M$ ) werden auf gut Glück nacheinander und ohne Zurücklegen zwei Kugeln gezogen. Wir definieren:

$A_1 :=$  "Die erste Kugel ist rot",  $A_2 :=$  "Die zweite Kugel ist rot".

Wie groß sind  $P(A_1)$  und  $P(A_2)$ ?

Die Wahrscheinlichkeit  $P(A_1)$  ist einfach zu bestimmen, da es sich beim ersten Ziehen um ein Laplace-Experiment handelt:

$$P(A_1) = \frac{R}{M}.$$

Wie groß die Wahrscheinlichkeit  $P(A_2)$  ist, ist nicht unmittelbar ersichtlich. Zu ihrer Berechnung greifen wir auf (5.6) zurück:

$$P(A_2) = P(A_2|A_1)P(A_1) + P(A_2|\bar{A}_1)P(\bar{A}_1)$$

Wir wissen bereits, dass  $P(A_1) = \frac{R}{M}$  und  $P(\bar{A}_1) = 1 - P(A_1) = \frac{S}{M}$  gilt.

Die bedingten Wahrscheinlichkeiten  $P(A_2|A_1)$  und  $P(A_2|\bar{A}_1)$  sind ebenfalls einfach zu bestimmen:

$$P(A_2|A_1) = \frac{R-1}{M-1} \quad , \quad P(A_2|\bar{A}_1) = \frac{R}{M-1}$$

Folglich ergibt sich

$$P(A_2) = \frac{R-1}{M-1} \cdot \frac{R}{M} + \frac{R}{M-1} \cdot \frac{S}{M} = \frac{R}{M}.$$

Diese Wahrscheinlichkeit ist genauso groß wie  $P(A_1)$ . Man beachte, dass man bei der Berechnung von  $P(A_2)$  nicht annimmt, dass bekannt ist, ob  $A_1$  eingetreten ist oder nicht.

Es seien  $(Z_i, i \in I)$  eine Zerlegung von  $\Omega$  wie in Aussage 5.6 und  $B \in \mathfrak{A}$  mit  $P(B) > 0$ .

**Aussage 5.9** (*„Bayes’sche Formel“*):

*Es gilt für jedes  $j \in I$ :*

$$P(Z_j|B) = \frac{P(B|Z_j)P(Z_j)}{\sum_{i \in I} P(B|Z_i)P(Z_i)}.$$

Beweis: Nach Definition 5.1 gilt

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

Auf den Nenner wird nunmehr der Satz der totalen Wahrscheinlichkeit angewandt, Aussage 5.6.  $\square$

**Beispiel 5.10** Reihenuntersuchung auf Tuberkulose (Tbc)

$A :=$  "Der Test auf Tbc ist beim untersuchten Patienten positiv"

$B :=$  "Der untersuchte Patient leidet unter Tbc"

Aus langen Testserien weiß man

$P(B) = \frac{1}{109}$ ,  $P(A|B) = 0,96$ ,  $P(\bar{A}|\bar{B}) = 0,92$  und somit insbesondere

$P(A|\bar{B}) = 0,08$ .

Mit  $Z_1 = B, Z_2 = \bar{B}$  ergibt sich

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})} = \frac{0,96 \cdot \frac{1}{109}}{0,96 \cdot \frac{1}{109} + 0,08 \cdot \frac{108}{109}} = 0,1.$$

Das bedeutet für die gegebenen Erfahrungswerte, dass bei positivem Testausgang die Wahrscheinlichkeit, dass der untersuchte Patient tatsächlich an Tbc leidet, nur 0,1 beträgt.

## Bedingte Wahrscheinlichkeiten im Laplace-Modell

Es sei  $(\Omega, \mathcal{P}(\mathfrak{A}), P)$  ein Laplace-Modell mit  $\#\Omega = N$ .  
Weiterhin seien  $A, B \subseteq \Omega$  mit  $N(B) = \#B > 0$ .

Angenommen, wir wissen bereits, dass das Ereignis  $B$  bei diesem zufälligen Versuch eintritt. Dann gibt es für uns nur noch  $N(B)$  mögliche Versuchsausgänge  $\omega$ . Alle diese Ausgänge sehen wir wieder als gleich wahrscheinlich an. Bei  $N(A \cap B) = \#(A \cap B)$  von ihnen tritt auch  $A$  ein.

Also gilt

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

Die Definition der bedingten Wahrscheinlichkeit  $P(B|A)$  ist also nicht nur aus der Sicht des empirischen Gesetzes der großen Zahlen, sondern auch aus der Sicht der Laplace-Modelle vernünftig.

**Beispiel 5.11** Im Nachbarraum werden zwei reguläre Würfel geworfen. Wir erhalten die Nachricht, dass die Augensumme kleiner als sieben ist. Wie groß ist unter dieser Bedingung die Wahrscheinlichkeit dafür, dass mindestens einer der Würfel eine Eins zeigt?  $\left( = \frac{3}{5} \right)$

## Bedingte Wahrscheinlichkeiten in mehrstufigen Versuchen

Wir haben in Abschnitt 4.7 die Wahrscheinlichkeiten  $P(A)$  für Ereignisse  $A$  berechnet, die mit mehrstufigen zufälligen Versuchen zusammenhängen. Mit den dortigen Bezeichnungen gilt:

$$P(A) = \sum_{\omega \in A} P(\{\omega\}) \text{ und} \tag{5.6}$$

$$P(\{\omega\}) = p^{(1)}(\omega_1)p_{\omega_1}^{(2)}(\omega_2) \cdots p_{\omega_1 \cdots \omega_{n-1}}^{(n)}(\omega_n) \tag{5.7}$$

für  $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ .

Dabei bildeten die Größen  $p^{(1)}(\omega_1)$  die Anfangsverteilung und Größen  $p_{\omega_1 \dots \omega_{t-1}}^{(k)}(\omega_k)$  die vorgegebene "Übergangswahrscheinlichkeiten", aus denen  $P(\{\omega\})$  berechnet wird. Auf der Grundlage der so konstruierten Wahrscheinlichkeitsverteilung  $P$  bestimmen wir nunmehr gewisse bedingte Wahrscheinlichkeiten.

Wir beginnen mit einem zweistufigen Versuch mit höchstens abzählbar vielen Ausgängen. Dann haben wir

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

$$P(\{\omega\}) = p^{(1)}(\omega_1)p_{\omega_1}^{(2)}(\omega_2), \quad \omega = (\omega_1, \omega_2),$$

wobei die Anfangsverteilung  $p^{(1)}(\omega_1), \omega_1 \in \Omega_1$ , und die Übergangswahrscheinlichkeiten  $p_{\omega_1}^{(2)}(\omega_2), \omega_1 \in \Omega_1, \omega_2 \in \Omega_2$ , gegeben seien.

Setzt man

$A := \{\omega \in \Omega \mid \text{In der zweiten Versuchsstufe erscheint der Ausgang } \omega_2\}$  und

$B := \{\omega \in \Omega \mid \text{In der ersten Versuchsstufe erscheint der Ausgang } \omega_1\}$ ,

so gilt

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{\sum_{\omega \in A \cap B} P(\{\omega\})}{\sum_{\omega \in B} P(\{\omega\})} = \\ &= \frac{p^{(1)}(\omega_1)p_{\omega_1}^{(2)}(\omega_2)}{\sum_{\omega'_2 \in \Omega_2} p^{(1)}(\omega_1)p_{\omega_1}^{(2)}(\omega'_2)} = p_{\omega_1}^{(2)}(\omega_2). \end{aligned} \tag{5.8}$$

Das heißt, die Wahrscheinlichkeit  $p_{\omega_1}^{(2)}(\omega_2)$  ist die bedingte Wahrscheinlichkeit  $P(A|B)$  bezüglich des Maßes  $P$ , was durchaus der Anschauung und Konstruktion entspricht.

Wir wollen im Folgenden zeigen, dass in jedem  $n$ -stufigen Versuch auch die Größen  $p_{\omega_1 \dots \omega_{k-1}}^{(k)}(\omega_k)$  als bedingte Wahrscheinlichkeiten bezüglich der Verteilung  $P$  auffassen lassen.

Definiert man

$A := \{\omega \in \Omega \mid \text{In der } k\text{-ten Versuchsstufe erscheint der Versuchsausgang } \omega_k\}$

und

$B := \{\omega \in \Omega \mid \text{In den ersten } k-1 \text{ Versuchen erscheinen jeweils } \omega_1, \omega_2, \dots, \omega_{k-1}\}$ ,  
so erhalten wir

### Aussage 5.12

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{\sum_{\omega \in A \cap B} P(\{\omega\})}{\sum_{\omega \in B} P(\{\omega\})} = \\ &= p_{\omega_1 \dots \omega_{k-1}}^{(k)}(\omega_k). \end{aligned}$$

Beweis: Wegen Formel (5.9) haben wir

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{\sum_{\omega \in A \cap B} P(\{\omega\})}{\sum_{\omega \in B} P(\{\omega\})} = \\ &= \frac{\sum_{\omega'_{k+1}, \dots, \omega'_n} p^{(1)}(\omega_1) \dots p_{\omega_1}^{(k)} \dots p_{\omega_{k-1}}(\omega_k) p_{\omega_1 \dots \omega_k}^{(k+1)}(\omega'_{k+1}) \dots p_{\omega_1 \dots \omega_{n-1}}^{(n)}(\omega'_n)}{\sum_{\omega'_k, \omega'_{k+1}, \dots, \omega'_n} p^{(1)}(\omega_1) \dots p_{\omega_1 \dots \omega_{k-1}}(\omega'_k) p_{\omega_1 \dots \omega'_k}^{(k+1)}(\omega'_{k+1}) \dots p_{\omega_1 \dots \omega'_{n-1}}^{(n)}(\omega'_n)} = \\ &= p_{\omega_1 \dots \omega_{k-1}}^{(k)}(\omega_k). \end{aligned}$$

□

Das bedeutet, die "Bausteine"  $p_{\omega_1 \dots \omega_{k-1}}^{(k)}(\omega_k)$ , aus denen sich  $P$  zusammen mit der Anfangsverteilung  $p^{(1)}(\cdot)$  ergibt, erweisen sich aus der Sicht von  $P$  gerade

als die bedingte Wahrscheinlichkeit dafür, dass der  $k$ -te Versuch mit  $\omega_k$  endet, unter der Bedingung, dass in den vorangegangenen Versuchen die Resultate  $\omega_1, \omega_2, \dots, \omega_{k-1}$  auftraten.

## 5.2 Unabhängigkeit

Wenn die Vorinformation, dass das Ereignisse  $B$  eintreten wird, die Wahrscheinlichkeit des Eintretens des Ereignisses  $A$  nicht verändert, so sagt man,  $A$  und  $B$  seien voneinander unabhängige Ereignisse.

Den Begriff der Unabhängigkeit kann man erweitern auf Zufallsgrößen und auf Ereignissysteme wie  $\sigma$ -Algebren. Er steht im Mittelpunkt der klassischen Wahrscheinlichkeitstheorie und Mathematischen Statistik und liegt klassischen Gesetzen der großen Zahlen und zentralen Grenzwertsätzen zugrunde, die wir später kennen lernen werden.

### 5.2.1 Unabhängigkeit von Ereignissen

Es seien  $(\Omega, \mathfrak{A}, P)$  ein Wahrscheinlichkeitsraum und  $A, B \in \mathfrak{A}$  mit  $P(B) \in (0, 1)$ .

Im Allgemeinen gilt  $P(A|B) \neq P(A)$ , d.h., die Kenntnis, dass  $B$  eingetreten ist, beeinflusst die Wahrscheinlichkeit des Eintretens von  $A$ .

In manchen Fällen gilt allerdings  $P(A|B) = P(A)$ . Das bedeutet, das Wissen, dass  $B$  eintritt, verändert die Wahrscheinlichkeit von  $A$  nicht. In diesem Fall gilt übrigens auch

$$\begin{aligned} P(A|\bar{B}) &= \frac{P(A \cap \bar{B})}{P(\bar{B})} = \frac{P(A) - P(A \cap B)}{P(\bar{B})} \\ &= \frac{P(A) - P(A|B)P(B)}{P(\bar{B})} = P(A). \end{aligned} \tag{5.9}$$

**Definition 5.13** Zwei Ereignisse  $A, B \in \mathfrak{A}$  heißen voneinander stochastisch unabhängig (bezüglich der Wahrscheinlichkeitsverteilung  $P$ ), falls  $P(A \cap B) = P(A)P(B)$  gilt. (Im Fall  $P(B) > 0$  bedeutet das  $P(A|B) = P(A)$ , und für  $P(A) > 0$  ergibt sich auch  $P(B|A) = P(B)$ .)

Die Unabhängigkeit ist eine symmetrische Eigenschaft bezüglich  $A$  und  $B$ . Statt stochastisch unabhängig sagt man auch einfach unabhängig.

**Aussage 5.14**  $A, B$  unabhängig  $\iff A, \bar{B}$  unabhängig.

Beweis:  $P(A \cap B) = P(A)P(B) \iff P(A \cap \bar{B}) =$

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

**Folgerungen 5.15**  $A, B$  unabhängig  $\iff A, \bar{B}$  unabhängig  $\iff \bar{A}, B$  unabhängig  $\iff \bar{A}\bar{B}$  unabhängig.

**Definition 5.16**  $(\Omega, \mathfrak{A}, P)$  sei ein Wahrscheinlichkeitsraum,  $I$  eine beliebige Indermenge und  $A_i \in \mathfrak{A}, i \in I$ .

- a) Die  $A_i, i \in I$ , heißen voneinander paarweise stochastisch unabhängig bezüglich der Wahrscheinlichkeitsverteilung  $P$ , wenn gilt:

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$$

für beliebige Indizes  $i, j \in I$  mit  $i \neq j$ .

- b) Die  $A_i$  heißen voneinander (in ihrer Gesamtheit) stochastisch unabhängig unter der Wahrscheinlichkeitsverteilung  $P$ , wenn gilt:

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

für je endlich viele verschiedene Indizes  $i_1, \dots, i_m \in I$ .

Statt in ihrer Gesamtheit unabhängiger Ereignisse sprechen wir auch einfach von *unabhängigen Ereignissen*.

Aus dieser Definition ergeben sich unmittelbar folgende Eigenschaften.

### Folgerungen 5.17

- 1) Ist  $(A_i, i \in I)$  eine Familie von unabhängigen Ereignissen, so sind auch die Ereignisse jeder Teilfamilie  $(A_i, i \in I'), I' \subseteq I$ , unabhängig.
- 2) Mit  $(A_i, i \in I)$  ist auch  $(C_i, i \in I)$ , wobei  $C_i = A_i$  oder  $C_i = \bar{A}_i, i \in I$ , gilt, eine Familie unabhängiger Ereignisse.

Der Beweis von 2) erfolgt analog zum Beweis der Aussage (5.14) und ist nur schreibtechnisch komplizierter.

Aus der Unabhängigkeit der Ereignisse  $(A_i, i \in I)$  folgt ihre paarweise Unabhängigkeit. Die Umkehrung gilt nicht, siehe Übungen.

**Aussage 5.18** *Ist  $(A_i, i \in I)$  mit  $I \subseteq N_0$  eine Folge voneinander unabhängiger Ereignisse, so gilt*

1)

$$P(\text{Alle } A_i, i \in I \text{ treten ein}) = P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i) \quad (5.10)$$

2)

$$\begin{aligned} P(\text{Mindestens eines der } A_i, i \in I, \text{ tritt ein}) &= P\left(\bigcup_{i \in I} A_i\right) \quad (5.11) \\ &= 1 - \prod_{i \in I} (1 - P(A_i)) \end{aligned}$$

Beweis: Es sei  $I_n = I \cap [0, n], n \geq 1$ .

- 1) Aus der Unabhängigkeit der Ereignisse  $A_i, i \in I$ , folgt  $P(\bigcap_{i \in I_n} A_i) = \prod_{i \in I_n} P(A_i)$  für jedes  $n \geq 1$ . Daraus folgt (...) durch Grenzübergang  $n \rightarrow \infty$  unter Verwendung der Stetigkeit von  $P$  bezüglich monotoner Ereignisfolgen.
- 2) Mit  $A_i, i \in I$ , sind auch die Ereignisse  $\bar{A}_i, i \in I$ , voneinander unabhängig. Folglich ist wegen der eben bewiesenen Eigenschaft a)

$$P\left(\bigcup_{i \in I} A_i\right) = 1 - P\left(\bigcap_{i \in I} \bar{A}_i\right) = 1 - \prod_{i \in I} P(\bar{A}_i), \text{ woraus sich (...) ergibt.} \quad \square$$

**Beispiel 5.19** Die Wahrscheinlichkeit dafür, dass beim rein zufälligen Austeilen eines Skatspieles zwei Buben im Skat liegen, beträgt  $p_1 = 0,0121$ . Niemand rechnet also bei einem einmaligen Spiel damit, dass dieses Ereignis tatsächlich eintritt. Die Wahrscheinlichkeit dafür, dass dieses Ereignis in einer Serie von  $n$  voneinander unabhängigen Spielen mindestens einmal stattfindet, ist dagegen  $p_n = 1 - (1 - p_1)^n$ . Für  $n = 20$  ergibt das  $p_{20} = 0,216$ , und für  $n = 50$  erhalten wir  $p_{50} = 0,456$ . Ab welchem  $n$  kann man darauf wetten, dass in einer Serie von  $n$  Skatspielen mindestens einmal zwei Buben im Skat liegen? (Vgl. Abschnitt 3.4a)

**Aussage 5.20** (2. Lemma von Borel-Cantelli)

Es sei  $(\Omega, \mathfrak{A}, P)$  ein Wahrscheinlichkeitsraum, und  $(A_n, n \geq 1)$  sei eine Folge voneinander unabhängiger Ereignisse aus  $\mathfrak{A}$ . Dann folgt aus  $\sum_{n \geq 1} P(A_n) = \infty$ ,

dass  $P(\limsup_{n \rightarrow \infty} A_n) = 1$  gilt.

Beweis:

$$\begin{aligned} 1 - P(\limsup A_n) &= P\left(\overline{\bigcap_n \bigcup_{m \geq n} A_m}\right) = \\ &= P\left(\bigcup_{n=1}^{\infty} \bigcap_{m \geq n} \bar{A}_m\right) = \lim_{n \rightarrow \infty} P\left(\bigcap_{m \geq n} \bar{A}_m\right) = \\ &= \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} P\left(\bigcap_{m=n}^k \bar{A}_m\right) = \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \prod_{m=n}^k (1 - P(A_m)). \end{aligned}$$

Unter Ausnutzung von  $1 - x \leq e^{-x}$ ,  $x \in R_1$ , folgt

$$1 - P(\limsup A_n) \leq \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \exp[-\sum_{m=n}^k P(A_m)] = 0.$$

□

### 5.2.2 Unabhängigkeit von $\sigma$ -Algebren

Wir erweitern in diesem Punkt den Unabhängigkeitsbegriff auf allgemeine Strukturen, da wir ihn später in dieser Form benötigen werden.

**Definition 5.21** *Es seien  $(\Omega, \mathfrak{A}, P)$  ein Wahrscheinlichkeitsraum,  $I$  eine Indexmenge und  $\gamma_i$  für jedes  $i$  ein Mengensystem mit  $\gamma_i \subseteq \mathfrak{A}$ . Die  $\gamma_i$  heißen voneinander unabhängig, falls für jede endliche Auswahl  $\mathfrak{J} \subseteq I$  und jede Auswahl  $A_j \in \gamma_j$ ,  $j \in \mathfrak{J}$ , gilt:*

$$P(\bigcap_{j \in \mathfrak{J}} A_j) = \prod_{j \in \mathfrak{J}} P(A_j). \tag{5.12}$$

#### Beispiel 5.22

$$\Omega = [0, 1), \mathfrak{A} = \mathfrak{B}_{[0,1)}, P = \lambda_{[0,1)}, \gamma_1 = \{[0, \frac{1}{2}), [\frac{1}{2}, 1)\},$$

$$\gamma_2 = \{[\frac{1}{4}, \frac{3}{4}), ([0, \frac{1}{4}) \cup [\frac{3}{4}, 1))\}$$

$\gamma_1$  und  $\gamma_2$  sind bez.  $P$  unabhängige Ereignissysteme.

Die folgende Aussage ist nützlich zum Nachweis, dass  $\sigma$ -Algebren voneinander unabhängig sind, da die sie erzeugenden Semialgebren im Allgemeinen von wesentlich einfacherer Gestalt sind.

**Aussage 5.23** *Sind die Mengensysteme  $\gamma_i, i \in I$ , aus der vorangegangenen Definition voneinander unabhängige Semiringe, so sind auch die von ihnen erzeugten  $\sigma$ -Algebren  $\mathfrak{A}_i = \sigma(\gamma_i)$ ,  $i \in I$ , voneinander unabhängig.*

Beweis: Es sei  $\mathfrak{J}$  eine endliche Teilmenge von  $I$  und für jedes  $j \in \mathfrak{J}$  sei  $A_j$  ein Element von  $\gamma_j$ . Nach Voraussetzung gilt

$$P\left(\bigcap_{j \in \mathfrak{J}} A_j\right) = \prod_{j \in \mathfrak{J}} P(A_j). \quad (5.13)$$

Fixiert man ein  $j_0 \in \mathfrak{J}$  und alle  $A_j$  mit  $j \in \mathfrak{J} \setminus \{j_0\}$ , so stellen beide Seiten von (8)  $\sigma$ -additive Mengenfunktionen auf der Semialgebra  $\gamma_{j_0}$  dar, die sich eindeutig zu Wahrscheinlichkeitsmaßen auf  $\mathfrak{A}_{j_0}$  fortsetzen lassen. Also gilt (...) bei festen  $A_j$ , ( $j \in \mathfrak{J} \setminus \{j_0\}$ ), für alle  $A_{j_0}$  aus  $\mathfrak{A}_{j_0}$ . Diese Überlegung wird sukzessiv fortgesetzt, und so ergibt sich die Behauptung.  $\square$

Auch die folgende Aussage wird im Weiteren nützlich sein.

**Aussage 5.24** *Es sei  $\mathfrak{A}_i, i \in I$ , eine Familie voneinander unabhängiger Teil- $\sigma$ -Algebren von  $\mathfrak{A}$ . Weiterhin seien  $\mathfrak{J}_1$  und  $\mathfrak{J}_2$  zwei disjunkte Teilmengen von  $I$ . Dann sind auch  $\mathfrak{A}_{\mathfrak{J}_1} := \sigma\left(\bigcup_{i \in \mathfrak{J}_1} \mathfrak{A}_i\right)$  und  $\mathfrak{A}_{\mathfrak{J}_2} := \sigma\left(\bigcup_{i \in \mathfrak{J}_2} \mathfrak{A}_i\right)$  voneinander unabhängig.*

Beweis:  $\gamma_1 = \bigcup_{i \in \mathfrak{J}_1} \mathfrak{A}_i$  und  $\gamma_2 = \bigcup_{i \in \mathfrak{J}_2} \mathfrak{A}_i$  sind nach Voraussetzung zwei unabhängige Ereignissysteme, die überdies Algebren sind. Daraus folgt auf Grund der Aussage 5.23, dass auch  $\mathfrak{A}_{\mathfrak{J}_1}$  und  $\mathfrak{A}_{\mathfrak{J}_2}$  unabhängig sind.  $\square$

### 5.2.3 Unabhängigkeit in mehrstufigen Experimenten

Es sei  $(\Omega, \mathfrak{A}, P)$  ein  $n$ -stufiges zufälliges Experiment mit den Einzelexperimenten  $(\Omega_k, \mathfrak{A}_k, P_k)$ ,  $k = 1, \dots, n$ . Das heißt, es seien

$$\Omega = \prod_1^{n \otimes} \Omega_k, \quad \mathfrak{A} = \prod_1^{n \otimes} \mathfrak{A}_k \quad \text{und} \quad P \text{ eine Wahrscheinlichkeitsverteilung auf } \mathfrak{A}.$$

Mögliche Ausgänge für das  $n$ -stufige Experiment sind somit die Folgen  $\omega = (\omega_1, \omega_2, \dots, \omega_n)$  von Ausgängen  $\omega_k$  der Einzelexperimente. Der Zusammenhang zwischen  $P$  und den  $P_k$  ist gegeben durch

$$P_k(A_k) = P(\Omega_1 \times \dots \times \Omega_{k-1} \times A_k \times \Omega_{k+1} \times \dots \times \Omega_n) \quad (5.14)$$

für alle  $A_k \in \mathfrak{A}_k$  und  $k = 1, \dots, n$ . Wir identifizieren jedes Ereignis  $A_k \in \mathfrak{A}_k$  mit dem ihm zugeordneten Ereignis  $A'_k$  im Gesamtexperiment  $(\Omega, \mathfrak{A}, P)$ :

$$A'_k := \Omega_1 \times \dots \times \Omega_{k-1} \times A_k \times \Omega_{k+1} \times \dots \times \Omega_n, \quad k = 1, \dots, n,$$

und setzen  $\mathfrak{A}'_k := \{A'_k \in \mathfrak{A} : A_k \in \mathfrak{A}_k\}$ ,

Dann bedeutet (...)

$$P_k(A_k) = P(A'_k), \quad A_k \in \mathfrak{A}_k, \quad k = 1, \dots, n.$$

Außerdem gilt mit dieser Terminologie

$$P(A'_1 \cap A'_2 \cap \dots \cap A'_n) = P(A_1 \times A_2 \times \dots \times A_n). \quad (5.15)$$

**Definition 5.25** Wir sagen, die Einzelexperimente  $(\Omega_k, \mathfrak{A}_k, P_k)$  seien im Rahmen des Gesamtexperimentes  $(\Omega, \mathfrak{A}, P)$  voneinander unabhängig, falls die  $\sigma$ -Algebren  $\mathfrak{A}'_k$  bezüglich  $P$  voneinander unabhängig sind, d.h., falls gilt:

$$P(A'_1 \dots \cap A'_n) = \prod_1^n P(A'_k), \quad A'_k \in \mathfrak{A}'_k, \quad k = 1, \dots, n. \quad (5.16)$$

**Aussage 5.26** Die Einzelexperimente  $(\Omega_k, \mathfrak{A}_k, P_k)$  sind im Rahmen des Gesamtexperimentes  $(\Omega, \mathfrak{A}, P)$  genau dann unabhängig, wenn  $P$  das Produktmaß der  $P_k$ ,  $k = 1, \dots, n$ , ist, d.h., falls für jede Auswahl  $A_k \in \mathfrak{A}_k$ ,  $k = 1, 2, \dots, n$  gilt:

$$P(A_1 \times \dots \times A_n) = \prod_1^n P_k(A_k). \quad (5.17)$$

Beweis: (...) ist mit (...) identisch, siehe (...) und (...).

Beide Seiten von (...) bilden  $\sigma$ -additive Mengenfunktionen auf der Semialgebra  $\gamma = \{A_1 \times A_2 \times \dots \times A_n \mid A_k \in \mathfrak{A}_k, k = 1, \dots, n\}$  und sind dort gleich. Wegen der eindeutigen Fortsetzbarkeit stimmen sie auch auf  $\mathfrak{A} = \sigma(\gamma)$  überein.

□

**Folgerungen 5.27** Sind alle  $\Omega_k$ ,  $k = 1, \dots, n$ , höchstens abzählbar,  $\mathfrak{A}_k = \mathfrak{P}(\Omega_k)$ , und ist  $P$  gegeben durch die Einzelwahrscheinlichkeiten

$$P(\{\omega\}) = p^{(1)}(\omega_1)p_{\omega_1}^{(2)}(\omega_2) \cdot \dots \cdot p_{\omega_1\omega_2\dots\omega_{n-1}}^{(n)}(\omega_n) \text{ mit } \omega = (\omega_1, \dots, \omega_n),$$

so sind die Einzelexperimente genau dann voneinander unabhängig (bez.  $P$ ), wenn die Übergangswahrscheinlichkeiten

$$p_{\omega_1\dots\omega_{k-1}}^{(k)}(\omega_k) =: p^{(k)}(\omega_k), \quad k = 2, \dots, n, \quad (5.18)$$

nicht von  $\omega_1, \dots, \omega_{k-1}$  abhängen.

In diesem Fall gilt

$$P_k(\{\omega_k\}) = p^{(k)}(\omega_k) \text{ und}$$

$$P(\{\omega\}) = \prod_{k=1}^n p^{(k)}(\omega_k), \quad \omega = (\omega_1, \dots, \omega_n).$$

Dieser Sachverhalt entspricht dem anschaulichen Unabhängigkeitsbegriff bei der Konstruktion des mehrstufigen Experimentes aus den Größen  $p_{\omega_1\dots\omega_{k-1}}^{(k)}$ .

## 5.2.4 Unabhängigkeit von Zufallsgrößen

Es seien  $X$  und  $Y$  zwei Zufallsgrößen auf einem Wahrscheinlichkeitsraum mit Werten in  $(E, \mathfrak{E})$  bzw.  $(F, \mathfrak{F})$ .

**Definition 5.28** Die Zufallsgrößen  $X$  und  $Y$  heißen voneinander unabhängig, falls für jede Wahl von Mengen  $C \in \mathfrak{E}$  und  $D \in \mathfrak{F}$  die Ereignisse  $\{X \in C\}$  und  $\{Y \in D\}$  voneinander unabhängig sind, d.h., falls gilt:

$$P(X \in C, Y \in D) = P(X \in C)P(Y \in D), \quad C \in \mathfrak{E}, D \in \mathfrak{F}. \quad (5.19)$$

Aus dieser Definition ist wegen  $\{X \in C, Y \in D\} = X^{-1}(C) \cap Y^{-1}(D)$  offenbar, dass die Unabhängigkeit von  $X$  und  $Y$  äquivalent damit ist, dass die  $\sigma$ -Algebren  $\mathfrak{A}^X = X^{-1}(\mathfrak{E}) = \{\{X \in C\} | C \in \mathfrak{E}\}$  und  $\mathfrak{A}^Y = Y^{-1}(\mathfrak{F}) = \{\{Y \in D\} | D \in \mathfrak{F}\}$  voneinander unabhängig sind.

Die Eigenschaft (...) lässt sich wegen  $\{X \in C, Y \in D\} = \{(X, Y) \in C \times D\}$  auch schreiben als

$$P^{(X,Y)}(C \times D) = P^X(C)P^Y(D), \quad C \in \mathfrak{E}, D \in \mathfrak{F}.$$

Die Unabhängigkeit der beiden Zufallsgrößen  $X$  und  $Y$  bedeutet also, dass ihre gemeinsame Wahrscheinlichkeitsverteilung  $P^{(X,Y)}$  gleich dem Produktmaß ihrer beiden Randverteilungen  $P^X$  und  $P^Y$  ist. Die Unabhängigkeit der zwei Zufallsgrößen  $X$  und  $Y$  ist somit eine Eigenschaft ihrer gemeinsamen Verteilungen  $P^{(X,Y)}$ . Im Fall der Unabhängigkeit von  $X$  und  $Y$  ist die gemeinsame Verteilung  $P^{(X,Y)}$  durch ihre Randverteilungen  $P^X$  und  $P^Y$  also eindeutig bestimmt.

Die folgende Aussage erlaubt es, die Unabhängigkeit zweier reellwertiger Zufallsgrößen anhand ihrer gemeinsamen Verteilungsfunktion zu prüfen.

**Aussage 5.29** *Zwei reellwertige Zufallsgrößen  $X$  und  $Y$  über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  sind genau dann unabhängig, wenn für ihre gemeinsame Verteilungsfunktion  $F_{(X,Y)}$  und ihre Randverteilungsfunktionen  $F_X$  und  $F_Y$  gilt:*

$$F_{(X,Y)}(x, y) = F_X(x)F_Y(y), \quad x, y \in R_1. \quad (5.20)$$

Beweis: Sind  $X$  und  $Y$  unabhängig, so gilt

$$F_{(X,Y)}(x, y) = P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) = F_X(x)F_Y(y).$$

Gilt dagegen (...), so ist mit  $F = F_{X,Y}$

$$\begin{aligned} P(X \in (a_1, b_1], Y \in (a_2, b_2]) &= \\ F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) &= \\ F_X(b_1)F_Y(b_2) - F_X(a_1)F_Y(b_2) - F_X(b_1)F_Y(a_2) + F_X(a_1)F_Y(a_2) &= \\ (F_X(b_1) - F_X(a_1))(F_Y(b_2) - F_Y(a_2)) &= \\ P(X \in (a_1, b_1]) \cdot P(Y \in (a_2, b_2]). & \end{aligned} \quad (5.21)$$

Es sei  $\gamma$  der Semiring aller halboffenen Intervalle  $(a, b] : \gamma = \{(a, b] \mid -\infty < a < b < \infty\}$ . Bekanntlich gilt  $\sigma(\gamma) = \mathfrak{B}_1$ . Die Gleichung (...) besagt, dass die

Semiringe  $X^{-1}(\gamma)$  und  $Y^{-1}(\gamma)$  unabhängig sind. Somit sind auch  $\sigma(X^{-1}(\gamma)) = X^{-1}(\sigma(\gamma)) = X^{-1}(\mathfrak{B}_1)$  und  $\sigma(Y^{-1}(\gamma)) = Y^{-1}(\sigma(\gamma)) = X^{-1}(\mathfrak{B}_1)$  unabhängig, was aber mit der Unabhängigkeit von  $X$  und  $Y$  äquivalent ist.  $\square$

Die Definitionen und Aussagen dieses Abschnittes lassen sich analog zu Abschnitt 5.2. auf beliebige Familien  $(X_i, i \in I)$  von Zufallsgrößen über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  verallgemeinern. Wir beschränken uns auf folgende

**Definition 5.30** *Es seien  $I$  irgendeine Indexmenge und  $X_i, i \in I$ , Zufallsgrößen über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  mit Werten in  $(E_i, \mathfrak{E}_i), i \in I$ . Die Zufallsgrößen  $X_i, i \in I$ , heißen voneinander unabhängig, falls für jede endliche Teilmenge  $\mathfrak{J}$  von  $I$  und jede Auswahl  $B_j \in \mathfrak{E}_j, j \in \mathfrak{J}$ , gilt:*

$$P\left(\bigcap_{j \in \mathfrak{J}} \{X_j \in B_j\}\right) = \prod_{j \in \mathfrak{J}} P(X_j \in B_j).$$

Wir betrachten noch die zwei Sonderfälle, dass  $(X, Y)$  diskret verteilt ist bzw. eine gemeinsame Dichte besitzt.

## Diskret verteilte zufällige Vektoren

Die gemeinsame Verteilung der Zufallsgrößen  $X, Y$  sei diskret und gegeben durch  $((x_i, y_j), p_{ij}), i \in I, j \in \mathfrak{J}$ .

a)  $X$  und  $Y$  sind genau dann unabhängig, wenn

$$p_{ij} = p_i \cdot p_{.j}, \quad i \in I, j \in \mathfrak{J} \quad (5.22)$$

gilt, wobei nach Definition  $p_i = \sum_j p_{ij}, i \in I$ , und  $p_{.j} = \sum_i p_{ij}, j \in \mathfrak{J}$ , die Einzelwahrscheinlichkeiten von  $X$  bzw.  $Y$  sind.

Beweis: (5.25) folgt aus (5.22) mit  $C = \{x_i\}, D = \{y_j\}$ . Umgekehrt ergibt sich (5.22) aus (5.25) mittels

$$P(X \in C, Y \in D) = \sum_{\substack{i: x_i \in C, \\ j: y_j \in D}} p_{ij} = \sum_{\substack{i: x_i \in C, \\ j: y_j \in D}} p_i \cdot p_j = \sum_{i: x_i \in C} p_i \cdot \sum_{j: y_j \in D} p_j = P(X \in C)P(Y \in D)$$

□

- b) Sind  $X$  und  $Y$  reellwertig und unabhängig, existieren die Erwartungswerte  $EX$ ,  $EY$  und sind sie beide endlich, so existiert auch der Erwartungswert  $E(XY)$  und ist endlich.

Überdies gilt in diesem Fall:

$$E(XY) = EX EY. \tag{5.23}$$

Beweis: Nach Voraussetzung ist  $\sum_i |x_i|p_i < \infty$  und  $\sum_j |y_j|p_j < \infty$ . Folglich gilt auch

$$\sum_{i,j} |x_i y_j| p_{ij} = \sum_{i,j} |x_i y_j| p_i \cdot p_j = \sum_i |x_i| p_i \cdot \sum_j |y_j| p_j < \infty.$$

Außerdem gilt für diese absolut konvergente Reihe

$$\sum_{i,j} x_i y_j p_{ij} = \sum_i x_i \sum_j y_j p_{ij} = \sum_i x_i p_i \cdot \sum_j y_j p_j.$$

□

Die Formel (5.26) impliziert, dass  $X$  und  $Y$  unter den genannten Voraussetzungen unkorreliert sind:

$$\text{Kov}(X, Y) = E(XY) - EXEY = 0.$$

Sind überdies die Streuungen  $D^2X$  und  $D^2Y$  endlich, so folgt aus der Unabhängigkeit von  $X$  und  $Y$  die Gleichung

$$D^2(X + Y) = D^2X + D^2Y.$$

Beweis:

$$\begin{aligned} D^2(X + Y) &= E(X + Y - EX - EY)^2 = D^2X + D^2Y + 2 \operatorname{Cov}(X, Y) \\ &= D^2X + D^2Y. \end{aligned}$$

□

- c) Die Einzelwahrscheinlichkeiten  $p_{ij}$  werden wie folgt in einer Tabelle aufgeschrieben.

$i \setminus j$	1	2	3	...	$j$	...	
1	$p_{11}$	$p_{12}$	...	...	$p_{1j}$	...	$p_{1\cdot}$
2	$p_{21}$	$p_{22}$					$p_{2\cdot}$
3	.						
.	.						
.	.						
.	.						
$i$	$p_{i1}$	...			...	$p_{ij}$	$p_{i\cdot}$
.	.						
	$p_{\cdot 1}$	$p_{\cdot 2}$			$p_{\cdot j}$		1

Teilt man die  $i$ -te Zeile durch den Wert  $p_{i\cdot}$  sofern  $p_{i\cdot} > 0$  gilt, so erhält man eine Wahrscheinlichkeitsverteilung

$$\left( \frac{p_{ij}}{p_{i\cdot}}, j \geq 1 \right),$$

für die gilt

$$\frac{p_{ij}}{p_{i\cdot}} = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} = P(Y = y_j | X = x_i).$$

Analog liefert die  $j$ -te Spalte, geteilt durch  $p_{\cdot j}$

$$\left( \frac{p_{ij}}{p_{.j}}, i \geq 1 \right)$$

$$\text{mit } \frac{p_{ij}}{p_{.j}} = P(X = x_i | Y = y_j).$$

Bei fest gewählten  $y_j$  bilden die  $P(X = x_i | Y = y_j), i \in \mathcal{J}$ , die Einzelwahrscheinlichkeiten einer diskreten Verteilung, die man als *bedingte Verteilung von  $X$  unter der Bedingung  $\{Y = y_j\}$*  bezeichnet. Analog definierten die  $P(Y = y_j | X = x_i)$  die bedingte Verteilung von  $Y$  unter der Bedingung  $\{X = x_i\}$ .

## Zufällige Vektoren mit gemeinsamer Dichte

- a) Besitzt  $(X, Y)$  eine Dichte  $f_{(X,Y)}$  und sind  $f_X$  und  $f_Y$  die entsprechenden Randverteilungsdichten, so sind  $X$  und  $Y$  genau dann unabhängig, wenn

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad \lambda \otimes \lambda - \text{f.ü.} \quad (5.24)$$

gilt.

Beweis: Aus (...) folgt (...) durch Differentiation nach  $x$  und nach  $y$ . Aus (...) ergibt sich (...) nach Definition der Dichten.  $\square$

Der Beweis folgt unmittelbar aus (...), wir kommen später im allgemeineren Rahmen darauf zurück.

- b) Zweidimensionale Normalverteilung: Es sei  $(X, Y)$  ein  $N(\mu, \Sigma)$ -verteilter zufälliger Vektor mit den Parametern  $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$ . Genau dann sind  $X$  und  $Y$  unabhängig, wenn  $\rho = 0$  gilt.



# Kapitel 6

## Bernoullischemata und Irrfahrten

Dieses Kapitel beginnt mit Folgen unabhängiger identisch verteilter Zufallsgrößen, die alle nur zwei Werte (zum Beispiel Null oder Eins) annehmen können, sogenannte *Bernoullischemata*. Solche Schemata bilden stochastische Modelle für zahlreiche reale Situationen.

Bernoullischemata sind eng verwandt mit Irrfahrten, die wir als mathematisches Modell des mehrfachen Werfens einer Münze in den Abschnitten 2.5 und 3.3 kennen gelernt haben.

### 6.1 Bernoullischemata

Es seien  $(\Omega, \mathfrak{A}, P)$  ein Wahrscheinlichkeitsraum,  $n \geq 1$  und  $p \in (0, 1)$ .

**Definition 6.1** *Es sei  $n \geq 2$ . Jede Folge  $X = (X_1, X_2, \dots, X_n)$  von Zufallsgrößen über  $(\Omega, \mathfrak{A}, P)$  mit*

1.  $X_1, X_2, \dots, X_n$  sind voneinander unabhängig,
2.  $P(X_k = 1) = p, P(X_k = 0) = 1 - p =: q, \quad k = 1, \dots, n,$

*heißt ein Bernoullischema mit den Parametern  $n$  und  $p$ . Wir bezeichnen es mit  $BS_n(p)$ .*

Ist  $X = (X_k, k \geq 1)$  eine unendliche Folge von Zufallsgrößen mit den Eigenschaften 1. und 2. für alle  $n \geq 2$ , so heißt  $(X_k, k \geq 1)$  ein unendliches Bernoullischema mit dem Parameter  $p$ . Wir verwenden dafür die Kurzschreibweise  $BS_\infty(p)$ .

Eigenschaft 2. aus Definition 6.1 kann man auch in der Form

$$P(X_k = i) = p^i(1-p)^{1-i}, \quad i \in \{0, 1\}, k = 1, \dots, n \quad (6.1)$$

schreiben.

Für jedes Bernoullischema  $BS_n(p)$  ist  $X = (X_1, \dots, X_n)$  ein zufälliger Vektor mit den möglichen Werten

$$x = (i_1, i_2, \dots, i_n) \in \{0, 1\}^n =: E.$$

Der zufällige Vektor  $X$  bildet  $\Omega$  in  $E$  ab. Er ist folglich diskret verteilt, und es gilt nach Definition 6.1 und mit (6.1) für jedes  $x = (i_1, \dots, i_n) \in E$ :

$$P(X = x) = P(X_1 = i_1, \dots, X_n = i_n) = \prod_{k=1}^n P(X_k = i_k) = p^{\sum_{k=1}^n i_k} (1-p)^{n - \sum_{k=1}^n i_k}. \quad (6.2)$$

**Bemerkung 6.2** Das Bernoullischema  $BS_n(p)$  entspricht einem  $n$ -stufigen zufälligem Experiment, dessen Einzelexperimente voneinander unabhängig sind und jeweils nur zwei mögliche Versuchsausgänge haben (wir bezeichnen sie mit 0 bzw. 1), die in jedem Telexperiment mit der Wahrscheinlichkeit  $q$  bzw.  $p$  auftreten. Die Zufallsgröße  $X_k$  gibt in diesem Zusammenhang das Ergebnis des  $k$ -ten Telexperimentes an.

Dabei wird das Ereignis  $\{X_k = 1\}$  häufig als *Erfolg im  $k$ -ten Experiment* bzw. das Ereignis  $\{X_k = 0\}$  als *Misserfolg im  $k$ -ten Experiment* gedeutet.

Mit jedem Bernoullischema sind gewisse konkrete Zufallsgrößen und ihre Wahrscheinlichkeitsverteilungen verbunden, von denen wir im Folgenden einige studieren werden.

## Anzahl $S_n$ der Erfolge bei $n$ Versuchen

Es sei  $X = (X_1, \dots, X_n)$  ein  $BS_n(p)$ . Die Zufallsgröße  $S_k$ , definiert durch  $S_k(\omega) = \sum_{l=1}^k X_l(\omega)$ , ist gleich der *Anzahl der Erfolge* in den ersten  $k$  Teilerperimenten des Bernoullischemas.

**Aussage 6.3** Die Zufallsgröße  $S_n$  ist binomialverteilt mit den Parametern  $n$  und  $p$ , d.h. es gilt

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n. \quad (6.3)$$

Beweis:  $P(S_n = k) = P(X_1 + \dots + X_n = k) =$

$$\begin{aligned} P^X(\{x \in E : \sum_{j=1}^n i_j = k\}) &= \sum_{\substack{x=(i_1, \dots, i_n): \\ \sum i_j = k}} \prod_{j=1}^n p^{i_j} (1-p)^{1-i_j} = \sum_{\substack{x=(i_1, \dots, i_n): \\ \sum i_j = k}} p^k (1-p)^{n-k} \\ &= \binom{n}{k} p^k (1-p)^{n-k}, \end{aligned}$$

da alle  $x \in E$  mit  $\sum_{j=1}^n i_j = k$  dieselbe Wahrscheinlichkeit  $p^k (1-p)^{n-k}$  besitzen (siehe (6.2)) und es  $\binom{n}{k}$  solcher  $x \in E$  gibt.  $\square$

Insbesondere ist  $ES_n = np$  und  $VarS_n = np(1-p)$ .

Die relative Häufigkeit der Erfolge in  $n$  Versuchen ist  $\frac{S_n}{n}$ . Somit gilt

$$E \frac{S_n}{n} = p \quad \text{und} \quad Var \frac{S_n}{n} = \frac{p(1-p)}{n}.$$

Für große  $n$  wird  $\frac{S_n}{n}$  mit hoher Wahrscheinlichkeit einen Wert in der Nähe von  $p$  annehmen. Das ergibt sich aus der Tschebyschevschen Ungleichung. Für jedes  $\varepsilon > 0$  gilt nämlich

$$P\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leq \frac{\text{Var}\left(\frac{S_n}{n}\right)}{\varepsilon^2} = \frac{p(1-p)}{n \cdot \varepsilon^2}. \quad (6.4)$$

Fixiert man  $\varepsilon > 0$ , so kann man durch Wahl einer genügend großen Versuchszahl  $n$  die rechte Seite von (6.4) beliebig klein machen. Diese Tatsache macht man sich bei unbekanntem  $p$  zunutze. Man verwendet das arithmetische Mittel  $\frac{S_n}{n}$  der beobachteten Werte als statistische Schätzung für  $p$ .

## Zeit $T_m$ des $m$ -ten Erfolges

Für die folgenden Betrachtungen nehmen wir ein unendliches Bernoullischema  $BS_\infty(p)$  als gegeben an.

Die Zufallsgrößen

$$T_1(\omega) := \min\{k > 0 \mid X_k(\omega) = 1\},$$

$$T_m(\omega) := \min\{k > T_{m-1}(\omega) \mid X_k(\omega) = 1\}, \quad m \geq 2$$

geben die Zeitpunkte an, zu denen der erste bzw. der  $m$ -te Erfolg im  $BS_\infty(p)$  eintritt. Dabei setzt man  $\min \emptyset := \infty$  und  $T_0(\omega) := 0$ .

Mit diesen Bezeichnungen ist  $T_{m+1} - T_m$  die Anzahl der Versuche nach dem  $m$ -ten bis zum  $(m+1)$ -ten Erfolg,  $T_m - m$  die Anzahl der Misserfolge bis zum  $m$ -ten Erfolg und  $T_{m+1} - T_m - 1$  ist die Anzahl der Misserfolge zwischen dem  $m$ -ten und dem  $(m+1)$ -ten Erfolg,  $m \geq 0$ .

**Aussage 6.4** *Es gilt*

$$\left. \begin{aligned} \{T_1 = 1\} &= \{X_1 = 1\} \\ \{T_1 = k\} &= \{X_1 = 0, X_2 = 0, \dots, X_{k-1} = 0, X_k = 1\}, \quad k > 1 \\ \{T_1 \leq k\} &= \bigcup_{l=1}^k \{X_l = 1\}, \quad k \geq 1 \end{aligned} \right\} \quad (6.5)$$

Der Beweis ergibt sich unmittelbar aus der Definition von  $T_1$ .

Analog erhalten wir

$$\{T_m = k\} = \left\{ \sum_{l=1}^k X_l(\omega) = m, X_k = 1 \right\}. \quad (6.6)$$

Wir erinnern daran, dass die kleinste  $\sigma$ -Algebra  $\mathfrak{A}_k := \sigma(X_1, \dots, X_k)$  von Teilmengen von  $\Omega$ , bezüglich der alle  $X_1, X_2, \dots, X_k$  meßbar sind, aus allen Teilmengen  $A$  von  $\Omega$  besteht, die die Form

$A = \{\omega \in \Omega | (X_1(\omega), \dots, X_k(\omega)) \in B\}$  für ein  $B \subseteq E = \{0, 1\}^k$  besitzen.

(Wir hatten  $\mathcal{E} = \mathfrak{P}(E)$  gesetzt.)

Man sagt,  $\mathfrak{A}_k$  enthält alle diejenigen Ereignisse, die mit dem Verlauf der Folge  $(X_1, X_2, \dots)$  bis zur Zeit  $k$  zusammenhängen.

**Folgerung 6.5** Für alle  $m \geq 1$  und alle  $k \geq m$  gilt

$$\{T_m = k\} \in \sigma(X_1, X_2, \dots, X_k) = \mathfrak{A}_k \quad (6.7)$$

Beweis: Die Eigenschaft (6.7) ist eine Konsequenz aus (6.6).  $\square$

**Aussage 6.6** *Es gilt:*

- a)  $T_1 - 1$  ist geometrisch verteilt mit dem Parameter  $p$ .
- b)  $(T_{k+1} - T_{k-1}), k \geq 0$ , sind voneinander unabhängig und identisch verteilt wie  $T_1 - 1$ .
- c)  $T_m - m$  besitzt eine negative Binomialverteilung mit den Parametern  $m$  und  $p$ . (Man beachte, dass insbesondere  $P(T_m < \infty) = \sum_{k=1}^{\infty} P(T_m = k) = 1$  gilt.)

- d) Die Folge  $(X_{T_m+k}, k \geq 1)$  bildet ein unendliches Bernoullischema mit dem Parameter  $p$  und ist unabhängig von  $\mathfrak{A}_{T_m}$ .

Hier bezeichnet  $\mathfrak{A}_{T_m}$  die  $\sigma$ -Algebra aller Ereignisse, die mit dem Verlauf der Folge  $(X_1, X_2, \dots)$  bis zur Zeit  $T_m$  zusammenhängen. Genauer, man definiert

$$\mathfrak{A}_{T_m} := \{A \in \sigma(X_1, X_2, \dots) \mid A \cap \{T_m = k\} \in \mathfrak{A}_k, k \geq 1\}.$$

Beweis:

- a) Es gilt wegen (6.3) und (6.2) für  $k \geq 0$

$$P(T_1 - 1 = k) = P(T_1 = k + 1) = (1 - p)^k p.$$

- b) Es sei  $m$  irgendeine natürliche Zahl mit  $m \geq 2$ . Das Ereignis

$$\bigcap_{k=0}^m \{T_{k+1} - T_k - 1 = t_k\} \text{ mit } t_k \in N_0 = \{0, 1, 2, \dots\}, k = 0, \dots, m,$$

tritt genau dann ein, wenn zwischen  $T_k$  und  $T_{k+1}$  genau  $t_k$  Misserfolge stattfinden,  $k = 0, \dots, m$ . Es ist folglich gleich dem Ereignis (mit der Bezeichnung  $s_\ell = t_0 + t_1 + \dots + t_\ell + \ell, 1 \leq \ell \leq m$ )

$$\bigcap_{\ell=1}^m \{X_{s_\ell} = 1\} \cap \bigcap_{\substack{j=1 \\ j \neq s_\ell, \\ \ell=1, \dots, m}}^{s_m} \{X_j = 0\}.$$

Seine Wahrscheinlichkeit ist auf Grund von (6.2) gleich

$$(1 - p)^{t_0} p (1 - p)^{t_1} p \cdot \dots \cdot (1 - p)^{t_m} p = (1 - p)^{t_0 + t_1 + \dots + t_m} p^m.$$

Daraus ergibt sich für jedes  $k$  mit  $0 \leq k \leq m$

$$P(T_{k+1} - T_k - 1 = t_k) = (1 - p)^{t_k} p$$

und somit

$$P\left(\bigcap_{k=0}^m \{T_{k+1} - T_k - 1 = t_k\}\right) = \prod_{k=0}^m P(T_{k+1} - T_k - 1 = t_k)$$

für alle  $t_k \in N_0, k = 0, \dots, m$ .

Das heißt, die  $(T_k - T_{k-1} - 1), k \geq 1$  sind voneinander unabhängig und alle geometrisch mit dem Parameter  $p$  verteilt.

- c) Das Ereignis  $\{T_m - m = k\}$  tritt genau dann ein, wenn in  $\{X_1 = i_1, \dots, X_{m+k} = i_{m+k}\}$  unter den  $i_j$  genau  $m$  mal eine Eins, sonst Nullen auftreten und  $i_{m+k} = 1$  gilt. Dafür gibt es  $\binom{m+k-1}{m-1}$  gleichwahrscheinliche Fälle mit jeweils der Wahrscheinlichkeit  $p^m(1-p)^k$ . Also ist

$$P(T_m - m = k) = \binom{m+k-1}{k} (1-p)^k p^m = \binom{-m}{k} (p-1)^k p^m.$$

Damit ist Teil c) bewiesen.

- d) Wir beweisen d) nur für  $m = 1$ . Der allgemeine Fall wird analog behandelt. Es sei  $A \in \mathfrak{A}_{T_1}$ . Dann gilt

$$A \cap \{T_1 = k\} \in \mathfrak{A}_k = \sigma(X_1, X_2, \dots, X_k), \quad k \geq 1,$$

und folglich gibt es für jedes  $k \geq 1$  ein  $B_k \subseteq \{0, 1\}^k$  mit

$$A \cap \{T_1 = k\} = \{(X_1, \dots, X_k) \in B_k\}.$$

Daraus ergibt sich für alle  $r \geq 1, i_\ell \in \{0, 1\}, \ell = 1, \dots, r$

$$P(A \cap \{T_1 = k\}, X_{T_1+1} = i_1, \dots, X_{T_1+r} = i_r) =$$

$$P((X_1, \dots, X_k) \in B_k, X_{k+1} = i_1, \dots, X_{k+r} = i_r). \quad (6.8)$$

Wegen der Unabhängigkeit der beiden Folgen  $(X_1, \dots, X_k)$  und  $(X_{k+1}, \dots, X_{k+r})$  ist dieser Wert gleich

$$P((X_1, \dots, X_k) \in B_k) \cdot P(X_{k+1} = i_1, \dots, X_{k+r} = i_r) =$$

$$P(A \cap \{T_1 = k\}) \cdot P(X_1 = i_1, \dots, X_r = i_r).$$

Daraus ergibt sich durch Summation über  $k \geq 1$  wegen  $P(T_1 < \infty) = 1$  die Gleichung

$$P(A \cap \{X_{T_1+1} = i_1, \dots, X_{T_1+r} = i_r\}) =$$

$$P(A)P(X_1 = i_1, \dots, X_r = i_r) = P(A)p^{\sum_{k=1}^r i_k} (1-p)^{r-\sum_{k=1}^r i_k}.$$

Damit ist die Aussage bewiesen. □

Ist in einem Bernoullischema  $BS_n(p)$  die Anzahl  $n$  der Versuche groß und die Erfolgswahrscheinlichkeit  $p$  in jedem einzelnen Versuch klein, so wird die Binomialverteilung durch die Poissonverteilung approximiert. Das heißt

$$b(n, p; k) \approx \frac{\lambda^k}{k!} e^{-\lambda} \quad (6.9)$$

wobei  $\lambda = np$  gesetzt wird. Das ist eine Konsequenz aus dem Poissonschen Grenzwertsatz.

$k$	$n = 10, p = 0,2$	$n = 50, p = 0,04$	$n = 100, p = 0,02$	$\lambda = 2$
0	0,1074	0,1299	0,1326	0,1353
1	0,2684	0,2706	0,2707	0,2707
2	0,3020	0,2762	0,2734	0,2707
3	0,2013	0,1842	0,1823	0,1804
4	0,0881	0,0902	0,0902	0,0902
5	0,0264	0,0346	0,0353	0,0361
6	0,0055	0,0108	0,0114	0,0120
7	0,0008	0,0028	0,0031	0,0034
8	0,0001	0,0006	0,0007	0,0009
9	0,0000	0,0001	0,0002	0,0002
10	0,0000	0,0000	0,0000	0,0000

Tabelle 1. Poissonapproximation der Binomialverteilung

## 6.2 Irrfahrten

Wir haben bereits die symmetrische Irrfahrt  $(S_k, k \leq n)$  kennen gelernt, die sich aus einem Laplace-Experiment ( $n$ -maliges Werfen einer regulären Münze) ergibt:

$$(\Omega, \mathfrak{A}, P) \text{ mit } \Omega = \{-1, +1\}^n, \mathfrak{A} = \mathfrak{P}(\Omega),$$

$$X_l(\omega) = x_l, \omega = (x_1, x_2, \dots, x_n) \in \Omega, l = 1, \dots, n,$$

$$S_k(\omega) = \sum_{l=1}^k X_l(\omega) \text{ und } P(\{\omega\}) = 2^{-n}, \omega \in \Omega.$$

Offenbar gilt

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = 2^{-n} \quad (x_1, x_2, \dots, x_n) \in \{-1, 1\}^n$$

und somit für jedes  $k$  mit  $1 \leq k \leq n$  und jedes  $x_k \in \{-1, +1\}$

$$P(X_k = x_k) = \sum_{x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n \in \{-1, 1\}} P(X_1 = x_1, \dots, X_n = x_n) = \frac{1}{2}. \quad (6.10)$$

Daraus folgt für alle  $(x_1, x_2, \dots, x_n) \in \{-1, 1\}^n$  die Gleichung

$$[P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)] = \prod_{k=1}^n P(X_k = x_k). \quad (6.11)$$

Das heißt aber gerade, dass die Zufallsgrößen  $X_1, X_2, \dots, X_n$  unter den angenommenen Voraussetzungen voneinander unabhängig sind und alle die gleiche Verteilung besitzen ( m.a.W., identisch verteilt sind).

Wir behalten die Unabhängigkeit und identische Verteilung der Zufallsgröße  $X_1, X_2, \dots$  bei, lassen aber als Einzelwahrscheinlichkeiten  $P(X_k = 1)$  und  $P(X_k = -1)$  irgend welche Werte  $p$  bzw.  $q = 1 - p$  mit  $p \in (0, 1)$  zu. Die Münze wird also nicht mehr als regulär vorausgesetzt. Demnach gilt

$$P(X_k = 1) = p, P(X_k = -1) = 1 - p, k = 1, 2, \dots, n,$$

in anderer Schreibweise

$$P(X_k = i) = p^{\frac{1+i}{2}} (1-p)^{\frac{1-i}{2}}, i \in \{-1, +1\}, 1 \leq k \leq n. \quad (6.12)$$

Die Ausgänge  $\omega = (x_1, x_2, \dots, x_n) \in \Omega$  der Wurfserie haben jedoch, falls  $p \neq q$  gilt, nicht mehr alle die gleiche Wahrscheinlichkeit, es gilt vielmehr für  $\omega = (x_1, x_2, \dots, x_n) \in \{-1, +1\}^n$

$$P(\{\omega\}) = \prod_{k=1}^n P(X_k = x_k) = p^{\frac{n+s_n}{2}} (1-p)^{\frac{n-s_n}{2}} \quad (6.13)$$

mit  $s_n = x_1 + \dots + x_n$ .

Durch

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

ist auf  $\mathfrak{A} = \mathfrak{P}(\Omega)$  eine Verteilung  $P$  definiert.

**Aussage 6.7** *Bezüglich  $P$  sind wegen (6.12) und (6.13) die Zufallsgrößen  $X_1, X_2, \dots, X_n$  unabhängig und identisch (gemäß (6.12)) verteilt.*

Das ergibt sich aus  $P(\{\omega\}) = P(X_1 = x_1, \dots, X_n = x_n)$  für  $\omega = (x_1, x_2, \dots, x_n)$  und (6.13).

Die Folge  $(Y_k, k \leq n)$  mit  $Y_k = \frac{X_k+1}{2}, 1 \leq k \leq n$ , bildet ein Bernoullischema  $BS_n(p)$ .

Im Folgenden gehen wir davon aus, dass  $(X_k, k \geq 1)$  eine unendliche Folge voneinander unabhängiger identisch verteilter Zufallsgrößen über einem geeigneten Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  mit der Verteilung (6.12) ist.

**Definition 6.8** Die Folge  $(S_n, n \geq 0)$  mit

$$S_0 := s_0, \quad S_n = \sum_{k=1}^n X_k, \quad n \geq 1, s_0 \in Z,$$

heißt eine Irrfahrt auf  $Z$  mit dem Parameter  $p$  und mit Start in  $s_0$ . Ist  $s_0 = 0$ , so nennt man  $(S_n, n \geq 0)$  einfach eine Irrfahrt mit dem Parameter  $p$ . Im Fall  $p = q = \frac{1}{2}$  spricht man von einer symmetrischen Irrfahrt auf  $Z$ . Dabei bezeichnet  $Z$  die Menge aller ganzen Zahlen.

Im Folgenden verwenden wir die Bezeichnung

$$\mathfrak{A}_n := \sigma(X_1, X_2, \dots, X_n), \quad n \geq 1,$$

d.h.,  $\mathfrak{A}_n$  ist die kleinste  $\sigma$ -Algebra von Teilmengen von  $\Omega$ , bezüglich der alle  $X_1, X_2, \dots, X_n$  messbar sind. Es gilt  $\mathfrak{A}_n \subseteq \mathfrak{A}$ , da alle  $X_k$  bezüglich  $\mathfrak{A}$  messbar sind, und  $\mathfrak{A}_n$  besteht hier aus allen Teilmengen  $A$  von  $\Omega$  der Form

$$A = \{\omega \in \Omega \mid (X_1(\omega), \dots, X_n(\omega)) \in B\}, \text{ wobei } B \text{ irgend eine Teilmenge von } \{-1, +1\}^n \text{ ist.}$$

Wegen der bijektiven Beziehung zwischen  $(X_1, X_2, \dots, X_n)$  und  $(S_1, S_2, \dots, S_n)$  haben wir auch

$$\mathfrak{A}_n = \sigma(S_1, S_2, \dots, S_n), \quad n \geq 1.$$

Eine fast offensichtliche Eigenschaft jeder Irrfahrt  $(S_n, n \geq 0)$  mit Parameter  $p$  ist die folgende:

**Aussage 6.9** Wählt man eine feste Zeit  $n_0 \geq 1$  und bildet man

$$\tilde{S}_n := S_{n+n_0} - S_{n_0} \quad (n \geq 0),$$

so ist  $(\tilde{S}_n, n \geq 0)$  wieder eine Irrfahrt mit demselben Parameter  $p$ . Außerdem ist  $(\tilde{S}_n, n \geq 0)$  unabhängig von  $\mathfrak{A}_{n_0}$ .

Beweis: Die Zufallsgrößen

$$\tilde{X}_k := X_{n_0+k}, \quad k \geq 1$$

sind voneinander unabhängig und haben alle die gleiche Verteilung:

$P(\tilde{X}_k = x_k) = p^{\frac{1+x_k}{2}}(1-p)^{\frac{1-x_k}{2}}$ ,  $x_k \in \{+1, -1\}$ . Weiterhin ist

$$P(\tilde{X}_1 = x_1, \dots, \tilde{X}_n = x_n) = P(X_{n_0+1} = x_1, \dots, X_{n_0+n} = x_n) =$$

$$\prod_{k=1}^n P(X_{n_0+k} = x_k) = \prod_{k=1}^n P(X_k = x_k) = p^{\frac{n+s_n}{2}}(1-p)^{\frac{n-s_n}{2}} \text{ mit}$$

$$s_n = \sum_{k=1}^n x_k.$$

Wegen  $\tilde{S}_n = S_{n_0+n} - S_{n_0} = \sum_{k=1}^n \tilde{X}_k$ ,  $n \geq 0$  ist folglich  $(\tilde{S}_n)$  eine Irrfahrt mit dem Parameter  $p$ . Ist  $A \in \mathfrak{A}_{n_0}$ , so gibt es ein  $B \subseteq \{-1, +1\}^{n_0}$  mit  $A = \{(X_1, \dots, X_{n_0}) \in B\}$ . Da alle  $(X_k, k \geq 1)$  voneinander unabhängig sind, sind es auch die beiden Folgen  $(X_1, X_2, \dots, X_{n_0})$  und  $(X_{n_0+1}, X_{n_0+2}, \dots)$  und somit auch  $A$  und  $(\tilde{S}_n, n \geq 1)$ . Damit ist die Aussage bewiesen.  $\square$

Praktisch nach jedem Zeitpunkt  $n_0$  beginnt also die Irrfahrt bei  $S_{n_0}$  startend von Neuem, sie hat kein Gedächtnis.

## Die Verteilung von $S_n$

Es sei  $(S_n, n \geq 0)$  eine Irrfahrt mit dem Parameter  $p$ . Interpretieren wir  $S_n$  als Lage eines Teilchens zur Zeit  $n$ , so befindet sich das Teilchen zu geraden Zeitpunkten  $n$  in einem geradzahligen Punkt, zu ungeraden Zeitpunkten  $n$  in einem ungeradzahligen Punkt.

**Aussage 6.10** Für die Wahrscheinlichkeitsverteilung von  $S_n$  gilt:

a)  $P(S_0 = 0) = 1$

b)  $n = 2m, \quad m \geq 1,$

$$P(S_{2m} = 2k) = \binom{2m}{k+m} p^{m+k} (1-p)^{m-k}, \quad -m \leq k \leq m$$

$$P(S_{2m} = l) = 0 \quad \text{sonst.}$$

$$c) \quad n = 2m + 1, \quad m \geq 0,$$

$$P(S_{2m+1} = 2k + 1) = \binom{2m+1}{k+m+1} p^{k+m+1} (1-p)^{m-k}, \quad -m-1 \leq k \leq m$$

$$P(S_{2m+1} = l) = 0 \quad \text{sonst.}$$

**Beweis:**

$$b) \quad P(S_{2m} = 2k) = P(\{\omega \in \Omega \mid S_{2m}(\omega) = 2k\}) = \sum_{\omega \in \Omega: S_{2m}(\omega) = 2k} P(\{\omega\}).$$

$S_{2m}(\omega) = 2k$  ist genau dann der Fall, wenn in  $\omega = (x_1, x_2, \dots, x_{2m})$  genau  $(m+k)$ -mal die  $+1$  auftritt. Jedes solche  $\omega$  hat wegen (6.13) die gleiche Wahrscheinlichkeit  $p^{m+k}(1-p)^{m-k}$  und es gibt  $\binom{2m}{m+k}$  Folgen  $\omega$  dieser Art.

c) Der Beweis erfolgt analog zu b).

□

## Zeit des ersten Erreichens des Zustandes $m$

In diesem Abschnitt werden die Zeiten  $V_m$  studiert, zu denen die Irrfahrt zum ersten Mal den Zustand  $m$  ( $m \geq 1$ ) erreicht.

Interpretiert man die Irrfahrt  $(S_n, n \geq 0)$  mit  $S_0 = 0$ , so wie wir das bereits in Abschnitt 2.5 getan haben, als Guthaben eines Spielers, so ist es u. a. von

Interesse, wann dieses Guthaben zum ersten Mal positiv wird bzw. zum ersten Mal den Betrag  $m$  erreicht. Das sind gerade die eben erwähnten zufälligen Zeiten  $V_1$  bzw.  $V_m$ .

**Definition 6.11** *Es sei  $m \geq 1$ . Die Zufallsgröße*

$$V_m(\omega) := \min\{k \geq 1 | S_k(\omega) = m\} \text{ mit } \min \emptyset := \infty, \omega \in \Omega$$

*heißt die Zeit des ersten Erreichens des Punktes  $m$  durch die Irrfahrt  $(S_n, n \geq 0)$ .*

Zeiten des ersten Erreichens sind Beispiele sogenannter zufälliger Zeiten und spielen sowohl in der Theorie zufälliger Prozesse als auch für praktische Anwendungen eine wichtige Rolle.

Der Fall  $V_m(\omega) = \infty$  tritt genau dann ein, wenn  $S_n(\omega) < m$  für alle  $n \geq 1$  gilt. Weiterhin gilt offenbar

$$S_{V_m(\omega)}(\omega) = m, \text{ falls } \omega \in \{V_m < \infty\}. \quad (6.14)$$

Um Aussagen über die Wahrscheinlichkeitsverteilungen von  $V_1, V_2, \dots$  zu erhalten, führen wir die Zufallsgrößen  $V_m$  auf  $(S_n, n \geq 1)$  bzw.  $(X_k, k \geq 1)$  zurück, deren Wahrscheinlichkeitsverteilungen wir bereits kennen.

**Lemma 6.12** *Es gilt für  $k \geq 0, r > k$*

$$\{V_1 = 2k + 1\} = \{S_1 \leq 0, S_2 \leq 0, \dots, S_{2k} = 0, S_{2k+1} = 1\} \quad (6.15)$$

$$\{V_1 = 2k + 1, V_2 = 2r\} =$$

$$\{S_1 \leq 0, S_2 \leq 0, \dots, S_{2k-1} \leq 0, S_{2k} = 0, S_{2k+1} = 1, S_{2k+2} \leq 1, \dots$$

$$S_{2r-2} \leq 1, S_{2r-1} = 1, S_{2r} = 2\} \quad (6.16)$$

und allgemein,

$$\{V_1 = k_1, V_2 = k_2, \dots, V_m = k_m\} =$$

$$\{S_1 \leq 0, S_2 \leq 0, \dots, S_{k_1-1} = 0, S_{k_1} = 1, S_{k_1+1} \leq 1, \dots, S_{k_m-1} = m-1, S_{k_m} = m\},$$

$$1 \leq k_1 < k_2 < \dots < k_m, m \geq 1.$$

Beweis: Der Beweis ergibt sich unmittelbar aus der Definition der  $V_1, V_2, \dots, V_m$ .

Bemerkung: Die rechte Seite von (6.15) ist so zu verstehen, dass  $S_\ell \leq 0$  für alle  $\ell \leq 2k - 1$  gelten soll. Für den Fall  $k = 0$  ist diese Bedingung nicht relevant und entfällt ohne weiteren Hinweis. Genauso verstehen wir im Weiteren analoge Ausdrücke, z.B. in (6.16).

Mit der folgenden Aussage zeigen wir, dass die Irrfahrt, unter der Bedingung, dass sie den Punkt Eins niemals erreicht, nach der Zeit  $V_1$  wieder von Neuem als Irrfahrt mit dem gleichen Parameter im Punkt Eins beginnt.

Zur Formulierung der Aussage führen wir folgende Bezeichnungen ein:

Für alle  $\omega \in \{V_1 < \infty\}$  setzen wir

$$X_n^*(\omega) := X_{V_1(\omega)+n}(\omega), \quad n \geq 1,$$

und lassen  $X_n^*$  undefiniert auf  $\{V_1 = \infty\}$ . Dann ist  $X_n^*$  für  $P^*$ -fast alle  $\omega$  definiert, wobei  $P^*$  die bedingte Wahrscheinlichkeitsverteilung  $P(\cdot | V_1 < \infty)$  bezeichnet. (Beachte  $P(V_1 < \infty) \geq P(X_1 = 1) = p > 0$  und  $P^*(\cdot) = P(\cdot)$ , falls  $P(V_1 < \infty) = 1$ .) Es sei

$$S_n^* = \sum_{k=1}^n X_k^* \quad \text{und} \quad S_0^* = 0.$$

Auf Grund dieser Definition und wegen (6.14) gilt für alle  $n \geq 0$

$$S_n^* = S_{V_1+n} - S_{V_1} = S_{V_1+n} - 1, \quad P^* - \text{fast sicher.}$$

**Aussage 6.13** Die Folge  $(S_n^*, n \geq 0)$  ist bezüglich der Verteilung  $P^*$  eine Irrfahrt mit dem Parameter  $p$ .

Beweis:

$$\begin{aligned} P(V_1 < \infty, X_1^* = x_1, \dots, X_n^* = x_n) &= \\ \sum_{k=0}^{\infty} P(V_1 = 2k + 1, X_1^* = x_1, \dots, X_n^* = x_n) &= \\ \sum_{k=0}^{\infty} P(V_1 = 2k + 1, X_{2k+2} = x_1, \dots, X_{2k+1+n} = x_n) &= \\ \sum_{k=0}^{\infty} P(S_1 \leq 0, S_2 \leq 0, \dots, S_{2k} = 0, S_{2k+1} = 1, X_{2k+2} = x_1, \dots, X_{2k+1+n} = x_n). \end{aligned}$$

Jetzt nutzen wir die Unabhängigkeit von  $(S_1, S_2, \dots, S_{2k+1})$  und  $(X_{2k+2}, \dots, X_{2k+1+n})$  aus sowie die Tatsache, dass  $(X_{2k+2}, \dots, X_{2k+1+n})$  die gleiche Verteilung wie  $(X_1, \dots, X_n)$  besitzt und erhalten die Summe

$$\begin{aligned} \sum_{k=0}^{\infty} P(V_1 = 2k + 1)P(X_1 = x_1, \dots, X_n = x_n) &= \\ P(V_1 < \infty) \cdot P(X_1 = x_1, \dots, X_n = x_n) &= \\ P(V_1 < \infty) \cdot p^{\frac{n+s_n}{2}} (1-p)^{\frac{n-s_n}{2}}. \end{aligned}$$

Somit gilt

$$P^*(X_1^* = x_1, \dots, X_n^* = x_n) = p^{\frac{n+s_n}{2}} (1-p)^{\frac{n-s_n}{2}}. \quad (6.17)$$

□

Das Lemma besagt, nach der zufälligen Zeit  $V_1$  (sofern diese endlich ist) verhält sich die Irrfahrt  $(S_n, n \geq 0)$ , als startete sie von Neuem mit demselben Parameter  $p$ , dieses Mal vom Zustand Eins.

Wir bemerken noch eine Eigenschaft, die wir später benutzen werden. Es sei

$V_1^*(\omega) := \min\{k \geq 1 \mid S_k^*(\omega) = 1\}$ ,  $\min \emptyset = \infty$ . Dann gilt

$$V_1^* = V_2 - V_1 \quad \text{auf} \quad \{V_1 < \infty\}. \quad (6.18)$$

Im Fall  $V_1(\omega) < \infty$  haben wir nämlich

$$\begin{aligned} V_1^*(\omega) + V_1(\omega) &= \min\{k \geq 1 \mid S_k^*(\omega) = 1\} + V_1(\omega) = \\ &= \min\{k + V_1(\omega) \geq V_1(\omega) + 1 \mid S_{V_1(\omega)+k}(\omega) - S_{V_1(\omega)}(\omega) = 1\} = \\ &= \min\{m > V_1(\omega) \mid S_m(\omega) = 2\} = \\ &= \min\{m \geq 1 \mid S_m(\omega) = 2\} = V_2(\omega). \end{aligned}$$

Wir haben oben gesehen, dass die Irrfahrt  $(\tilde{S}_n, n \geq 0)$  mit  $\tilde{S}_n = S_{n+n_0} - S_{n_0}$  unabhängig von  $(X_1, \dots, X_{n_0})$ , also von  $\mathfrak{A}_{n_0} = \sigma(X_1, \dots, X_{n_0})$  ist. Ein analoger Sachverhalt gilt auch für  $(S_n^*, n \geq 0)$  mit  $S_n^* = S_{V_1+n} - S_{V_1}$ . Allerdings ist er etwas komplizierter in der Formulierung, da jetzt die Zeit  $V_1$  eine Zufallsgröße ist, die überdies den Wert Unendlich annehmen kann. Wir führen folgendes Ereignissystem aus  $\mathfrak{A}$  ein:

$$\mathfrak{A}_{V_1} := \{A \in \mathfrak{A} \mid A \cap \{V_1 = 2k + 1\} \in \mathfrak{A}_{2k+1}, k \geq 0\}.$$

**Lemma 6.14**  $\mathfrak{A}_{V_1}$  ist eine Teil- $\sigma$ -Algebra von  $\mathfrak{A}$  und  $V_1$  ist bezüglich  $\mathfrak{A}_{V_1}$  meßbar.

**Beweis:**  $\emptyset$  und  $\Omega$  gehören zu  $\mathfrak{A}_{V_1}$ , die übrigen Eigenschaften einer  $\sigma$ -Algebra folgen unter Zuhilfenahme der Tatsache, dass alle  $\mathfrak{A}_{2k+1}, k \geq 0$ ,  $\sigma$ -Algebren sind. Man beachte, dass  $\{V_1 = 2k + 1\} \in \mathfrak{A}_{2k+1}, k \geq 0$ , gilt. Die Meßbarkeit von  $V_1$  bez.  $\mathfrak{A}_{V_1}$  ergibt sich aus

$$\{V_1 = 2m + 1\} \cap \{V_1 = 2k + 1\} \in \mathfrak{A}_{2k+1}, k, m \geq 0.$$

□

Nun können wir die angekündigte Eigenschaft der Irrfahrt  $(S_n^*, n \geq 0)$  formulieren.

**Aussage 6.15** *Die Folge  $(S_n^*, n \geq 1)$  ist bezüglich  $P^*$ , also bezüglich  $P$  unter der Bedingung  $V_1 < \infty$ , eine Irrfahrt mit dem Parameter  $p$ , die unabhängig von  $\mathfrak{A}_{V_1}$  ist.*

**Beweis:** Es sei  $A \in \mathfrak{A}_{V_1}$ . Nach Definition von  $\mathfrak{A}_{V_1}$  gibt es für jedes  $k \geq 0$  eine Teilmenge  $B_k$  von  $\{-1, 1\}^{2k+1}$  mit  $A \cap \{V_1 = 2k + 1\} = \{(X_1, \dots, X_{2k+1}) \in B_k\}$ . Folglich gilt für alle  $x = (x_1, x_2, \dots, x_n) \in \{-1, +1\}^n$ :

$$\begin{aligned}
& P(A \cap \{X_1^* = x_1, \dots, X_n^* = x_n\} \cap \{V_1 < \infty\}) = \\
& \sum_{k=0}^{\infty} P(A \cap \{V_1 = 2k + 1\} \cap \{X_1^* = x_1, \dots, X_n^* = x_n\}) = \\
& \sum_{k=0}^{\infty} P((X_1, \dots, X_{2k+1}) \in B_k, X_{2k+2} = x_1, \dots, X_{2k+n+1} = x_n) = \\
& \sum_{k=0}^{\infty} P((X_1, \dots, X_{2k+1}) \in B_k) \cdot P(X_1 = x_1, \dots, X_n = x_n) = \\
& \sum_{k=0}^{\infty} P(\{V_1 = 2k + 1\} \cap A) P(X_1 = x_1, \dots, X_n = x_n) = \\
& P(A \cap \{V_1 < \infty\}) P(X_1 = x_1, \dots, X_n = x_n).
\end{aligned}$$

Teilen wir Anfangs- und Endterm dieser Gleichung durch  $P(V_1 < \infty)$  und berücksichtigen wir die Formel (6.17), so erhalten wir

$$\begin{aligned}
& P^*(A \cap \{X_1^* = x_1, \dots, X_n^* = x_n\}) = \\
& P^*(A) \cdot P^*(X_1^* = x_1, \dots, X_n^* = x_n).
\end{aligned}$$

□

**Folgerungen 6.16** Bezüglich  $P^*$ , also bezüglich  $P$  unter der Bedingung  $\{V_1 < \infty\}$ , sind  $V_1$  und  $V_2 - V_1$  unabhängige Zufallsgrößen mit

$$P^*(V_2 - V_1 = 2k + 1) = P^*(V_1^* = 2k + 1) = P(V_1 = 2k + 1), k \geq 0. \quad (6.19)$$

Beweis: Mittels Aussage 6.15 folgt wegen (6.18)  $\{V_1 = 2k + 1\} \in \mathfrak{A}_{V_1}$  die Gleichung

$$P^*(V_1 = 2k + 1, V_2 - V_1 = 2m + 1) =$$

$$P^*(V_1 = 2k + 1, V_1^* = 2m + 1) = P^*(V_1 = 2k + 1) \cdot P^*(V_1^* = 2m + 1), k, m \geq 0.$$

Durch Summation über  $k \geq 0$  und wegen  $P^*(V_1 < \infty) = 1$  und Aussage 6.13 ergibt sich die Folgerung.  $\square$

## Die Wahrscheinlichkeitsverteilung der zufälligen Zeit $V_1$

Wir bestimmen die Wahrscheinlichkeiten  $P(V_1 = 2k + 1)$  mittels der erzeugenden Funktion  $g_1$  von  $V_1$ :

$$g_1(s) := Es^{V_1} = \sum_{k=0}^{\infty} s^{2k+1} P(V_1 = 2k + 1), |s| \leq 1.$$

Es gilt  $P(V_1 = 1) = p$ , und für  $k \geq 1$  haben wir

$$P(V_1 = 2k + 1) =$$

$$P(X_1 = -1, S_2 \leq 0, \dots, S_{2k} = 0, S_{2k+1} = 1) =$$

$$P(X_1 = -1, S_2 - X_1 \leq 1, \dots, S_{2k} - X_1 = 1, S_{2k+1} - X_1 = 2) =$$

$$P(X_1 = -1)P(S_1 \leq 1, S_2 \leq 1, \dots, S_{2k-1} = 1, S_{2k} = 2) =$$

$$(1 - p)P(V_2 = 2k).$$

Somit ist mit der Bezeichnung  $q = 1 - p$

$$\begin{aligned} g_1(s) &= ps + qs \sum_{k=0}^{\infty} s^{2k} P(V_2 = 2k) \\ &= ps + qs E s^{V_2}, \quad |s| \leq 1. \end{aligned} \quad (6.20)$$

Wegen  $s^{V_2(\omega)} = 0$ , falls  $V_2(\omega) = \infty$  und  $|s| < 1$ , gilt

$$\begin{aligned} E s^{V_2} &= E[\mathbb{1}_{\{V_1 < \infty\}} s^{V_2}] = E[\mathbb{1}_{\{V_1 < \infty\}} s^{V_2 - V_1} s^{V_1}] = \\ E^* [s^{V_2 - V_1} s^{V_1}] P(V_1 < \infty) &= E^* s^{V_1^*} E^* s^{V_1} \cdot P(V_1 < \infty) = \\ E s^{V_1} \cdot E s^{V_1} &= [E s^{V_1}]^2 = g_1^2(s). \end{aligned} \quad (6.21)$$

Dabei wurden die Unabhängigkeit von  $V_1$  und  $V_1^* = V_2 - V_1$  bezüglich  $P^*$  (Folgerung 6.16), die Definition

$E^*[Z] = E[Z \cdot \mathbb{1}_{\{V_1 < \infty\}}] / P(V_1 < \infty)$  und die Gleichung

$E^* s^{V_1^*} = E s^{V_1}$  (Aussage 6.15) benutzt.

Wegen (6.20) und (6.21) genügt  $g_1(\cdot)$  der Gleichung

$$g_1^2(s) - \frac{1}{qs} g_1(s) + \frac{p}{q} = 0, \quad |s| < 1, s \neq 0.$$

Als Lösung ergibt sich auf Grund der Beschränktheit von  $g_1(\cdot)$  auf  $(-1, 1)$ :

$$g_1(s) = \frac{1}{2qs} [1 - (1 - 4pqs^2)^{\frac{1}{2}}], \quad |s| < 1, s \neq 0.$$

Anhand dieser erzeugenden Funktion berechnen wir die Einzelwahrscheinlichkeiten  $P(V_1 = 2k + 1)$ , die Wahrscheinlichkeit  $P(V_1 < \infty)$ , Eins jemals zu erreichen, und  $EV_1$ .

Zunächst gilt

$$P(V_1 < \infty) = \sum_{k=0}^{\infty} P(V_1 = 2k + 1) = \lim_{s \uparrow 1} g_1(s),$$

also

$$P(V_1 < \infty) = \frac{1 - (1 - 4pq)^{\frac{1}{2}}}{2q} = \frac{1 - |p - q|}{2q} = \begin{cases} 1 & \text{falls } p \geq \frac{1}{2}, \\ \frac{p}{q} & \text{falls } p < \frac{1}{2}. \end{cases} \quad (6.22)$$

Für die Einzelwahrscheinlichkeiten der Verteilung von  $V_1$  ergibt sich durch Entwicklung der erzeugenden Funktion  $g_1(\cdot)$  in eine Potenzreihe

$$P(V_1 = 2k - 1) = \frac{(2pq)^k}{2qk!} (2k - 3)!!, \quad k \geq 1 \quad (6.23)$$

mit  $m!! = 1 \cdot 3 \cdot 5 \cdot \dots \cdot m$  für ungerades  $m$  (Übung).

Wegen  $EV_1 = \lim_{s \uparrow 1} \frac{d}{ds} g_1(s)$  erhalten wir

$$EV_1 = \begin{cases} \frac{1}{|p - q|}, & \text{falls } p > q \\ \infty, & \text{falls } p \leq q \end{cases} \quad (6.24)$$

**Folgerungen 6.17** Für die symmetrische Irrfahrt  $(p = q) = \frac{1}{2}$  gelten die Gleichungen

$$P(V_1 < \infty) = 1 \text{ und } EV_1 = \infty. \quad (6.25)$$

# Kapitel 7

## Erwartungswert und Integral

Für diskret verteilte Zufallsgrößen haben wir Erwartungswerte in Kapitel vier kennengelernt. Der Begriff des Erwartungswertes war auch Grundlage für die Definition der Varianz einer Zufallsgröße, deren Momente sowie der Kovarianz zweier Zufallsgrößen.

Um sich von der Voraussetzung zu lösen, dass die zugrundeliegenden Zufallsgrößen diskret verteilt sind, erweitern wir den Begriff des Erwartungswertes auf eine möglichst große Klasse von Zufallsgrößen. Das gelingt mit Hilfe der Maß- und Integrationstheorie und soll in diesem Kapitel geschehen. Wir verzichten hier weitgehend auf Beweise, die Darstellung dient nur der Festlegung der Terminologie und der Vorstellung derjenigen Teile der Maß- und Integrationstheorie, die im Rahmen dieser Vorlesung benötigt werden. Für ausführlichere Darstellungen siehe die Vorlesung "Maßtheorie" sowie die Bücher von Bauer (1990), Jacod, Protter (2000) oder Siraev (1988).

### 7.1 Definitionen

Es sei  $(\Omega, \mathfrak{A}, P)$  ein Wahrscheinlichkeitsraum.

#### Einfache Zufallsgrößen

**Definition 7.1** *Eine reellwertige Zufallsgröße  $X$  über  $(\Omega, \mathfrak{A}, P)$  heißt einfach (in der Maßtheorie: Elementarfunktion) falls gilt*

$$X(\omega) = \sum_{i=1}^n a_i \mathbb{1}_{A_i}(\omega), \quad \omega \in \Omega \quad (7.1)$$

für gewisse  $n \geq 1, a_i \in R_1, A_i \in \mathfrak{A}, i = 1, 2, \dots, n$ .

Die Darstellung (7.1) ist nicht eindeutig, da die  $a_i$  nicht notwendig verschieden und die  $A_i$  nicht notwendig disjunkt sind. Man kann jedoch immer eine Darstellung finden mit  $a_i \neq a_j$  und  $A_i \cap A_j = \emptyset$  für  $i \neq j$ .

Jede einfache Zufallsgröße ist diskret verteilt mit der Menge der möglichen Werte  $\{a_i, i = 1, \dots, n\}$  und den Einzelwahrscheinlichkeiten  $P(X = a_i)$ .

## 1. Etappe: Erwartungswert einfacher Zufallsgrößen

**Definition 7.2** *Es sei  $X$  eine einfache Zufallsgröße über  $(\Omega, \mathfrak{A}, P)$  der Form (7.1). Als Erwartungswert von  $X$  oder als Integral über  $X$  bezüglich  $P$  bezeichnet man die Zahl*

$$EX := \sum_{i=1}^n a_i P(X = a_i) \quad (7.2)$$

Für  $EX$  schreibt man auch  $\int_{\Omega} X(\omega)P(d\omega)$  oder kurz  $\int_{\Omega} XdP$ .

Der Erwartungswert  $EX$  hängt nicht von der Darstellung (7.1) ab. Genauer,

gelten (7.1) und  $X(\omega) = \sum_{j=1}^m b_j \mathbb{1}_{B_j}(\omega), \omega \in \Omega$ , so haben wir

$$EX = \sum_{i=1}^n a_i P(X = a_i) = \sum_{j=1}^m b_j P(X = b_j).$$

Offenbar gelten

$$E\mathbb{1}_A = P(A), \quad A \in \mathfrak{A} \quad \text{und} \quad E\mathbb{1} = 1. \quad (7.3)$$

Die hier gegebene Definition stimmt mit der im Abschnitt 4.3. eingeführten Definition des Erwartungswertes diskret verteilter Zufallsgrößen überein. Einfache Zufallsgrößen sind diskret verteilt.

Die Menge aller einfachen Zufallsgrößen über  $(\Omega, \mathfrak{A}, P)$  bildet einen linearen Raum, d.h. mit  $X$  und  $Y$  sind auch alle Linearkombinationen  $\alpha X + \beta Y$  ( $\alpha, \beta \in \mathbb{R}_1$ ) einfache Zufallsgrößen.

Die Erwartungswertbildung ist eine *lineare Operation* auf diesem Raum, m.a.W. es gilt

$$E(\alpha X + \beta Y) = \alpha EX + \beta EY. \quad (7.4)$$

Außerdem ist die Erwartungswertbildung eine *monotone Operation*. Sind nämlich  $X$  und  $Y$  einfache Zufallsgrößen über  $(\Omega, \mathfrak{A}, P)$ , so gilt

$$X(\omega) \leq Y(\omega), \omega \in \Omega \Rightarrow EX \leq EY. \quad (7.5)$$

Zum Beweis von (7.4) und (7.5) wählt man eine Zerlegung  $\{C_i, i = 1, \dots, n\}$  von  $\Omega$  in Teilmengen  $C_i$  aus  $\mathfrak{A}$  mit  $X = \sum_1^n a_i \mathbb{1}_{C_i}$  und  $Y = \sum_1^n b_i \mathbb{1}_{C_i}$ .

## Nichtnegative Zufallsgrößen

Im nächsten Schritt werden wir den Begriff des Erwartungswertes auf nichtnegative Zufallsgrößen über  $(\Omega, \mathfrak{A}, P)$  erweitern. Dazu verwenden wir folgendes Lemma.

**Lemma 7.3** *Ist  $X$  eine nichtnegative Zufallsgröße über  $(\Omega, \mathfrak{A}, P)$ , so gibt es eine Folge  $(X_n)$  einfacher Zufallsgrößen über  $(\Omega, \mathfrak{A}, P)$  mit*

$$\left. \begin{array}{l} 0 \leq X_n(\omega) \leq X_{n+1}(\omega) \leq X(\omega), \omega \in \Omega, n \geq 1 \\ \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega), \omega \in \Omega. \end{array} \right\} \quad (7.6)$$

Beweis:

Man wähle für jedes  $n \geq 1$  und jedes  $\omega \in \Omega$

$$X_n(\omega) := \begin{cases} k \cdot 2^{-n}, & \text{falls } X(\omega) \in [k \cdot 2^{-n}, (k+1)2^{-n}) \\ & \text{und } 0 \leq k \leq n2^n - 1 \\ n, & \text{falls } X(\omega) \geq n. \end{cases}$$

Für jedes  $n \geq 1$  ist  $X_n$  eine Zufallsgröße, also Borel-messbar, da  $X$  es ist. Nunmehr ist (7.6) offensichtlich.  $\square$

Jede Folge  $(X_n)$  mit der Eigenschaft (7.6) nennen wir eine *die nichtnegative Zufallsgröße  $X$  approximierende Folge einfacher Zufallsgrößen*.

## 2. Etappe: Erwartungswert nichtnegativer Zufallsgrößen

**Definition 7.4** *Es seien  $X$  eine nichtnegative Zufallsgröße über  $(\Omega, \mathfrak{A}, P)$  und  $(X_n)$  eine  $X$  approximierende Folge einfacher Zufallsgrößen. Als Erwartungswert  $EX$  von  $X$  bezeichnen wir die Zahl*

$$EX := \lim_{n \rightarrow \infty} EX_n. \quad (7.7)$$

Der Erwartungswert  $EX$  existiert folglich für jede nichtnegative Zufallsgröße  $X$  und ist eventuell gleich Unendlich.

**Aussage 7.5** *Sind  $(X_n)$  und  $(X'_n)$  zwei die nichtnegative Zufallsgröße  $X$  approximierende Folgen, so gilt*

$$\lim_{n \rightarrow \infty} EX_n = \lim_{n \rightarrow \infty} EX'_n = EX$$

Für den Erwartungswert  $EX$  nichtnegativer Zufallsgrößen  $X$  gelten die Linearitätseigenschaft (7.4) (zumindest für  $\alpha, \beta \geq 0$ ) und die Monotonieeigenschaft (7.5) sinngemäß.

### 3. Etappe: Erwartungswert reellwertiger Zufallsgrößen

Im dritten und letzten Schritt erweitern wir den Erwartungswertbegriff auf reellwertige Zufallsgrößen über  $(\Omega, \mathfrak{A}, P)$ .

Ist  $X$  irgend eine solche Zufallsgröße, so zerlegt man sie durch

$$X = X^+ - X^-$$

mit  $X^+(\omega) := \max(X(\omega), 0)$  und  $X^-(\omega) := -\min(X(\omega), 0)$ ,  $\omega \in \Omega$ ,

in zwei nichtnegative Zufallsgrößen  $X^+$  und  $X^-$ . Wir bemerken, dass mit

$|X|(\omega) := |X(\omega)|$ ,  $\omega \in \Omega$ , außerdem die Gleichung  $|X| = X^+ + X^-$  gilt.

**Definition 7.6** *Man sagt, die Zufallsgröße  $X$  über  $(\Omega, \mathfrak{A}, P)$  hat einen endlichen Erwartungswert  $EX$ , falls  $EX^+ < \infty$  und  $EX^- < \infty$  gelten. Der Erwartungswert  $EX$  wird in diesem Fall definiert als  $EX := EX^+ - EX^-$ .*

*Gilt  $EX^+ < \infty$  oder  $EX^- < \infty$  so sagt man,  $X$  besitze einen Erwartungswert und setzt ebenfalls  $EX = EX^+ - EX^-$ . In diesem Fall kann  $EX = \infty$  bzw.  $-\infty$  gelten. Ist  $EX^+ = EX^- = \infty$ , so heißt es,  $X$  habe keinen Erwartungswert.*

In anderer Sprechweise sagt man, falls  $X$  einen endlichen Erwartungswert hat,  $X$  sei bezüglich  $P$  integrierbar und schreibt für  $EX$  auch

$$\int_{\Omega} X(\omega)P(d\omega) \text{ oder kurz } \int_{\Omega} X dP.$$

Der Erwartungswert  $EX$  von  $X$  wird in diesem Zusammenhang auch als *Integral über  $X$  bez.  $P$* , kurz  *$P$ -Integral über  $X$* , bezeichnet.

Existiert  $EX$ , so existiert für jedes  $A \in \mathfrak{A}$  auch  $E(X\mathbb{1}_A)$ , wir schreiben dafür auch  $\int_A X dP$ .

Gilt für eine nichtnegative Zufallsgröße  $X$  die Gleichung  $EX = 0$ , so folgt  $P(X = 0) = 1$ , die Zufallsgröße  $X$  hat also eine "entartete" Verteilung, sie nimmt mit Wahrscheinlichkeit Eins den Wert Null an. Ein Beispiel dafür haben wir bei der Einführung der gleichmäßigen Verteilung auf  $[0, 1)$  gesehen. Für

$X = \mathbb{1}_Q$  mit  $Q =$  Menge der rationalen Zahlen aus  $[0, 1)$  gilt  $EX = \lambda_{[0,1)}(Q) = 0$ . Die Abbildung  $X$  ist deswegen aber nicht identisch Null, sondern nur  $P$ -fast sicher gleich Null.

Insbesondere folgt für jede Zufallsgröße  $X$  mit  $E|X| = 0$  die Eigenschaft  $P(X = 0) = 1$ .

## $P$ -Äquivalenzklassen von Zufallsgrößen

**Definition 7.7** *Zwei Zufallsgrößen  $X$  und  $Y$  über  $(\Omega, \mathfrak{A}, P)$  heißen  $P$ -äquivalent oder einfach äquivalent, falls gilt*

$$P(\{\omega \in \Omega | X(\omega) \neq Y(\omega)\}) = 0.$$

*Alle zueinander  $P$ -äquivalenten Zufallsgrößen fasst man zu einer Äquivalenzklasse zusammen.*

Sind zwei Zufallsgrößen  $X$  und  $Y$   $P$ -äquivalent und existiert der Erwartungswert  $EX$ , so existiert auch  $EY$  und beide sind einander gleich. Der Erwartungswert ist also ein Funktional auf der Menge aller Äquivalenzklassen.

## 7.2 Einige Eigenschaften des Erwartungswertes

Es seien  $(\Omega, \mathfrak{A}, P)$  irgendein Wahrscheinlichkeitsraum und  $X, Y, \dots$  reellwertige Zufallsgrößen über  $(\Omega, \mathfrak{A}, P)$ .

Mit  $\mathfrak{L}_1(\Omega, \mathfrak{A}, P)$  bezeichnen wir die Menge aller reellwertigen Zufallsgrößen über  $(\Omega, \mathfrak{A}, P)$  mit endlichem Erwartungswert. Wir fassen einige Eigenschaften des Erwartungswertes in folgender Aussage zusammen.

### Aussage 7.8

- a)  $X, Y \in \mathfrak{L}_1 \implies \alpha X + \beta Y \in \mathfrak{L}_1$  und  $E(\alpha X + \beta Y) = \alpha EX + \beta EY$ , ( $\alpha, \beta \in \mathbb{R}_1$ )  
 ( $\mathfrak{L}_1$  ist ein linearer Raum und  $X \rightarrow EX$  ein lineares Funktional auf  $\mathfrak{L}_1$ )

- b)  $X \in \mathfrak{L}_1, X \leq Y$  P – f.s.  $\implies EX \leq EY \leq \infty$   
*(insbesondere folgt aus  $Y \geq 0$  P – f.s. die Ungleichung  $EY \geq 0$ )*
- c)  $X \in \mathfrak{L}_1 \iff |X| \in \mathfrak{L}_1$ , in diesem Fall gilt  $|EX| \leq E|X|$
- d) Ist  $X$  P – f.s. beschränkt ( $|X| \leq C$  P – f.s. für ein  $C > 0$ ), so besitzt  $X$  einen endlichen Erwartungswert  $EX$ .

## Ungleichungen

Im Folgenden stellen wir einige Ungleichungen den Erwartungswert von Zufallsgrößen betreffend zusammen, die in der Wahrscheinlichkeitstheorie relevant sind.

- a) *Ungleichung von Tschebychev:*  
 Ist  $X$  eine nichtnegative Zufallsgröße, so gilt für jeden  $\varepsilon > 0$

$$P(X \geq \varepsilon) \leq \frac{EX}{\varepsilon} \quad (7.8)$$

Beweis:

$$EX \geq E(\mathbb{1}_{\{X \geq \varepsilon\}} \cdot X) \geq E\varepsilon \mathbb{1}_{\{X \geq \varepsilon\}} = \varepsilon P(X \geq \varepsilon)$$

□

- b) *Ungleichung von Cauchy-Schwarz:*

Ist  $E(X^2) < \infty$  und  $E(Y^2) < \infty$ , dann gilt  $E|XY| < \infty$   
 und

$$(E(XY))^2 \leq E(X^2) \cdot E(Y^2) \quad (7.9)$$

Das Gleichheitszeichen gilt genau dann, wenn  $aX + bY = 0$  für gewisse  $a, b \in R_1$  P – f.s. gilt.

Beweis: O.B.d.A. sei  $EX^2 > 0, EY^2 > 0$ , wir setzen

$$\tilde{X} := \frac{X}{\sqrt{EX^2}}, \tilde{Y} := \frac{Y}{\sqrt{EY^2}}.$$

Wegen  $E(\tilde{X} - \tilde{Y})^2 \geq 0$ ,  $E(\tilde{X} + \tilde{Y})^2 \geq 0$ , und  $E\tilde{X}^2 = E\tilde{Y}^2 = 1$

gilt

$$-1 \leq E\tilde{X}\tilde{Y} \leq 1$$

mit  $|E\tilde{X}\tilde{Y}| = 1$  genau dann, wenn  $\tilde{X} = \tilde{Y}$  oder  $\tilde{X} = -\tilde{Y}$  P - f.s.

Daraus folgt die Behauptung.  $\square$

Bemerkung: Die Ungleichung (7.9) bleibt erhalten, wenn man auf der linken Seite  $E|XY|$  an Stelle  $E(XY)$  setzt. Der Beweis verläuft analog.

c) *Ungleichung von Jensen:*

Es seien  $g$  eine von unten konvexe und Borel-messbare Funktion auf  $R_1$  und  $X$  eine reellwertige Zufallsgröße mit  $E|X| < \infty$ . Dann gilt

$$g(EX) \leq Eg(X) \leq \infty. \quad (7.10)$$

Beweis: Da  $g$  von unten konvex ist, gibt es zu jedem  $x_0 \in R_1$  eine Zahl  $\lambda(x_0)$  mit

$$g(x_0) + (x - x_0)\lambda(x_0) \leq g(x).$$

Wir setzen  $x = X, x_0 = EX$  und erhalten damit

$$g(EX) + (X - EX)\lambda(EX) \leq g(X),$$

daraus folgt  $g(EX) \leq Eg(X)$ .  $\square$

Die Jensen'sche Ungleichung impliziert zwei weitere Ungleichungen, die wir hier nur angeben, für einen Beweis siehe z. B. Siraev (1988), Kap. II, § 6.

- d) *Hölder-Ungleichung*: Es sei  $1 < p < \infty, 1 < q < \infty, \frac{1}{p} + \frac{1}{q} = 1$ . Wenn  $E|X|^p < \infty, E|Y|^q < \infty$ , so ist  $E|XY| < \infty$ , und es gilt

$$E|XY| \leq (E|X|^p)^{\frac{1}{p}} (E|Y|^q)^{\frac{1}{q}}$$

( $p = q = 2$ : Cauchy-Schwarz-Ungleichung)

- e) *Minkovski-Ungleichung*: Wenn  $E|X|^p < \infty, E|Y|^p < \infty$ , für ein  $p$  mit  $1 \leq p < \infty$ , dann gilt  $E|X + Y|^p < \infty$  und

$$(E|X + Y|^p)^{\frac{1}{p}} \leq (E|X|^p)^{\frac{1}{p}} + (E|Y|^p)^{\frac{1}{p}}.$$

## Die Räume $L^p$

Es sei  $p \in [1, \infty)$  und  $\mathfrak{L}^p(\Omega, \mathfrak{A}, P)$  die Menge aller reellwertigen Zufallsgrößen  $X$  über  $(\Omega, \mathfrak{A}, P)$  mit  $E(|X|^p) < \infty$ . Die Menge aller Äquivalenzklassen von Zufallsgrößen  $X$  aus  $\mathfrak{L}^p$  werde mit  $L^p(\Omega, \mathfrak{A}, P)$  bezeichnet (siehe Definition 7.6).

**Aussage 7.9** *Es sei  $p \in [1, \infty)$*

- a) *Die Menge  $\mathfrak{L}^p = \mathfrak{L}^p(\Omega, \mathfrak{A}, P)$  ist ein linearer Raum.*  
 b)  *$L^p(\Omega, \mathfrak{A}, P)$  ist mit der Norm*

$$\|X\|_p := (E|X|^p)^{\frac{1}{p}}, \quad X \in L^p$$

*ein normierter Raum, sogar ein Banachraum.*

- c) *Es gilt für alle  $p, p'$  mit  $1 \leq p < p' < \infty$*

$$L^{p'}(\Omega, \mathfrak{A}, P) \subseteq L^p(\Omega, \mathfrak{A}, P) \text{ und}$$

$$\|X\|_p \leq \|X\|_{p'}, \quad X \in L^{p'}, \quad (\text{Ungleichung von Ljapunov}) \quad (7.11)$$

*Insbesondere gilt*

$$E|X| \leq (EX^2)^{\frac{1}{2}}. \quad (7.12)$$

## Vertauschung von Grenzwert und Erwartungswert

Es sei  $(X_n, n \geq 1)$  eine Folge reellwertiger Zufallsgrößen über  $(\Omega, \mathfrak{A}, P)$ .

**Definition 7.10** *Man sagt, die Folge  $(X_n, n \geq 1)$  konvergiert  $P$ -fast sicher gegen eine Zufallsgröße  $X$  über  $(\Omega, \mathfrak{A}, P)$ , falls*

$$P(\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1$$

Die folgenden drei Aussagen betreffen das Verhältnis zwischen Grenzwerten und Erwartungswerten.

**Aussage 7.11** *(Satz von der majorisierten Konvergenz) Konvergiert  $(X_n, n \geq 1)$   $P$ -fast sicher gegen  $X$  und gibt es eine  $P$ -integrierbare Zufallsgröße  $Z$  mit  $|X_n| \leq Z$   $P$ -f.s. für alle  $n \geq 1$ , so ist auch  $X$  bezüglich  $P$  integrierbar, und es gilt*

$$\lim_{n \rightarrow \infty} EX_n = E \lim_{n \rightarrow \infty} X_n = EX.$$

**Aussage 7.12** *(Satz von der monotonen Konvergenz) Ist  $(X_n, n \geq 1)$  eine monoton wachsende Folge  $P$ -integrierbarer Zufallsgrößen, so gilt für  $X := \lim_{n \rightarrow \infty} X_n$  die Beziehung*

$$EX = \lim_{n \rightarrow \infty} EX_n \leq \infty.$$

**Aussage 7.13 (Lemma von Fatou):** *Sind  $Y$  und  $Z$  zwei  $P$ -integrierbare Zufallsgrößen, so gilt*

$$X_n \leq YP - \text{f.s. für alle } n \geq 1 \implies E(\limsup_{n \rightarrow \infty} X_n) \geq \limsup_{n \rightarrow \infty} EX_n$$

$$X_n \geq ZP - \text{f.s. für alle } n \geq 1 \implies E(\liminf_{n \rightarrow \infty} X_n) \leq \liminf_{n \rightarrow \infty} EX_n$$

Die folgende Aussage gestattet es, die Berechnung des Erwartungswertes einer Zufallsgröße auf ein Integral bezüglich ihrer Wahrscheinlichkeitsverteilung zurückzuführen.

**Aussage 7.14 (Substitutionsformel):** Es sei  $X$  eine Zufallsgröße über  $(\Omega, \mathfrak{A}, P)$  mit Werten in  $(E, \mathfrak{E})$  und der Wahrscheinlichkeitsverteilung  $P^X$  (siehe ...). Weiterhin sei  $h$  eine  $\mathfrak{E} - \mathfrak{B}_1$ -messbare Abbildung von  $E$  in  $\mathbb{R}_1$ . Dann gilt:

- a)  $h(X)$  ist  $P$ -integrierbar genau dann, wenn  $h(\cdot)$  bezüglich  $P^X$  integrierbar ist.
- b) Im Falle von a) gilt

$$Eh(X) = \int_{\Omega} h(X(\omega))P(d\omega) = \int_E h(x)P^X(dx). \quad (7.13)$$

Beweis: Wir gehen zurück auf die Definition von  $P^X$ . Es gilt

$$P^X(B) = P(X^{-1}(B)), \quad B \in \mathfrak{E}.$$

Daraus folgt

$$E(\mathbb{1}_B(X)) = P(X^{-1}(B)) = P^X(B) = \int_E \mathbb{1}_B(x)P^X(dx) \quad (7.14)$$

Ist  $h$  eine einfache Funktion (Elementarfunktion, endliche Linearkombination aus messbaren Indikatorfunktionen) so folgt aus (7.14) die Eigenschaft (7.13) auf Grund der Linearität der Erwartungswertoperation.

Wenn  $h$  nichtnegativ ist, so wählen wir eine  $h$  approximierende Folge  $h_n$  einfacher Funktionen:

$$0 \leq h_n \leq h_{n+1} \leq h$$

$$\lim_{n \rightarrow \infty} h_n(x) = h(x), \quad x \in E.$$

Dann gilt  $h_n(X) \uparrow h(X)$  und wegen des Satzes (7.11) von der monotonen Konvergenz (zweimal angewandt)

$$Eh(X) = E \lim h_n(X) = \lim Eh_n(X) = \lim \int_E h_n(x)P^X(dx) = \int_E h(x)P^X(dx)$$

Das beweist a) und b) für nichtnegative  $h$ . Für beliebiges  $h$  benutzen wir wieder die Zerlegung  $h = h^+ - h^-$ .  $\square$

## Varianz, Kovarianz und Korrelation

Wir haben bei diskret verteilten Zufallsgrößen gesehen, dass zur Beurteilung einer Wahrscheinlichkeitsverteilung neben dem Erwartungswert, der den "Schwerpunkt" der Verteilung beschreibt, auch die Varianz oder Streuung von Bedeutung ist. Sie ist eine Maßzahl, wie breit die möglichen Werte der Zufallsgröße um den Erwartungswert (mit ihren Wahrscheinlichkeiten gewichtet) gelagert sind bzw. wie stark Realisierungen einer zugrundeliegenden Zufallsgröße um ihren Mittelwert "streu"en".

Der Begriff der Varianz oder der Streuung überträgt sich mit dem nunmehr bereit stehenden Begriff des Erwartungswertes beliebiger Funktionen von Zufallsgrößen problemlos auf unseren allgemeinen Fall.

**Definition 7.15** Für jedes  $X \in L_2$  wird durch

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

die Varianz (oder die Streuung) von  $X$  definiert. Sie wird häufig auch mit  $\sigma_X^2$  bzw. einfach mit  $\sigma^2$  bezeichnet. Die Zahl  $\sigma_X = (\sigma_X^2)^{\frac{1}{2}}$  heißt Standardabweichung der Zufallsgröße  $X$ .

Es gilt

$$D^2(X) = E((X - EX)^2) = EX^2 - 2EXEX + (EX)^2 = EX^2 - (EX)^2 \quad (7.15)$$

### Die Wirkung linearer Transformationen

Hat die Zufallsgröße  $X$  einen endlichen Erwartungswert, so gilt für alle  $a, b \in \mathbb{R}_1$  die Gleichung  $E(aX + b) = aEX + b$ .

Ist  $D^2X < \infty$ , so besitzt für jede reelle Zahl  $a$  die Zufallsgröße  $aX$  die Varianz

$$D^2(aX) = a^2 D^2X,$$

und für jedes  $b \in R_1$  gilt

$$D^2(aX + b) = a^2 D^2 X.$$

Ist  $D^2 X > 0$ , so bildet

$$X^* := \frac{X - EX}{\sqrt{D^2 X}} \quad (7.16)$$

eine *standardisierte* Zufallsgröße, d. h., es gilt

$$EX^* = 0 \text{ und } D^2 X^* = 1.$$

**Bemerkung 7.16** Hat eine Zufallsgröße  $X$  eine positive Streuung  $D^2 X$  (oder ist diese gleich Unendlich), so handelt es sich um eine echte Zufallsgröße in dem Sinne, dass ihr Wert vor Ausführung des zugrunde liegenden Experimentes unbestimmt ist. Ihre möglichen Werte besitzen eine "echte" Wahrscheinlichkeitsverteilung, die Gesamtwahrscheinlichkeit Eins verteilt sich auf mehrere verschiedene mögliche Werte.

Dagegen gilt  $D^2 X = 0$  genau dann, wenn  $P(X = EX) = 1$  erfüllt ist, wenn also  $X$  mit Wahrscheinlichkeit Eins nur einen einzigen Wert annehmen kann, der dann natürlich der Erwartungswert von  $X$  ist.

Aus Formel (7.8) folgt die

**Aussage 7.17 (Tschebyschev'sche Ungleichung)** *Ist  $D^2 X < \infty$ , so gilt für jedes  $\varepsilon > 0$*

$$P(|X - EX| \geq \varepsilon) \leq \frac{D^2 X}{\varepsilon^2}.$$

Ist die Streuung  $D^2 X$  positiv aber klein, so besagt die Tschebyschev'sche Ungleichung, dass die möglichen Werte von  $X$ , die weit von  $EX$  entfernt liegen, bei einer Realisierung der Zufallsgröße  $X$  nur mit sehr kleiner Wahrscheinlichkeit (die aber durchaus positiv ist) auftreten werden.

### 7.3 Dichten eindimensionaler Verteilungen

In diesem und im folgenden Abschnitt erweitern wir den in Abschnitt 3.5 eingeführten Begriff der Dichte einer Wahrscheinlichkeitsverteilung. Wir stützen uns dabei auf Vorkenntnisse über das Lebesguemaß  $\lambda$  auf  $(R_1, \mathcal{B}_1)$  aus der Maßtheorie-Vorlesung.

Die Integration über reellwertige Borel-messbare Funktionen  $f$  auf  $R_1$  bez. des Lebesguemaßes definiert man völlig analog zur Definition des Erwartungswertes, d. h. des Integrales bezüglich des Wahrscheinlichkeitsmaßes  $P$  in Abschnitt 7.1.

Statt  $\int_{R_1} f(x)\lambda(dx)$  schreiben wir  $\int_{R_1} f(x)dx$ .

**Definition 7.18** *Ist  $Q$  ein Wahrscheinlichkeitsmaß auf  $(R_1, \mathcal{B}_1)$ , und existiert eine nichtnegative Borelfunktion  $f$  auf  $R_1$ , so dass*

$$F_Q(x) := Q((-\infty, x]) = \int_{(-\infty, x]} f(y)dy = \int_{R_1} f(y)\mathbb{1}_{(-\infty, x]}(y)dy, x \in R_1 \quad (7.17)$$

*gilt, so heißt  $f$  die Dichte des Maßes  $Q$ . Ist  $Q = P^X$  für eine Zufallsgröße  $X$ , so nennt man  $f$  auch die Dichte der Zufallsgröße  $X$ .*

Aus (7.17) folgt wie üblich mit Hilfe des Erweiterungssatzes für  $\sigma$ -additive Mengenfunktionen

$$Q(B) = \int_{R_1} f(y)\mathbb{1}_B(y)dy =: \int_B f(y)dy \quad (7.18)$$

für jedes  $B \in \mathcal{B}_1$ .

**Aussage 7.19** *Genau dann besitzt eine Wahrscheinlichkeitsverteilung  $Q$  auf  $(R_1, \mathcal{L}_1)$  eine Dichte  $f$  wenn ihre Verteilungsfunktion  $F_Q$  Lebesgue-fast überall differenzierbar ist. In diesem Fall gilt  $\frac{dF_Q}{dx} = f(x)$  Lebesgue-fast überall.*

Der Beweis dieser Aussage ist Gegenstand der Analysis monotoner Funktionen auf  $R_1$ , siehe z. B. I.P.Natanson, Theorie der Funktionen einer reellen Veränderlichen, Akademie Verlag, 1961.

**Aussage 7.20** Eine nichtnegative Borelmeßbare Funktion  $f$  auf  $R_1$  ist die Dichte einer Wahrscheinlichkeitsverteilung  $Q$  auf  $\mathcal{B}_1$  genau dann, wenn gilt

$$\int_{R_1} f(x)dx = 1.$$

Die Verteilung  $Q$  ist in diesem Fall durch die Formel in (7.18) gegeben. Besitzt eine Wahrscheinlichkeitsverteilung  $Q$  auf  $(R_1, \mathcal{B}_1)$  eine Dichte  $f$ , so bestimmt  $f$  das Maß  $Q$  eindeutig. Andererseits ist für je zwei Dichten  $f_1$  und  $f_2$  von  $Q$

$$\lambda(\{x \in R_1 | f_1(x) \neq f_2(x)\}) = 0,$$

d. h.,  $f_1$  und  $f_2$  sind Lebesgue - fast überall gleich.

Beweisskizze:

$$\int_{R_1} f(x)dx = \lim_{y \rightarrow \infty} Q((-\infty, y]) = 1.$$

Wenn  $f \geq 0$  gegeben ist, so setzt man

$$Q(B) := \int_{R_1} f(x) \mathbb{1}_B(x) dx = \int_B f(x) dx.$$

Sind  $f_1$  und  $f_2$  Dichten von  $Q$ , so gilt

$$\int_{R_1} \mathbb{1}_{\{f_1 < f_2\}}(x) (f_2(x) - f_1(x)) dx = 0,$$

folglich ist  $\lambda(\{f_1 < f_2\}) = 0$ , und somit auch  $\lambda(\{f_1 \neq f_2\}) = 0$ .

Mit Hilfe der folgenden Aussage gelingt es, Erwartungswerte der Form  $Eg(X)$  auf Integrale bezüglich des Lebesguemaßes zurückzuführen.

**Aussage 7.21** *Es sei  $X$  eine reellwertige Zufallsgröße mit der Dichte  $f$ . Ist  $g$  eine Borel-messbare Funktion auf  $R_1$ , so gilt*

- a)  *$g(\cdot)$  ist bezüglich  $P^X$  integrierbar genau dann, wenn  $g(\cdot)f(\cdot)$  bezüglich des Lebesguemaßes integrierbar ist,*
- b) *im Fall a) gilt*

$$Eg(X) = \int_{R_1} g(x)P^X(dx) = \int_{R_1} g(x)f(x)dx. \quad (7.19)$$

Beweis: Für  $g = \mathbb{1}_B$  mit  $B \in \mathcal{B}_1$  hat (7.19) die Form  $E\mathbb{1}_B(X) = P^X(B) = \int_{R_1} \mathbb{1}_B(x)f(x)dx$ .

Diese Gleichung ist aber auf Grund von (7.18) und  $E\mathbb{1}_B(X) = P(X \in B)$  richtig, man setze  $Q = P^X$ .

Wegen der Linearität der Erwartungswertbildung folgt damit (7.19) für alle einfachen Funktionen  $g(\cdot)$  (Elementarfunktionen). Für allgemeines nichtnegatives  $g$  ergibt sich (7.19) und auch a) aus dem Satz über die monotone Konvergenz.

Der Fall beliebiger Funktionen  $g$  folgt wie üblich mittels  $g = g^+ - g^-$ .  $\square$

## Lebesgue- und Riemannintegrale

Der Einfachheit und Allgemeinheit der Definition von  $P$ -Integralen steht die Kompliziertheit ihrer konkreten Ausrechnung auf der Grundlage ihrer Definition gegenüber. Andererseits verfügt man mit der Theorie des Riemannintegrals und seiner zahlreichen Berechnungsmethoden über ein sehr leistungsfähiges Werkzeug zur Berechnung von Integralen. Wir geben im Folgenden die Beziehungen zwischen beiden Integralarten an und gewinnen damit die Möglichkeit, in vielen Fällen Erwartungswerte, Streuungen und andere Kenngrößen von Verteilungen konkret ausrechnen zu können. Die Beweise findet man in der Literatur zur Maß- und Integrationstheorie, siehe z. B. Elstrodt (1996), Bauer (1992) oder die Vorlesung Maßtheorie.

**Aussage 7.22** *Es sei  $f$  eine beschränkte Borel-messbare Funktion auf dem endlichen Intervall  $[a, b]$ . Dann gilt:*

- a)  $f$  ist  $L$ -integrierbar,
- b)  $f$  ist  $R$ -integrierbar genau dann, wenn  $\{x \in [a, b] : f \text{ ist unstetig bei } x\}$  das Lebesguemaß Null hat.

Im Fall b) gilt

$$(R) - \int_{[a,b]} f(x)dx = (L) - \int_{[a,b]} f(x)dx.$$

Im Fall eines unendlichen Integrationsbereiches, z. B.  $I = (\infty, a]$ ,  $I = (a, \infty)$  oder  $I = R_1$ , hat man für Funktionen  $f$ , die auf jedem kompakten Intervall  $[a, b] \subseteq I$  Riemannintegrierbar sind, den Begriff des *uneigentlichen Riemannintegrals*. Man sagt (hier für  $I = R_1$  aufgeschrieben), das uneigentliche Riemannintegral über  $f$  existiert, falls der Grenzwert

$$(R) - \int_{-\infty}^{\infty} f(x)dx := \lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} (R) - \int_{[a,b]} f(x)dx,$$

existiert und endlich ist.

Wir vergleichen uneigentliche Riemannintegrale mit Lebesgueintegralen und bemerken als Erstes die folgende

**Aussage 7.23** *Ist  $f$  eine nichtnegative Funktion auf  $I_{[a,\infty)}$ , und ist  $f$  auf jedem Intervall  $[a, b]$  für  $b > a$   $R$ -integrierbar, so gilt*

$$\lim_{b \rightarrow \infty} (R) - \int_{[a,b]} f(x)dx = (L) - \int_{R_1} f(x)dx.$$

*Anders ausgedrückt, das uneigentliche  $R$ -Integral über eine nichtnegative Funktion  $f$  existiert genau dann, wenn das  $L$ -Integral existiert und endlich ist. In diesem Fall sind beide gleich.*

Beweis: Die Folge  $(f_n)$ , definiert durch  $f_n := f \cdot \mathbb{1}_{[a,n]}$  konvergiert monoton gegen  $f$ . Die Aussagen 7.12 und 7.21 implizieren

$$(L) - \int_{[a,\infty)} f(x)dx = \lim_{n \rightarrow \infty} (L) - \int_{[a,\infty)} f_n(x)dx =$$

$$\lim_{n \rightarrow \infty} (R) - \int_{[a,n]} f(x)dx = (R) - \int_{[a,\infty)} f(x)dx.$$

□

Wir setzen den Vergleich beider Integralarten fort mit der folgenden Bemerkung:

Das uneigentliche Riemannintegral kann existieren und endlich sein, obwohl  $f$  nicht Lebesgueintegrierbar ist.

**Beispiel 7.24** Für  $f$ , definiert durch  $f(x) = \frac{\sin x}{x}$ ,  $x > 0$ , gilt

$$(R) - \int_0^{\infty} f(x)dx = \sum_{k=0}^{\infty} \int_{[\pi k, \pi(k+1)]} \frac{\sin x}{x} dx.$$

Die Reihe konvergiert, da sie alternierend ist und die Reihenglieder gegen Null konvergieren. Das Lebesgueintegral  $(L) - \int_{[0,\infty)} f(x)dx$  existiert nicht, da  $f^+$  und  $f^-$  kein endliches Lebesgueintegral besitzen.

Die folgende Aussage gibt eine Bedingung an, unter der Erwartungswerte der Form  $Eh(X)$  mit Hilfe von Riemannintegralen berechnet werden können.

**Aussage 7.25** *Es sei  $X$  eine reellwertige Zufallsgröße mit der Dichte  $f$  und  $h$  eine Funktion von  $R_1$  in sich.*

*Sind  $h$  und  $f$  Lebesgue-fast-überall stetig und ist  $h$  nichtnegativ, so gilt*

$$Eh(X) = (R) - \int_{R_1} h(x)f(x)dx \quad (7.20)$$

Die Gleichung (7.20) gilt auch für  $h$  mit  $E|h(X)| < \infty$  oder, äquivalent,

$$(R) - \int_{R_1} |h(x)|f(x)dx < \infty.$$

Der Beweis ergibt sich unmittelbar aus Aussage 7.22.

Es sei  $X$  eine reellwertige Zufallsgröße mit der Dichte  $f$ .

**Folgerung 7.26** Das  $n$ -te Moment  $\mu_n := E(X^n)$  der Zufallsgröße  $X$  existiert und ist endlich genau dann, wenn  $\int_{R_1} x^n f(x)dx$  existiert und endlich ist. In diesem Fall gilt

$$E(X^n) = \int_{R_1} x^n f(x)dx. \quad (7.21)$$

Insbesondere ergibt sich

$$EX = \int_{R_1} xf(x)dx \text{ und } D^2X = \int_{R_1} (x - EX)^2 f(x)dx. \quad (7.22)$$

Dabei sind die Integrale als Lebesgueintegrale zu verstehen, die unter geeigneten Voraussetzungen (s. oben) auch zu Riemannintegralen werden.

Beispiele für Dichten auf  $R_1$ 

Obwohl jede nichtnegative Borel-messbare Funktion  $f$  mit  $\int_{R_1} f(x)dx = 1$  Dichte einer Wahrscheinlichkeitsverteilung auf  $(R_1, \mathcal{B}_1)$  ist, sind viele theoretisch und praktisch wichtige Dichten stetig oder stückweise stetig. Wir geben einige davon an. Ihre Bedeutung wird im weiteren Verlauf der Vorlesung noch diskutiert.

Es sei  $X$  eine reellwertige Zufallsgröße mit der Dichte  $f$ .  
Man sagt  $X$  besitze eine

- a) *gleichmäßige Verteilung auf  $[a, b]$* , falls

$$f(x) = \mathbb{1}_{[a,b]}(x) \frac{1}{b-a}, \quad x \in R_1.$$

Bezeichnung:  $X \sim U([a, b])$

$$EX = \frac{a+b}{2}, \quad D^2X = \frac{(b-a)^2}{12}$$

- b) *Exponentialverteilung mit dem Parameter  $\lambda$  ( $\lambda > 0$ )*, falls

$$f(x) = \lambda \cdot \mathbb{1}_{[0,\infty)}(x) \exp(-\lambda x), \quad x \in R_1.$$

Bezeichnung:  $X \sim Exp(\lambda)$

$$EX = \frac{1}{\lambda}, \quad D^2X = \frac{1}{\lambda^2}$$

- c) *Gammaverteilung mit den Parametern  $\alpha, \lambda$  ( $\alpha > 0, \lambda > 0$ )*, falls

$$f(x) = \mathbb{1}_{(0,\infty)}(x) \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x \in R.$$

Bezeichnung:  $X \sim \Gamma(\alpha, \lambda)$

$$EX = \frac{\alpha}{\lambda}, \quad D^2X = \frac{\alpha}{\lambda^2}$$

Für  $\alpha = \frac{n}{2}, \lambda = \frac{1}{2}$  ist diese Verteilung auch als " $\chi^2$ -Verteilung mit  $n$  Freiheitsgraden" ( $n \geq 1$ ) bekannt.

- d) *Normal- oder Gaußsche Verteilung mit den Parametern  $\mu$  und  $\sigma^2$  ( $\mu \in R_1, \sigma^2 > 0$ ),*

$$\text{falls } f(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right), x \in R_1$$

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

$$EX = \mu, \quad D^2X = \sigma^2$$

Im Fall  $\mu = 0, \sigma^2 = 1$  spricht man von einer "*Standardnormalverteilung*". Ihre Verteilungsfunktion wird mit  $\Phi$  bezeichnet:

$$\Phi(x) = P(X \leq x) = \int_{(-\infty, x]} (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{y^2}{2}\right) dy, \quad x \in R_1.$$

Sie ist nicht explizit berechenbar und ist deshalb vertafelt.

- e) *Cauchyverteilung, mit dem Parameter  $a \in R_1$  falls*

$$f(x) = \frac{1}{\pi} \frac{1}{1+(x-a)^2}, x \in R_1.$$

Erwartungswert und Streuung der Cauchyverteilung existieren nicht.

## Transformationssatz für Dichten

Es sei  $X$  eine reellwertige Zufallsgröße über  $(\Omega, \mathfrak{A}, P)$  mit der Dichte  $f$ . Häufig hat man die Verteilung einer Zufallsgröße  $Y$  zu berechnen, die eine Funktion von  $X$  ist. Dazu nehmen wir an,  $h$  sei eine Borel-messbare Funktion von  $R_1$  in sich, und es gelte

$$Y(\omega) := h(X(\omega)), \quad \omega \in \Omega.$$

Offenbar gilt für die Verteilungsfunktion  $F_Y$  von  $Y$

$$F_Y(y) = P(Y \leq y) = P(\{\omega \in \Omega | h(X(\omega)) \leq y\}) =$$

$$P(\{\omega \in \Omega | X(\omega) \in h^{-1}((-\infty, y])\}) = P(\{\omega \in \Omega | X(\omega) \in \{x : h(x) \leq y\}\}) =$$

$$\int_{\{x: h(x) \leq y\}} f(s) ds. \quad (7.23)$$

Aus dieser Gleichung gewinnen wir folgende

**Aussage 7.27** *Ist  $f_X$  eine stetige Dichte von  $X$ ,  $\{x \in R_1, |f(x) > 0\}$  ein Intervall  $I$ , und ist  $h$  eine stetig differenzierbare, streng monotone Funktion von  $I$  in  $R_1$  mit  $h'(x) \neq 0$  für alle  $x \in I$ , gilt  $Y = h(X)$  und setzt man  $g(y) = h^{-1}(y)$ , so besitzt  $Y$  ebenfalls eine Dichte  $f_Y$ , und es gilt*

$$f_Y(y) = f_X(g(y)) |g'(y)|, \quad y \in R_1. \quad (7.24)$$

Beweis: Es sei  $h$  monoton wachsend. Dann gilt

$$F_Y(y) = P(Y \leq y) = P(h(X) \leq y) =$$

$$P(X \leq g(y)) = \int_{(-\infty, g(y)]} f_X(x) dx.$$

Darauf folgt, dass  $F_Y$  differenzierbar ist, und dass gilt

$$f_Y(y) := F'_Y(y) = f_X(g(y)) \cdot g'(y).$$

Ist  $h$  monoton fallend, so haben wir  $f_Y(y) = f_X(g(y)) \cdot (-g'(y))$ . Somit ergibt sich die Aussage.  $\square$

Beispiele:

- 1) Es sei  $h(x) = ax + b$  mit  $a > 0$ ,  $Y = aX + b$ .

Dann ist

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right) \text{ und } f_X(x) = a \cdot f_Y(ax + b).$$

- 2) Ist  $D^2X < \infty$ , so bezeichnet man

$$X^* := \frac{X - EX}{\sqrt{D^2X}}$$

als die zu  $X$  gehörende standardisierte Zufallsgröße. Es gilt

$$EX^* = 0, \quad D^2X^* = 1 \text{ und}$$

$$X = \sqrt{D^2X} X^* + EX.$$

Somit haben wir

$$f_X(x) = \frac{1}{\sqrt{D^2X}} f_{X^*}\left(\frac{x - EX}{\sqrt{D^2X}}\right).$$

- 3) Es sei  $X$  eine  $N(\mu, \sigma^2)$ -verteilte Zufallsgröße. Dann besitzt  $Y = \exp(X)$  eine Dichte der Form

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2y}} \exp\left[-\frac{1}{2\sigma^2}(\ln y - \mu)^2\right], \quad y > 0$$

$$f_Y(y) = 0, \quad y \leq 0.$$

Die Verteilung mit dieser Dichte nennt man *logarithmische Normalverteilung* mit den Parametern  $\mu$  und  $\sigma^2$ . Es gilt

$$EY = e^{\mu + \frac{\sigma^2}{2}}, \quad D^2Y = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$$

## Die eindimensionale Normalverteilung $N(\mu, \sigma^2)$

Es sei  $X \sim N(\mu, \sigma^2)$ . Dann gilt

$$P(a < X \leq b) = P\left(\frac{a - \mu}{\sigma} < X^* \leq \frac{b - \mu}{\sigma}\right) = \\ \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right), \quad a, b \in \mathbb{R}_1, a < b.$$

Insbesondere erhalten wir für alle  $c > 0$ :

$$P(\mu - c\sigma < X \leq \mu + c\sigma) = P(|X^*| < c) = \\ \Phi(c) - \Phi(-c) = 2\Phi(c) - 1$$

Für  $c = 3$  ergibt sich:

$$P(|X - \mu| < 3\sigma) = 0,9974$$

(3 -  $\sigma$ -Regel für die Normalverteilung)

## 7.4 Die Kovarianzmatrix eines zufälligen Vektors

**Definition 7.28** Es sei  $X = (X_1, \dots, X_n)^T$  ein zufälliger Vektor über  $(\Omega, \mathfrak{A}, P)$  mit  $E|X_i| < \infty, i = 1, \dots, n$ .

Dann heißt der Vektor  $\mu$ , definiert durch

$$\mu := (\mu_1, \dots, \mu_n)^T, \quad \mu_i = EX_i, \quad i = 1, \dots, n,$$

der Erwartungswertvektor von  $X$ . Er wird auch mit  $EX$  bezeichnet:

$$EX := (EX_1, \dots, EX_n)^T.$$

Gilt  $EX_i^2 < \infty, i = 1, \dots, n$ , so ist wegen der Cauchy-Schwarz'schen Ungleichung (7.9) auch  $E|X_i X_j| < \infty, i, j = 1, \dots, n$ . Folglich sind alle Kovarianzen  $Kov(X_i, X_j)$  mit  $i, j = 1, \dots, n$  endlich und es gilt:

$$Kov(X_i, X_j) = E(X_i - \mu_i)(X_j - \mu_j) = EX_i X_j - \mu_i \mu_j.$$

Offenbar gilt  $Kov(X_i, X_i) = D^2 X_i = \sigma_i^2$ .

**Definition 7.29** Die Matrix  $\sum_X := (Kov(X_i, X_j))_{i,j=1,\dots,n}$  heißt Kovarianzmatrix des zufälligen Vektors  $X$ .

Mit der Schreibweise

$$E(XX^T) := (EX_iX_j)_{i,j=1,\dots,n}$$

gilt

$$\sum_X = E[(X - \mu)(X - \mu)^T].$$

Auf Grund der Linearität der Erwartungswertbildung haben wir

$$\sum_X = E(XX^T) - \mu\mu^T = (E(X_iX_j) - EX_iEX_j)_{i,j=1,\dots,n}. \quad (7.25)$$

**Aussage 7.30** Die Kovarianzmatrix  $\sum_X$  ist symmetrisch und nichtnegativ definit. Für jeden Vektor  $a = (a_1, \dots, a_n)^T \in R_n$  gilt

$$E(a^T X) = a^T (EX) \text{ und } D^2(a^T X) = E(a^T (X - \mu))^2 = a^T \sum_X a \geq 0. \quad (7.26)$$

Sind die  $X_1, \dots, X_n$  paarweise unkorreliert, so ist  $\sum_X$  eine Diagonalmatrix, und umgekehrt.

Beweis:

$$E(a^T X) = E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i EX_i = a^T EX.$$

Die Symmetrie von  $\sum_X$  folgt aus  $Kov(X_i, X_j) = Kov(X_j, X_i)$ . Für jedes  $a \in R_n$  haben wir auf Grund der Linearität der Erwartungswertbildung

$$E(a^T X) = a^T EX \quad \text{und}$$

$$a^T \sum_X a = a^T E[(X - \mu)(X - \mu)^T]a =$$

$$E[(a^T(X - \mu)) \cdot ((X - \mu)^T a)] = E(a^T(X - \mu))^2 =$$

$$D^2(a^T X) \geq 0.$$

Insbesondere bedeutet dies die nichtnegative Definitheit von  $\sum_X$ . Der letzte Teil der Aussage ist offensichtlich.  $\square$

Die Kovarianzmatrix  $\sum_X$  ist das mehrdimensionale Analogon zur Varianz  $\sigma_X^2$  für reellwertige Zufallsgrößen  $X$ . Im mehrdimensionalen Fall ist die Varianz des zufälligen Vektors richtungsabhängig und i. a. nicht mehr durch eine einzige Zahl zu charakterisieren. Ist  $e = (e_1, \dots, e_n)^T$  ein Vektor der Länge Eins, so ist nach der vorangegangenen Aussage  $e^T \sum_X e = E((e^T(X - \mu))^2)$  die Varianz der Projektion  $e^T X$  von  $X$  auf die durch  $e$  gegebene Richtung.

## Lineare Transformationen

Die folgende Aussage wird in der linearen Algebra bewiesen.

**Aussage 7.31** Die Kovarianz  $\sum_X$  ist singular genau dann, wenn es einen Vektor  $x = (x_1, \dots, x_n)^T \neq 0$  gibt mit  $x^T \sum_X x = 0$ .

Für jedes solche  $x$  gilt also wegen (7.26), dass  $D^2(x^T X) = 0$  gilt.

Ist  $Y$  eine lineare Transformation des  $n$ -dimensionalen zufälligen Vektors  $X$ , d. h. gilt  $Y = AX + b$  für eine  $m \times n$ -Matrix  $A$  und einen  $m$ -dimensionalen Vektor  $b$ , so ist

$$EY = AEX + b \text{ und } \sum_Y = A \sum_X A^T. \quad (7.27)$$

Da  $\sum_X$  symmetrisch ist, gibt es eine orthogonale Matrix  $\mathcal{O}$ , so dass  $\mathcal{O} \sum_X \mathcal{O}^T = D$  eine Diagonalmatrix ist:

$$D = \begin{pmatrix} d_1 & & 0 \\ & d_2 & \\ & & \ddots \\ 0 & & & d_n \end{pmatrix}$$

Die Diagonalelemente  $d_i$  sind die Eigenwerte von  $\sum_X$  und nichtnegativ wegen der nichtnegativen Definitheit von  $\sum_X$ . Der zufällige Vektor  $Y := \mathcal{O}X$  besitzt gemäß (7.27) mit  $\sum_Y = E\mathcal{O}XX^T\mathcal{O}^T$  die Matrix  $D$  als Kovarianzmatrix. Seine Komponenten sind somit unkorreliert.

## Regressionsgerade

Es sei  $(U, V)$  ein zufälliger Vektor reellwertiger Zufallsgrößen  $U$  und  $V$ .

Genau wie im Kapitel 4.5 definiert man die Regressionsgerade für  $V$  auf der Basis von  $U$  durch

$$y = EV + Kor(U, V) \frac{\sigma_2}{\sigma_1} (x - EU)$$

Die Zufallsgröße  $\hat{V}$ , definiert durch

$$\hat{V} := EV + Kor(U, V) \frac{\sigma_2}{\sigma_1} (U - EU)$$

ist die im quadratischen Mittel beste Vorhersage von  $V$  auf der Basis von  $U$ , d. h., es gilt

$$E(V - \hat{V})^2 = \min_{a, b \in \mathbb{R}_1} E(V - aU - b)^2.$$

Diese Regressionsgerade ist für alle Paare reellwertiger Zufallsgrößen  $(U, V)$  definiert, für die  $\sigma_1^2 = D^2U < \infty$  und  $\sigma_2^2 = D^2V < \infty$  gilt.

Für den Vorhersagefehler  $V - \hat{V}$  erhalten wir

$$E(V - \hat{V}) = 0 \text{ und}$$

$$D^2(V - \hat{V}) = D^2V(1 - Kor(U, V))^2.$$

Außerdem haben wir

$$Kov(U, V - \hat{V}) = E((U - EU)[(V - EV) - Kor(U, V) \frac{\sigma_1}{\sigma_2} (U - EU)]) =$$

$$Kov(U, V) - \sigma_1 \sigma_2 Kor(U, V) = 0.$$

## 7.5 Dichten mehrdimensionaler Verteilungen

In diesem Punkt studieren wir zufällige  $n$ -dimensionale Vektoren, die eine Dichte besitzen.

**Definition 7.32** *Ist  $Q$  ein Wahrscheinlichkeitsmaß auf  $(R_n, \mathcal{B}_n)$ , und existiert eine Borelfunktion  $f$  auf  $R_n$ , so dass mit der Bezeichnung  $(-\infty, x] := (-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_n]$ , wobei  $x = (x_1, x_2, \dots, x_n)$  sei, gilt*

$$Q((-\infty, x]) = \int \int_{(-\infty, x]} \dots \int f(x) dx_1, dx_2 \dots dx_n, \quad x \in R_n,$$

dann heißt  $f$  eine Dichte des Maßes  $Q$ . Ist  $Q = P^X$  für einen  $n$ -dimensionalen zufälligen Vektor  $X$ , so nennt man  $f$  auch die Dichte von  $X$ .

Dabei versteht sich das Integral als Integral bezüglich des  $n$ -dimensionalen Lebesguemaßes  $\lambda(dx) = \lambda(dx_1, \dots, dx_n) = dx_1 dx_2 \dots dx_n$ .

Auch hier haben wir für jedes  $B \in \mathcal{B}_n$  die Gleichung

$$Q(B) = \int_{R_n} f(y) \mathbb{1}_B(y) dy =: \int_B f(y) dy$$

Analog zum Fall des  $R_1$  ist eine nichtnegative Borel-messbare Funktion  $f$  auf  $R_n$  Dichte einer Wahrscheinlichkeitsverteilung auf  $(R_n, \mathcal{B}_n)$  genau dann, wenn

$$\int_{R_n} f(x) dx = 1$$

gilt.

In diesem Fall bestimmt  $f$  die Verteilung eindeutig, andererseits ist die Dichte  $f$  einer  $n$ -dimensionalen Wahrscheinlichkeitsverteilung bis auf eine Menge vom  $n$ -dimensionalen Lebesguemaß Null eindeutig bestimmt.

**Aussage 7.33 (Erwartungswertregel)** *Es seien  $X$  ein  $n$ -dimensionaler zufälliger Vektor mit der Dichte  $f$  und  $h$  eine Borel-messbare reellwertige Funktion auf  $(R_n, \mathcal{B}_n)$ . Dann gilt:*

- a)  $h(\cdot)$  ist bezüglich  $P^X$  integrierbar genau dann, wenn  $h(\cdot)f(\cdot)$  bezüglich des  $n$ -dimensionalen Lebesguemaßes integrierbar ist,
- b) in diesem Fall gilt

$$Eh(X) = \int_{R_n} h(x)P^X(dx) = \int_{R_n} h(x)f(x)dx.$$

Wir haben im Fall diskreter Verteilungen die Kovarianz zweier Zufallsgrößen in Kapitel 4 berechnet. Hier wollen wir die entsprechenden Formeln für den Fall angeben, dass der zufällige Vektor  $X$  eine Dichte  $f$  besitzt.

In diesem Fall gilt nach der Erwartungswertregel

$$\mu_i = EX_i = \int_{R_n} x_i f(x_1, \dots, x_n) dx_1, \dots, x_n, \quad i = 1, \dots, n,$$

$$Kov(X_i, X_j) = E(X_i - \mu_i)(X_j - \mu_j) =$$

$$\int_{R_n} (x_i - \mu_i)(x_j - \mu_j) f(x_1, \dots, x_n) dx_1 \dots dx_n =$$

$$\int_{R_n} x_i x_j f(x_1, \dots, x_n) dx_1 \dots dx_n - \mu_i \mu_j.$$

Wie man Integrale über Funktionen im  $R_n$  ausrechnet, werden wir im folgenden Kapitel 8 kennen lernen.

Wir beschränken uns im Weiteren auf den Fall  $n = 2$ .

Es sei  $X = (Y, Z)^T$  ein zufälliger Vektor mit Werten in  $(R_2, \mathcal{B}_2)$  und der Dichte  $f$ .

### Aussage 7.34

- a)  $Y$  und  $Z$  haben Dichten  $f_Y$  bzw.  $f_Z$ , die sich mittels  $f$  wie folgt berechnen lassen:

$$f_Y(y) = \int_{R_1} f(y, z) dz, \quad y \in R_1$$

$$f_Z(z) = \int_{R_1} f(y, z) dy, \quad z \in R_1$$

b)  $Y$  und  $Z$  sind genau dann voneinander unabhängig, falls

$$f(y, z) = f_Y(y)f_Z(z) \quad (y, z) \in R_2, \quad \lambda_2 - \text{fast überall}$$

c) für jedes  $y$  mit  $f_Y(y) > 0$  ist durch

$$f_{Y=y}(z) := \frac{f(y, z)}{f_Y(y)}, \quad z \in R_1$$

eine Dichte definiert. Sie heißt "bedingte Dichte von  $Z$  unter der Bedingung  $Y = y$ ."

Man nennt  $f_Y$  und  $f_Z$  die Randverteilungsdichten von  $f$ .

Beweis:

a)  $P(Y \leq y) = P(Y \leq y, Z \in R_1) =$

$$\iint_{(-\infty, y] \times R_1} f(s, t) ds dt = \int_{(-\infty, y]} \left( \int_{R_1} f(s, t) dt \right) ds.$$

Folglich gilt die erste Formel von a), analog folgt die zweite (es wurde der Satz von Fubini benutzt.)

b) Wenn  $f = f_Y f_Z$ , so ist

$$P(Y \in B, Z \in C) = \iint_{B \times C} f(y, z) dy dz =$$

$$\int_{R_1} \mathbb{1}_C f_Z(z) \left( \int_{R_1} \mathbb{1}_B(y) f_Y(y) dy \right) dz =$$

$$P(Y \in B)P(Z \in C), \quad B, C \in \mathcal{B}_1.$$

Also sind  $Y$  und  $Z$  unabhängig.

Umgekehrt, sind  $Y$  und  $Z$  unabhängig, so gilt

$$F(y, z) = P(Y \leq y, Z \leq z) = F_Y(y)F_Z(z) =$$

$$\int_{(-\infty, y]} f_Y(y) dy \int_{(-\infty, z]} f_Z(z) dz = \iint_{(-\infty, y] \times (-\infty, z]} f_Y(y) f_Z(z) dy dz$$

(Fubini; Tonelli, Hobson)). Wegen der Eindeutigkeit der Dichte besitzt  $(Y, Z)^T$  eine Dichte  $f$ , und es gilt

$$f(y, z) = f_Y(y) \cdot f_Z(z) \quad , \quad \lambda - \text{f.ü.}$$

c) Es gilt  $f_{Y=y}(z) \geq 0$  und  $\int_{R_1} f_{Y=y}(z) dz = 1$ .

Bemerkung: Interpretation von c):

Es sei  $f(y, z)$  stetig und streng positiv in  $(y_0, z_0)$ .

Dann ist  $f(x, y) > 0$  in Umgebung  $U$  von  $(y_0, z_0)$  (z.B.  $U = (y_0 - \Delta, y_0 + \Delta) \times (z_0 - \Delta, z_0 + \Delta)$  für genügend kleines  $\Delta > 0$ ).

Wir erhalten für jedes  $z \in R_1$

$$P(Z \leq z | Y \in (y_0 - \Delta, y_0 + \Delta)) = \frac{P(Y \in (y_0 - \Delta, y_0 + \Delta), Z \leq z)}{P(Y \in (y_0 - \Delta, y_0 + \Delta))} =$$

$$\begin{aligned}
& \int_{(-\infty, z]} \left( \int_{(y_0 - \Delta, y_0 + \Delta)} f(s, t) ds \right) dt \Big/ \int_{R_1} \left( \int_{(y_0 - \Delta, y_0 + \Delta)} f(s, t) ds \right) dt = \\
& \sim \int_{(-\infty, z]} (2\Delta f(y_0, t)) dt \Big/ \int_{R_1} 2\Delta f(y_0, t) dt \\
& = \int_{(-\infty, z]} f_{Y=y_0}(s) ds
\end{aligned}$$

und sehen darin eine Interpretation von  $f_{Y=y_0}(z)$  als Dichte von  $Z$  unter der Bedingung  $Y = y_0$ .

**Beispiel 7.35**  $X = (X_1, X_2)^T$  besitze eine 2-dimensionale Normalverteilung mit den Parametern  $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$ . Dann gilt

$$f_X(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)} \left[ \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 \right]\right)$$

$$f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(x_i - \mu_i)^2\right)$$

$$EX_i = \int_{R_2} x_i f_X(x_1, x_2) dx_1 dx_2 = \int_{R_1} x_i f_{X_i}(x_i) dx_i = \mu_i$$

$$D^2 X_i = \int_{R_2} (x_i - \mu_i)^2 f_X(x_1, x_2) dx_1 dx_2 = \int_{R_1} (x_i - \mu_i)^2 f_{X_i}(x_i) dx_i = \sigma_i^2$$

$$Kov(X_1, X_2) = \int_{R_2} (x_1 - \mu_1)(x_2 - \mu_2) f_X(x_1, x_2) dx_1 dx_2 =$$

$$\int_{R_1} (x_2 - \mu_2) \left( \int_{R_1} (x_1 - \mu_1) \cdot f_X(x_1, x_2) dx_1 \right) dx_2 = \rho\sigma_1\sigma_2.$$

Damit ist die Bedeutung der Parameter der 2-dimensionalen Normalverteilung geklärt. Folglich haben wir für die Kovarianzmatrix des Vektors  $X$

$$\Sigma_X = \begin{pmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

und  $\varrho$  ist gleich dem Korrelationskoeffizienten  $Kor(X_1, X_2)$ .

Man prüft leicht nach, dass sich die Dichte  $f_X$  in diesem Beispiel folgendermaßen schreiben läßt ( $\mu = (\mu_1, \mu_2)^T$ ,  $x = (x_1, x_2)^T$ ):

$$f_X(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp \left[ -\frac{1}{2}(x-\mu)^T \Sigma_X^{-1} (x-\mu) \right]$$

Für die bedingte Dichte  $f_{X_1=x_1}(x_2)$  ergibt sich

$$f_{X_1=x_1}(x_2) = \frac{1}{\sqrt{2\pi\sigma_2^*}} \exp \left[ -\frac{1}{2\sigma_2^*} (x_2 - \mu_2^*)^2 \right]$$

mit

$$\mu_2^* = \mu_2 + \varrho \frac{\sigma_2}{\sigma_1} (x_1 - \mu_1) \text{ und}$$

$$\sigma_2^{*2} = \sigma_2^2 (1 - \varrho^2)$$

Beachte, dass  $\sigma_2^{*2}$  nicht von  $x_1$  abhängt.

Die Komponenten  $X_1$  und  $X_2$  sind genau dann unabhängig, falls sie unkorreliert sind, also  $\varrho = 0$  gilt. In der Tat, genau in diesem Fall gilt

$$f_X(x) = f_{X_1}(x_1)f_{X_2}(x_2), \quad x = (x_1, x_2)^T.$$

Die Transformationsformel für  $n$ -dimensionale Dichten (Aussage (3.59) und Beispiel (3.60) bleiben in diesem allgemeinen Fall gültig.



# Kapitel 8

## Produktmaße und Summen unabhängiger Zufallsgrößen

Die Wahrscheinlichkeitsverteilungen voneinander unabhängiger Zufallsgrößen sind Produktmaße. In diesem Kapitel geht es zunächst um den Zusammenhang zwischen Integralen über Funktionen bez. Produktmaßen und sogenannten iterierten Integralen über diese Funktionen bez. der Einzelmaße. Der diesbezügliche Satz von Fubini mit samt seiner Folgerung, die nach Tonelli und Hobson benannt ist, bildet ein in der Wahrscheinlichkeitstheorie oft benutztes Werkzeug.

Im zweiten Abschnitt wird dann die Verteilung der Summen unabhängiger Zufallsgrößen untersucht. Sie ergibt sich als sogenannte Faltung der Verteilungen der einzelnen Zufallsgrößen.

### 8.1 Der Satz von Fubini

Es seien  $X$  und  $Y$  zwei Zufallsgrößen über  $(\Omega, \mathfrak{A}, P)$  mit Werten in  $(E, \mathfrak{E})$  bzw.  $(F, \mathfrak{F})$ . Der zufällige Vektor  $(X, Y)^T$  ist eine Zufallsgröße mit Werten in  $(E \times F, \mathfrak{E} \otimes \mathfrak{F})$ , wobei  $\mathfrak{E} \otimes \mathfrak{F}$  die Produkt- $\sigma$ -Algebra von  $\mathfrak{E}$  und  $\mathfrak{F}$  in  $E \times F$  ist, d. h. die kleinste  $\sigma$ -Algebra von Teilmengen von  $E \times F$ , die alle "Rechtecke"  $B \times C$  mit  $B \in \mathfrak{E}$  und  $C \in \mathfrak{F}$  umfasst.

Sind  $X$  und  $Y$  voneinander unabhängig, so gilt für die Verteilung  $P^{(X,Y)}$  von  $(X, Y)^T$  die Beziehung

$$P^{(X,Y)}(B \times C) = P(X \in B, Y \in C) = P(X \in B)P(Y \in C) = P^X(B)P^Y(C), \quad B \in \mathfrak{E}, C \in \mathfrak{F}. \quad (8.1)$$

Die gemeinsame Verteilung  $P^{(X,Y)}$  von  $(X, Y)^T$  ist also das *Produktmaß* der Randverteilungen  $P^X$  und  $P^Y$ . Ist  $h$  eine Borel-messbare Funktion von  $(E \times F, \mathfrak{E} \otimes \mathfrak{F})$  in  $(R_1, \mathcal{B}_1)$ , so haben wir gemäß der Substitutionsformel (Aussage 7.14)

$$Eh(X, Y) = \int_{E \times F} h(x, y) P^{(X,Y)}(dx, dy). \quad (8.2)$$

Der folgende Satz von Fubini liefert Bedingungen, unter denen man dieses Integral im Fall der Unabhängigkeit von  $X$  und  $Y$  auf Einzelintegrale bez.  $P^X$  bzw.  $P^Y$  zurückführen kann, die sich z. B. im Falle der Existenz von Dichten wiederum einfacher berechnen lassen. Für den Beweis dieses Satzes und den Aussagen dieses Abschnittes siehe z.B. Elstrodt (1999), Bauer (1990) bzw. die Vorlesung Maßtheorie.

Anstelle von  $P^X$  und  $P^Y$  verwenden wir im Folgenden irgend zwei Wahrscheinlichkeitsmaße  $Q_1$  und  $Q_2$  auf  $(E, \mathfrak{E})$  bzw.  $(F, \mathfrak{F})$ .

Wir beginnen mit einer Aussage über Existenz und Eindeutigkeit des Produktmaßes  $Q_1 \otimes Q_2$ .

**Aussage 8.1** *Es seien  $Q_1$  und  $Q_2$  zwei Wahrscheinlichkeitsmaße auf  $(E, \mathfrak{E})$  bzw.  $(F, \mathfrak{F})$ . Dann besitzt die durch*

$$R(A \times B) := Q_1(A)Q_2(B) \quad (8.3)$$

*auf der Menge  $\mathcal{R}$  aller Rechtecke  $A \times B$  mit  $A \in \mathfrak{E}, B \in \mathfrak{F}$  definierte Mengenfunktion  $R$  eine eindeutige Fortsetzung zu einem Wahrscheinlichkeitsmaß auf  $\mathfrak{E} \otimes \mathfrak{F}$ . Diese Fortsetzung wird als Produktmaß  $Q_1 \otimes Q_2$  von  $Q_1$  und  $Q_2$  bezeichnet.*

Bemerkung: Die Aussage bleibt richtig, falls  $Q_1$  und  $Q_2$  keine Wahrscheinlichkeitsmaße, sondern sogenannte  $\sigma$ -finite Maße sind. Das Produktmaß  $Q_1 \otimes Q_2$  ist dann ebenfalls ein  $\sigma$ -finites Maß. Das gilt insbesondere für die Lebesguemaße in  $R_1$  bzw.  $R_2$  und allgemeiner, in  $R_n$ .

**Satz 8.2 (Satz von Fubini)** *Es sei  $h$  eine reellwertige,  $(\mathfrak{E} \otimes \mathfrak{F})$ - $\mathcal{B}_1$ -messbare Funktion, die nichtnegativ ist oder für die*

$$\int_{E \times F} |h(x, y)| Q_1 \otimes Q_2(dx, dy) < \infty \quad (8.4)$$

*richtig ist.*

*Dann gelten folgende zwei Aussagen:*

1) *Die Funktion  $x \rightarrow \int_F h(x, y) Q_2(dy)$  ist  $\mathfrak{E}$ - $\mathcal{B}_1$ -messbar. Sie ist nichtnegativ, falls  $h$  nichtnegativ ist, und sie ist  $Q_1$ -integrierbar, falls (8.4) gilt.*

*Die Funktion  $y \rightarrow \int_E h(x, y) Q_1(dx)$  ist  $\mathfrak{F}$ - $\mathcal{B}_1$ -messbar. Sie ist nichtnegativ, falls  $h$  nichtnegativ ist, und sie ist  $Q_2$ -integrierbar, falls (8.4) gilt.*

2) *Ist  $h$  nichtnegativ oder gilt (8.4), so haben wir*

$$\begin{aligned} \int_{E \times F} h(x, y) Q_1 \otimes Q_2(dx, dy) &= \int_E \left( \int_F h(x, y) Q_2(dy) \right) Q_1(dx) \\ &= \int_F \left( \int_E h(x, y) Q_1(dx) \right) Q_2(dy) \end{aligned} \quad (8.5)$$

Bemerkung: Der Satz von Fubini gilt auch für Maße  $Q_1, Q_2$ , die nicht notwendig Wahrscheinlichkeitsmaße sind. Insbesondere für Lebesguemaße.

Der Satz von Fubini besagt also, dass aus der Integrierbarkeit von  $h$  bezüglich des Produktmaßes  $Q_1 \otimes Q_2$  folgt, dass die iterierten Integrale in (8.5) existieren und gleich dem Integral bez. dem Produktmaß sind.

Im allgemeinen folgt aus der Endlichkeit der iterierten Integrale, selbst wenn sie gleich sind, noch nicht die Eigenschaft (8.5). (Vgl. Elstrodt (1996), Kap. V §2).

Es gilt aber die

**Folgerung 8.3 (Tonelli, Hobson)** *Wenn gilt*

$$\int_E \left( \int_F |h(x, y)| Q_2(dy) \right) Q_1(dx) < \infty \quad \text{oder} \quad (8.6)$$

$$\int_F \left( \int_E |h(x, y)| Q_1(dx) \right) Q_2(dy) < \infty, \quad (8.7)$$

dann sind die Voraussetzungen des Satzes von Fubini erfüllt.

Beweis: Für jede Indikatorfunktion  $h = \mathbb{1}_C$  mit  $C \in \mathfrak{E} \otimes \mathfrak{F}$  gilt der Satz von Fubini, d. h.  $x \rightarrow Q_2(C(x))$  ist  $\mathfrak{E}$ -messbar und es besteht die Beziehung

$$\int_{E \times F} h dQ_1 \otimes Q_2 = \int_E Q_2(C(x)) Q_1(dx) < \infty$$

wobei  $C(x)$  die sogenannte "Schnittmenge" von  $C$  bei  $x$  ist:

$$C(x) := \{y \in F \mid (x, y) \in C\}, \quad x \in E.$$

Damit ist auch für jede nichtnegative Elementarfunktion  $h$

$$\int_{E \times F} h dQ_1 \otimes Q_2 = \int_E \left( \int_F h dQ_2 \right) dQ_1 < \infty.$$

Ist nun  $h$  nichtnegativ mit der Eigenschaft (8.6) und  $(h_n)$  eine  $h$  approximierende Folge von nichtnegativen Elementarfunktionen, so ergibt sich aus dem Satz über die monotone Konvergenz, dass

$$\begin{aligned} \int_{E \times F} h dQ_1 \otimes Q_2 &= \lim_{n \rightarrow \infty} \int_{E \times F} h_n dQ_1 \otimes Q_2 = \\ &= \lim_{n \rightarrow \infty} \int_E \left( \int_F h_n dQ_2 \right) dQ_1 \end{aligned}$$

richtig ist.

Wegen

$$\int_F h_n dQ_2 \uparrow \int_F h dQ_2$$

(Satz über monotone Konvergenz) erhalten wir wegen (8.6) die Eigenschaft (8.4).

Im allgemeinen Fall nutzen wir die Zerlegung  $h = h^+ - h^-$ .

Der Fall (7) wird analog behandelt. □

## 8.2 Faltungsformeln

Es seien  $X$  und  $Y$  zwei voneinander unabhängige reellwertige Zufallsgrößen über  $(\Omega, \mathfrak{A}, P)$  und  $Z := X + Y$ .

Für die Wahrscheinlichkeitsverteilung  $P^Z$  der reellwertigen Zufallsgröße  $Z$  gilt wegen der Substitutionsformel (7.13), angewandt auf  $h(X, Y) = \mathbb{1}_B(X + Y)$ ,

$$\begin{aligned} P^Z(B) &= P(X + Y \in B) = E \mathbb{1}_B(X + Y) = \\ &= \int_{\Omega} \mathbb{1}_B(X + Y) dP = \int_{\mathbb{R}_2} \mathbb{1}_B(x + y) P^X \otimes P^Y(dx, dy) \end{aligned}$$

Auf Grund des Satzes von Fubini ist dieser Wert gleich

$$\int_{R_1} P^Y(\{y : x + y \in B\}) P^X(dx).$$

Für  $B = (-\infty, z]$  ergibt sich

$$\begin{aligned} F_Z(z) &= P^Z((-\infty, z]) = P(Z \leq z) = \\ &= \int_{R_1} P^Y((-\infty, z - x]) P^X(dx) = \\ &= \int_{R_1} F_Y(z - x) P^X(dx). \end{aligned} \tag{8.8}$$

Bemerkung: Da die Verteilungsfunktion  $F_X$  die Wahrscheinlichkeitsverteilung  $P^X$  eindeutig bestimmt, schreibt man häufig statt

$$\int_{R_1} f(x) P^X(dx) \text{ auch } \int_{R_1} f(x) F_X(dx).$$

Wir haben dann also anstelle (8.8) die Gleichung

$$F_Z(z) = \int_{R_1} F_Y(z - x) F_X(dx) \tag{8.8'}$$

**Definition 8.4** Die durch die rechte Seite von (8.8') aus  $F_X$  und  $F_Y$  gebildete Verteilungsfunktion  $F_Z$  bezeichnet man als Faltung der beiden Verteilungsfunktionen  $F_X$  und  $F_Y$  und schreibt  $F_Z = F_X * F_Y$ .

Offensichtlich gilt  $F_X * F_Y = F_Y * F_X$ .

### Spezialfälle 8.5

- a) Sind  $X$  und  $Y$  unabhängig und diskret verteilt mit Werten aus den ganzen Zahlen, gilt also

$$P(X = k) = p_k, \quad P(Y = k) = q_k, \quad k \in \Gamma := \{0, \pm 1, \pm 2, \dots\}$$

mit  $p_k, q_k \geq 0$  und  $\sum p_k = \sum q_k = 1$ , so erhalten wir

$$P(Z = k) = \sum_{l \in \Gamma} q_{k-l} p_l = \sum_{l \in \Gamma} p_{k-l} q_l, \quad k \in \Gamma. \quad (8.9)$$

Beweis: Man verwende

$$P(Z = k) = F_Z(k) - F_Z(k - \frac{1}{2}) \text{ und } \int_{R_1} h(x) F_X(dx) = \sum_k h(k) p_k$$

bzw.  $\int_{R_1} h(x) F_Y(dx) = \sum_k h(k) q_k$  für alle nichtnegativen Funktionen  $h$  auf der Menge der ganzen Zahlen.  $\square$

- b) Sind  $X$  und  $Y$  unabhängig und haben sie Dichten  $f_X$  bzw.  $f_Y$ , so hat auch  $Z = X + Y$  eine Dichte  $f_Z$ , und es gilt

$$f_Z(z) = \int_{R_1} f_Y(z-x) f_X(x) dx = \int_{R_1} f_X(z-y) f_Y(y) dy. \quad (8.10)$$

Beweis: Mittels der Substitutionsformel (7.13) folgt aus (8.8) die Beziehung

$$F_Z(z) = \int_{R_1} \left( \int_{(-\infty, z-x]} f_Y(y) dy \right) f_X(x) dx =$$

$$\int_{R_1} \left( \int_{(-\infty, z]} f_Y(y-x) dy \right) f_X(x) dx$$

und der Satz von Fubini mit seiner Folgerung von Tonelli-Hobson ergibt damit

$$F_Z(z) = \int_{(-\infty, z]} \left( \int_R f_Y(y-x) f_X(x) dx \right) dy.$$

Daraus folgt nach Definition der Dichten die Behauptung.  $\square$

**Bemerkung 8.6** Bereits wenn nur eine der beiden unabhängigen Zufallsgrößen  $X$  und  $Y$  eine Dichte besitzt, hat auch  $Z = X + Y$  eine Dichte. (Übung)

**Beispiele 8.7** Es seien  $X$  und  $Y$  unabhängige Zufallsgrößen.

1. Sind  $X$  und  $Y$  binomialverteilt mit den Parametern  $(n, p)$  bzw.  $(m, p)$ , so ist  $X + Y$  binomialverteilt mit den Parametern  $(n + m, p)$ .
2. Sind  $X$  und  $Y$  Poissonverteilt mit den Parametern  $\lambda$  bzw.  $\mu$ , so ist  $X + Y$  Poissonverteilt mit dem Parameter  $\lambda + \mu$ .
3. Sind  $X$  und  $Y$  negativ binomialverteilt mit den Parametern  $(p, v_X)$  bzw.  $(p, v_Y)$ , so ist  $X + Y$  negativ binomialverteilt mit den Parametern  $(p, v_X + v_Y)$ .
4. Sind  $X$  und  $Y$  normalverteilt mit den Parametern  $(\mu_X, \sigma_X^2)$  bzw.  $(\mu_Y, \sigma_Y^2)$ , so ist  $X + Y$  normalverteilt mit den Parametern  $(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$ .
5. Sind  $X$  und  $Y$  Gammaverteilt mit den Parametern  $(\alpha_X, \lambda)$  bzw.  $(\alpha_Y, \lambda)$ , so ist  $X + Y$  Gammaverteilt mit den Parametern  $(\alpha_X + \alpha_Y, \lambda)$ .

Der Beweis von 1. - 3. ergibt sich sofort aus der Formel

$$g_{X+Y}(s) = E s^{X+Y} =$$

$$E s^X \cdot E s^Y = g_X(s) \cdot g_Y(s), |s| < 1$$

für die erzeugenden Funktionen.

Für den Beweis von 4. und 5. verwendet man die sogenannten charakteristischen Funktionen, die wir später definieren werden.

Die Faltungsformeln (8.9) und (8.10) führen auch zum Ziel, sind aber häufig mit längeren Rechnungen verbunden.

**Definition 8.8** Eine Familie  $(F_\vartheta, \vartheta \in \Theta \subseteq R_k)$  von Verteilungsfunktionen auf  $R_1$  heißt *faltungsstabil*, falls für alle  $\vartheta, \eta \in \Theta$  ein  $\xi \in \Theta$  existiert mit  $F_\vartheta * F_\eta = F_\xi$ .

Die Familie aller eindimensionalen Normalverteilungen ist zum Beispiel faltungsstabil.

Wir schließen dieses Kapitel mit einigen Eigenschaften von Erwartungswerten und Varianzen unabhängiger Zufallsgrößen.

**Aussage 8.9** Sind  $X$  und  $Y$  zwei voneinander unabhängige reellwertige Zufallsgrößen über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  und gilt  $E|X| < \infty$ ,  $E|Y| < \infty$ , so folgt

$$E|XY| < \infty \text{ und } E(XY) = (EX)(EY). \quad (8.11)$$

Beweis: Wegen der Unabhängigkeit ist die gemeinsame Verteilung von  $X$  und  $Y$  gleich dem Produktmaß  $P^X \otimes P^Y$ . Der Satz von Fubini (Folgerung 8.3) impliziert

$$E(XY) = \int_{R_2} xy P^X \otimes P^Y(dx, dy) =$$

$$\int_{R_1} \left( \int_{R_1} xy P^X(dx) \right) P^Y(dy) = \int_{R_1} y \left( \int_{R_1} x P^X(dx) \right) P^Y(dy) = (EX)(EY).$$

**Folgerung 8.10** Sind  $X_1, X_2, \dots, X_n$  voneinander unabhängige reellwertige Zufallsgrößen über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  und gilt  $E|X_k| < \infty$ ,  $k = 1, \dots, n$ , so folgt

$$E|X_1 \cdot X_2 \cdot \dots \cdot X_n| < \infty \text{ und}$$

$$E\left(\prod_{k=1}^n X_k\right) = \prod_{k=1}^n EX_k.$$

Die Gleichung (8.11) hat zur Konsequenz, dass für zwei unabhängige Zufallsgrößen  $X$  und  $Y$ , deren Erwartungswerte endlich sind, die Kovarianz auch endlich und gleich Null ist.

$$Kov(X, Y) = E(XY) - EXEY = 0.$$

Unabhängige Zufallsgrößen mit endlichem Erwartungswert sind also unkorreliert.

Daraus ergibt sich die

**Aussage 8.11** Sind  $X_1, X_2, \dots, X_n$  voneinander unabhängige Zufallsgrößen mit  $D^2 X_k < \infty$ ,  $k = 1, \dots, n$ , so gilt

$$D^2\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n D^2 X_k \tag{8.12}$$

Beweis:

$$D^2\left(\sum_{k=1}^n X_k\right) = E\left(\sum_{k=1}^n (X_k - EX_k)\right)^2 =$$

$$\sum_{k,\ell=1}^n E(X_k - EX_k)(X_\ell - EX_\ell) =$$

$$\sum_{k=1}^n D^2 X_k + 2 \sum_{k<\ell} Kov(X_k, X_\ell) = \sum_{k=1}^n D^2 X_k.$$

**Folgerung 8.12** Unter den gleichen Voraussetzungen wie in der eben bewiesenen Aussage gilt für die Varianz des arithmetischen Mittels:

$$D^2\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n^2} \sum_{k=1}^n D^2 X_k \leq \frac{1}{n} \max_{k=1,\dots,n} D^2 X_k.$$

Beispiel: Wirft man einen regelmäßigen Spielwürfel unabhängig voneinander 100mal und bildet das arithmetische Mittel  $M_{100}$  der auftretenden Augenzahlen, so gilt

$$EM_{100} = 3,5 \quad \text{und} \quad D^2 M_{100} = \frac{2,9167}{100} = 0,0292.$$

Das arithmetische Mittel wird bei häufiger Wiederholung von 100 Würfeln also meist in der Nähe von 3,5 liegen. Die Tschebyschev'sche Ungleichung liefert nämlich

$$P(|M_{100} - 3,5| \geq 0,5) \leq \frac{0,0292}{0,25} = 0,12.$$

Das arithmetische Mittel  $M_{100}$  ist also "weniger zufällig" als jeder einzelne Wurf des Würfels.



# Kapitel 9

## Charakteristische Funktionen

Jeder Wahrscheinlichkeitsverteilung auf  $(R_1, \mathfrak{B}_1)$  (allgemeiner:  $(R_n, \mathfrak{B}_n)$ ) ist eine komplexwertige Funktion, ihre charakteristische Funktion, zugeordnet, durch die sie wiederum auch eindeutig bestimmt ist. Alle Eigenschaften der Verteilung spiegeln sich in Eigenschaften ihrer charakteristischen Funktion wider. In vielen Fällen, z.B. bei der Herleitung von Grenzwertsätzen, bilden charakteristische Funktionen ein leistungsfähiges Werkzeug.

Es sei  $X$  eine reellwertige Zufallsgröße über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  mit der Verteilung  $P^X$  und der Verteilungsfunktion  $F_X$ , die wir auch kurz mit  $F$  bezeichnen:

$$F(x) = F_X(x) = P^X((-\infty, x]) = P(X \leq x), \quad x \in R_1.$$

**Definition 9.1** Als charakteristische Funktion  $\varphi$  der Zufallsgröße  $X$  (bzw. der Verteilungsfunktion  $F$ ) bezeichnet man die Funktion

$$\varphi(u) = E \cos(uX) + iE \sin(uX), \quad u \in R_1, \quad i \text{ imaginäre Einheit.} \quad (9.1)$$

Man schreibt auch

$$\begin{aligned} \varphi(u) &= E e^{iuX}, \text{ oder} \\ \varphi(u) &= \int_{R_1} e^{iux} P^X(dx) =: \int_{R_1} e^{iux} F(dx), \quad u \in R_1. \end{aligned}$$

Statt  $\varphi$  verwenden wir die Bezeichnung  $\varphi_X$  oder  $\varphi_F$ , wenn wir die Zufallsgröße  $X$  oder die Verteilungsfunktion  $F$  hervorheben sollen.

Die charakteristische Funktion  $\varphi_X$  ist für jede reellwertige Zufallsgröße  $X$  definiert und gleich der Fouriertransformierten des Maßes  $P^X$  (bis auf eventuell einen konstanten Faktor).

Bevor wir Eigenschaften charakteristischer Funktionen untersuchen, zeigen wir die Gültigkeit der folgenden Ungleichung (9.2). Definiert man für komplexwertige Funktionen  $f$  auf  $R_1$  den Erwartungswert  $Ef(X)$  durch  $Ef(X) := E(\operatorname{Re} f(X)) + i E(\operatorname{Im} f(X))$  (Endlichkeit beider Erwartungswerte der rechten Seite sei vorausgesetzt), so gilt

$$|Ef(X)| \leq E|f(X)|. \quad (9.2)$$

Beweis: Ist  $f$  eine Elementarfunktion

$$f(x) = \sum_1^n \alpha_i \mathbb{1}_{A_i}(x) \quad , \quad x \in R,$$

$\alpha_i$  komplex, so gilt

$|Ef(X)| = \left| \sum_1^n \alpha_i P(A_i) \right| \leq \sum_1^n P(A_i) |\alpha_i| = E|f(X)|$ . Für allgemeine Borel-messbare  $f$  mit  $E|f(X)| < \infty$  folgt (9.2) wie üblich, mit der Approximationsmethode.

Eine sofortige Konsequenz aus der Definition der charakteristischen Funktion ist

$$\varphi_{aX+b}(u) = e^{iub} \varphi_X(au), \quad u \in R_1, \quad (9.3)$$

für alle  $a, b \in R_1$ . Insbesondere, falls  $0 < \sigma^2 = D^2X < \infty$  gilt, haben wir

$$\varphi_{X^*}(u) = e^{-iu\frac{\mu}{\sigma}} \varphi_X\left(\frac{u}{\sigma}\right), \quad u \in R_1. \quad (9.3')$$

Dabei bezeichnet  $X^*$  die *standardisierte Zufallsgröße*  $\frac{X-\mu}{\sigma_X}$  mit  $\mu = EX$ .

**Beispiele 9.2**

a)  $X \sim B(n, p)$  (Binomialverteilung)

$$\varphi_X(u) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \cdot e^{iku} = (pe^{iu} + q)^n, \quad u \in \mathbb{R}_1,$$

b)  $X \sim P(\lambda)$  (Poissonverteilung)

$$\varphi_X(u) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \cdot e^{iuk} = e^{\lambda(e^{iu}-1)}, \quad u \in \mathbb{R}_1,$$

c)  $X \sim NB(v, p)$  (Negative Binomialverteilung)

$$\varphi_X(u) = \left( \frac{p}{1 - qe^{iu}} \right)^v \quad \text{mit } q = 1 - p, \quad u \in \mathbb{R}_1,$$

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

$$\varphi_X(u) = \exp\left(-\frac{\sigma^2 u^2}{2} + i\mu u\right), \quad u \in \mathbb{R}_1,$$

e)  $X \sim \Gamma(\alpha, \lambda)$  (Gammaverteilung)

$$\varphi_X(u) = \left( \frac{\lambda}{\lambda - iu} \right)^\alpha, \quad u \in \mathbb{R}_1,$$

f)  $X \sim$  Cauchyverteilung,  $a > 0$ , Dichte  $f(x) = \frac{1}{\pi} \frac{a}{a^2 + x^2}$ ,  $x \in \mathbb{R}_1$ ,

$$\varphi_X(u) = \exp(-a|u|), \quad u \in \mathbb{R}_1,$$

g) Ausgeartete Verteilung:  $X = a \in \mathbb{R}_1$

$$\varphi_X(u) = e^{iua}, \quad u \in \mathbb{R}_1.$$

**Aussage 9.3** *Es sei  $\varphi$  die charakteristische Funktion einer Zufallsgröße  $X$ . Dann gelten folgende Eigenschaften:*

a)  $|\varphi(u)| \leq \varphi(0) = 1, \quad u \in \mathbb{R}_1,$

- b)  $\varphi(\cdot)$  ist gleichmäßig stetig auf  $R_1$ .
- c)  $\overline{\varphi(u)} = \varphi(-u)$ ,  $u \in R_1$ .
- d)  $\varphi(\cdot)$  ist genau dann eine reellwertige Funktion, falls die Verteilung  $P^X$  symmetrisch ist, d.h., falls  $P^X(-B) = P^X(B)$  für alle  $B \in \mathfrak{B}_1$ , wobei  $-B := \{-x : x \in B\}$  gesetzt wird.
- e) Falls für ein  $n \geq 1$  das  $n$ -te Moment von  $X$  endlich ist, d.h., falls  $E|X|^n < \infty$  gilt, so ist  $\varphi(\cdot)$   $n$ -mal stetig differenzierbar, und es gilt für alle  $k$  mit  $1 \leq k \leq n$

$$\varphi^{(k)}(u) = \frac{d^k}{du^k} \varphi(u) = \int_R (ix)^k e^{iux} F(dx), \text{ insbesondere}$$

$$EX^k = \frac{\varphi^{(k)}(0)}{i^k}.$$

Weiterhin gilt

$$\varphi(u) = \sum_{k=0}^n \frac{(iu)^k}{k!} EX^k + \frac{(iu)^n}{n!} \varepsilon_n(u), \quad u \in R_1, \quad (9.4)$$

mit einer Funktion  $\varepsilon_n(u)$ ,  $u \in R_1$ , für die

$$|\varepsilon_n(u)| \leq 3E|X|^n \text{ und } \lim_{u \rightarrow 0} \varepsilon_n(u) = 0$$

erfüllt ist.

- f) Wenn für ein  $m \geq 1$  die Ableitung  $\varphi^{(2m)}$  bei 0 existiert und endlich ist, so folgt  $E(X^{2m}) < \infty$ .
- g) Wenn  $E|X|^n < \infty$  für alle  $n \geq 1$  und  $\frac{1}{R} := \overline{\lim}_{n \rightarrow \infty} \left[ \frac{1}{n} (E|X|^n)^{\frac{1}{n}} \right] < \infty$  gelten, so haben wir für alle  $u$  mit  $|u| < R$  die Gleichung

$$\varphi(u) = \sum_{n=0}^{\infty} \frac{(iu)^n}{n!} EX^n.$$

Zum Beweis:

a)  $|\varphi(u)| \stackrel{\leq}{=} E|e^{iuX}| = 1 = \varphi(0)$  wegen (9.2).

b)  $|\varphi(u+h) - \varphi(u)| \leq \int_{R_1} |e^{ixu}| |e^{ixh} - 1| F(dx) =$   
 $= E|e^{ixh} - 1| \xrightarrow{h \rightarrow 0} 0, \quad u \in R_1$

auf Grund des Satzes von der majorisierten Konvergenz.

c)  $Ee^{-iuX} = \overline{Ee^{iuX}}$ .

d) Ist  $P^X$  symmetrisch, so gilt für jede beschränkte, schiefsymmetrische Borel-messbare Funktion  $g$  (d.h.  $g(-s) = -g(s)$ ,  $s \in R_1$ ) die Beziehung  $Eg(X) = 0$ . (Sie ist richtig nach Definition, falls  $g$  auf  $(0, \infty)$  gleich einer Indikatorfunktion ist. Den allgemeinen Fall beweist man wie üblich mittels Approximation.) Also ist

$$\int_{R_1} \sin(ux) \cdot P^X(dx) = 0, \text{ d.h.}$$

$$\varphi_X(u) = E \cos(uX) = \int_{R_1} \cos(ux) F(dx) \quad u \in R_1.$$

Ist umgekehrt  $\varphi_X$  reellwertig, so folgt wegen c) die Gleichung  $\varphi_X(u) = \varphi_{-X}(u)$ ,  $u \in R_1$ .

Aus dem Eindeigkeitssatz für charakteristische Funktionen (siehe unten) ergibt sich, dass die Verteilungsfunktionen von  $X$  und  $-X$  übereinstimmen:

$$F_X = F_{-X}.$$

Das bedeutet  $P^X = P^{-X}$ , also

$$P(X \in B) = P(-X \in B) = P(X \in -B), \quad B \in \mathfrak{B}_1.$$

e) Wir nehmen zunächst  $n = 1$  an. Dann gilt mit  $\varphi = \varphi_X$

$$\frac{\varphi(u+h) - \varphi(u)}{h} = E\left(e^{iuX} \frac{(e^{ihX} - 1)}{h}\right) \text{ und}$$

$$\text{aus } \frac{|e^{ihx} - 1|}{h} \leq |x| \text{ sowie } E|X| < \infty$$

folgt mittels des Satzes von der majorisierten Konvergenz, dass  $\varphi'(u)$  existiert und endlich ist:

$$\varphi'(u) = \lim_{h \rightarrow 0} E\left(e^{iuX} \frac{(e^{ihX} - 1)}{h}\right) = iE(Xe^{iuX}).$$

Die Stetigkeit von  $\varphi'$  zeigt man wie in b) die von  $\varphi$ .

Für  $n > 1$  folgt nunmehr der Beweis mittels vollständiger Induktion analog.

Zum Beweis von (9.4) erinnern wir daran, dass für jedes  $x \in R_1$  und für jedes  $n$  reelle Zahlen  $\vartheta_1, \vartheta_2$  mit  $|\vartheta_i| \leq 1; i = 1, 2$ , existieren, so dass

$$e^{ix} = \cos x + i \sin x =$$

$$\sum_{k=0}^{n-1} \frac{(ix)^k}{k!} + \frac{(ix)^n}{n!} [\cos \vartheta_1 x + i \sin \vartheta_2 x]$$

richtig ist. Daraus ergibt sich für  $x = uX$  die Gleichung

$$Ee^{iuX} = \sum_{k=0}^{n-1} \frac{(iu)^k}{k!} EX^k + \frac{(iu)^n}{n!} [EX^n + \varepsilon_n(u)]$$

mit

$$\varepsilon_n(u) = E[X^n (\cos[\vartheta_1(\omega) \cdot uX] + i \sin[\vartheta_2(\omega) \cdot uX] - 1)].$$

Weiter folgt damit, dass gilt

$$|\varepsilon_n(u)| \leq 3E|X|^n, \quad u \in R_1,$$

und mit dem Satz von der majorisierten Konvergenz erhalten wir

$$\varepsilon_n(u) \rightarrow 0 \text{ für } u \rightarrow 0.$$

Die Eigenschaften f) und g) werden wir im Weiteren nicht benutzen. Für ihren Beweis sei deshalb z. B. auf Siraev (1988), Kap. II; 12 verwiesen.

Die Aussage f) ist i. a. nicht richtig für ungerade  $n = 2m + 1$ , d. h. aus  $|\varphi^{(2m+1)}(0)| < \infty$  folgt noch nicht  $E|X|^{2m+1} < \infty$ . Ein Gegenbeispiel findet man z. B. in Galambos, J., *Advanced Probability Theory*, Marcel-Dekker (1998), Chapter 3.  $\square$

Im Folgenden führen wir vier Eigenschaften charakteristischer Funktionen an, die sie, zusammen mit der eben formulierten Aussagen, zu einem nützlichen Werkzeug der Wahrscheinlichkeitstheorie machen. Ihre Beweise überschreiten den Rahmen dieser Vorlesung. Man findet sie u. a. in Siraev (1988), Kap. II, 12.

**Eindeutigkeitssatz 9.4** *Sind  $F$  und  $G$  zwei Verteilungsfunktionen auf  $R_1$  mit*

$$\varphi_F(u) = \varphi_G(u), \quad u \in R_1,$$

*dann gilt  $F(x) = G(x)$  für alle  $x \in R_1$ .*

**Umkehrformel 9.5** *Es sei  $F$  eine Verteilungsfunktion auf  $R_1$  mit der charakteristischen Funktion  $\varphi_F$ . Dann gilt:*

a) *Für alle  $a, b \in R_1$  mit  $a < b$ , in denen  $F$  stetig ist, gilt*

$$F(b) - F(a) = \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-iau} - e^{-ibu}}{iu} \varphi(u) du. \quad (9.5)$$

b) Ist  $\int_{R_1} |\varphi_F(u)| du < \infty$ , so besitzt  $F$  eine Dichte  $f$ , und es gilt:

$$f(x) = \frac{1}{2\pi} \int_{R_1} e^{-ixu} \varphi_F(u) du, \quad x \in R_1. \quad (9.6)$$

c) Ist  $F$  die Verteilungsfunktion einer diskreten, auf  $N = \{0, 1, 2, \dots, n, \dots\}$  konzentrierten Verteilung ( $p(k), k \geq 0$ ), so gilt

$$p(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-iku} \varphi_F(u) du, \quad k \geq 0. \quad (9.7)$$

**Stetigkeitssatz 9.6** Ist  $(F_n, n \geq 1)$  eine Folge von Verteilungsfunktionen auf  $R_1$ ,  $\varphi_n$  die charakteristische Funktion von  $F_n$ :

$$\varphi_n(u) = \int_{R_1} e^{iux} F_n(dx), \quad u \in R_1,$$

so gilt:

a) Wenn  $w - \lim_{n \rightarrow \infty} F_n = F$  für eine Verteilungsfunktion  $F$  richtig ist, so gilt

$$\lim_{n \rightarrow \infty} \varphi_n(u) = \varphi_F(u) = \int_{R_1} e^{iux} F(dx), \quad u \in R_1.$$

b) Wenn  $\lim_{n \rightarrow \infty} \varphi_n(u) =: \varphi(u)$  für alle  $u \in R_1$  existiert, und wenn die so definierte Funktion  $\varphi$  bei  $u = 0$  stetig ist, so ist  $\varphi$  die charakteristische Funktion einer Verteilungsfunktion  $F$ , und es gilt

$$w - \lim_{n \rightarrow \infty} F_n = F.$$

Bemerkung:  $w - \lim_{n \rightarrow \infty} F_n = F$  bedeutet für Verteilungsfunktionen  $F_n$  und  $F$

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

für alle  $x$ , die Stetigkeitspunkte von  $F$  sind.

**Faltungssatz 9.7** Sind  $X_1$  und  $X_2$  zwei unabhängige reellwertige Zufallsgrößen, so gilt

$$\varphi_{X_1+X_2}(u) = \varphi_{X_1}(u)\varphi_{X_2}(u), \quad u \in R_1. \quad (9.8)$$

Der Beweis dieser Gleichung ergibt sich auf Grund der Unabhängigkeit leicht aus

$$Ee^{iu(X_1+X_2)} = Ee^{iuX_1}Ee^{iuX_2}, \quad u \in R_1.$$

Wir formulieren eine wichtige Folgerung aus dem Faltungssatz.

**Folgerung 9.8** Sind  $X_1, X_2, \dots, X_n$  voneinander unabhängige, identisch verteilte Zufallsgrößen mit der charakteristischen Funktion  $\varphi$ , so gilt für  $M_n :=$

$\frac{1}{n} \sum_{k=1}^n X_k$  die Gleichung

$$\varphi_{M_n}(u) = \left[ \varphi\left(\frac{u}{n}\right) \right]^n, \quad u \in R_1. \quad (9.9)$$

Für eine Ausdehnung des Begriffes der charakteristischen Funktion auf zufällige Vektoren und seine Untersuchung siehe z. B. Jacod, Protter (2000), Chapter 13 oder Siraev (1988), Kap. II, 12. Wir geben hier nur die Definition und eine oft benutzte Aussage an.

Ist  $X = (X_1, \dots, X_n)^T$  ein zufälliger Vektor, so definiert man als seine charakteristische Funktion wie folgt:

$$\varphi_X(u_1, \dots, u_n) := E \exp(iu^T X) \quad , \quad u = (u_1, \dots, u_n)^T \in R_n.$$

Von besonderem Interesse ist dabei die

**Aussage 9.9**  $X_1, \dots, X_n$  sind genau dann voneinander unabhängig, wenn gilt

$$\varphi_X(u_1, \dots, u_n) = \prod_{k=1}^n \varphi_{X_k}(u_k), \quad u = (u_1, \dots, u_n)^T \in R_n,$$

wobei  $\varphi_{X_k}(u_k) = E \exp(iu_k X_k)$  die charakteristische Funktion von  $X_k$  ist,  $k = 1, \dots, n$ .

Zum Beweis siehe die angegebene Literatur.



# Kapitel 10

## Gesetze der großen Zahlen

### 10.1 Einführung

Im ersten Kapitel wurde auf eine Erfahrungstatsache im Umgang mit zufälligen Erscheinungen aufmerksam gemacht, die man gewöhnlich als empirisches Gesetz der großen Zahlen bezeichnet. Gemeint ist die Beobachtung, dass sich bei häufiger Wiederholung eines zufälligen Experimentes der Zufall "ausmittelt", das heißt, dass sich die relativen Häufigkeiten des Eintretens eines mit dem Versuch verbundenen Ereignisses mit wachsender Versuchsanzahl stabilisieren. Eng verbunden damit ist die Beobachtung, dass auch die arithmetischen Mittel der beobachteten Werte einer wiederholt realisierten zufälligen Größe in ähnlicher Weise einem festen Wert zuzustreben scheinen (siehe Abschnitt 3.1).

Diese Erfahrungen sollten sich in einer Wahrscheinlichkeitstheorie als Theoreme wiederfinden. In der Tat liefert die Theorie eine Gruppe von Aussagen über die Konvergenz arithmetischer Mittel von Zufallsgrößen, die man gemeinhin als Gesetze der großen Zahlen bezeichnet. Sie unterscheiden sich in der Konvergenzart der arithmetischen Mittel und in der Art der Voraussetzungen an die zugrunde liegenden Zufallsgrößen.

Die Gesetze der großen Zahlen klären im Rahmen der Kolmogorov'schen Axiomatik der Wahrscheinlichkeitstheorie die Bedingungen an die untersuchten Zufallsgrößen, unter den ihre arithmetischen Mittel im geeigneten Sinne konvergieren und identifizieren ihren Grenzwert. Der Grenzwert steht dabei häufig in

Verbindung mit den Erwartungswerten der betrachteten Zufallsgrößen.

In den folgenden Abschnitten seien  $(X_n, n \geq 1)$  eine Folge reellwertiger Zufallsgrößen über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$ ,  $S_n := \sum_{k=1}^n X_k$  ihre  $n$ -te Partialsumme und  $M_n = \frac{1}{n}S_n$  ihr  $n$ -tes arithmetisches Mittel,  $n \geq 1$ .

Die Fragestellung der Gesetze der großen Zahlen ist:

*Unter welchen Bedingungen an die Zufallsgrößen  $X_n, n \geq 1$ , konvergiert die Folge  $(M_n, n \geq 1)$  in welchem Sinne gegen welchen Grenzwert?*

Gesetze der großen Zahlen gibt es in sehr vielen Varianten. Wir geben hier nur einige wenige exemplarisch an. Weitere interessante Versionen mit samt ihren Anwendungen findet man z. B. in den Monographien zur Wahrscheinlichkeitstheorie von Siraev (1988), Jacod, Protter(2000) oder Bauer (1991).

## 10.2 Schwache Gesetze der großen Zahlen

Als schwache Gesetze der großen Zahlen bezeichnet man gewöhnlich Aussagen, die die stochastische Konvergenz der arithmetischen Mittel  $M_n, n \geq 1$  betreffen.

Wir stellen zunächst die Definition und einige Eigenschaften der stochastischen Konvergenz voran. Die Beweise findet man z. B. in Siraev (1988) oder Jacod, Protter (2000).

**Definition 10.1** *Eine Folge  $(Y_n, n \geq 1)$  reellwertiger Zufallsgrößen über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  heißt stochastisch konvergent gegen eine reellwertige Zufallsgröße  $Y$ , falls  $\lim_{n \rightarrow \infty} P(|Y_n - Y| > \varepsilon) = 0$  für alle  $\varepsilon > 0$ .*

Symbolisch schreibt man in diesem Fall  $Y_n \xrightarrow{P} Y$ .

Die stochastische Konvergenz ist identisch mit der aus der Maßtheorie bekannten Konvergenz dem Maß  $P$  nach.

Mitunter spricht man auch von der *Konvergenz in Wahrscheinlichkeit*.

Offenbar konvergiert eine Folge  $(Y_n)$  genau dann stochastisch gegen  $Y$ , falls

$$\lim_{n \rightarrow \infty} \sup_{m \geq n} P(|Y_m - Y| > \varepsilon) = 0. \quad (10.1)$$

Aus der Maßtheorie ist bekannt, dass die stochastische Konvergenz einer Folge  $(Y_n)$  gegen eine Zufallsgröße  $Y$  äquivalent damit ist, dass man in jeder Teilfolge  $(Y_{n_k})$  eine Unterfolge  $(Y_{n_{k_l}})$  finden kann, die  $P$ -fast sicher gegen  $Y$  konvergiert.

Aus der stochastischen Konvergenz von  $(Y_n)$  gegen  $Y$  ergibt sich die schwache Konvergenz ihrer Verteilungen:

$$\lim_{n \rightarrow \infty} \int_{R_1} f(x) P^{Y_n}(dx) = \int_{R_1} f(x') P^Y(dx), \quad f \in \mathbb{C}(R_1), \quad (10.2)$$

wobei  $\mathbb{C}(R_1)$  die Menge aller stetigen und beschränkten Funktionen auf  $R_1$  bezeichnet.

Die Beziehung (10.2) gilt genau dann, wenn die Verteilungsfunktionen  $F_n$  von  $Y_n$  in allen Punkten  $x$ , in denen die Verteilungsfunktion  $F$  der Zufallsgröße  $Y$  stetig ist, gegen  $F(x)$  konvergieren. Man bezeichnet diese Art der Konvergenz auch als *Konvergenz in Verteilung* und schreibt symbolisch  $Y_n \xrightarrow{d} Y$  bzw.  $F_n \xrightarrow{d} F$ .

Aus der schwachen Konvergenz der Verteilungen von  $(Y_n)$  folgt umgekehrt im Allgemeinen noch nicht die stochastische Konvergenz der Zufallsgrößen  $(Y_n)$ . Wir haben aber die

**Aussage 10.2** *Gilt für alle  $f \in \mathbb{C}(R_1)$  und ein  $x_0 \in R_1$*

$$\lim_{n \rightarrow \infty} \int_{R_1} f(x) P^{Y_n}(dx) = f(x_0),$$

so konvergiert  $(Y_n)$  stochastisch gegen die Zufallsgröße  $Y$ , die  $P$ -fast sicher gleich  $x_0$  ist.

Beweis: Für alle  $\varepsilon > 0$  und für  $f_\varepsilon \in \mathbb{C}(R_1)$ , definiert durch

$$f_\varepsilon(x) = \mathbb{1}_{[x-\varepsilon, x+\varepsilon]^c}(x) + \frac{1}{\varepsilon} \mathbb{1}_{(x-\varepsilon; x+\varepsilon)}(x) |x - x_0|$$

gilt  $f_\varepsilon(x_0) = 0$  und

$$P(|Y_n - x_0| > \varepsilon) \leq \int_{R_1} f_\varepsilon(x) P^{Y_n}(dx) \longrightarrow f_\varepsilon(x_0) \text{ für } \varepsilon \downarrow 0.$$

□

Wir kommen nun zur Formulierung zweier schwacher Gesetze der großen Zahlen.

**Aussage 10.3 (1. Schwaches Gesetz der großen Zahlen)** *Gilt  $D^2 X_n \leq C < \infty, n \geq 1$  für ein  $C > 0$  und  $\text{Kov}(X_k, X_l) = 0$  für alle  $k, l \geq 1$  mit  $k \neq l$ , so konvergieren die zentrierten arithmetischen Mittel  $(M_n - EM_n, n \geq 1)$  stochastisch gegen Null:*

$$\lim_{n \rightarrow \infty} P(|M_n - EM_n| > \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

*Sind insbesondere alle Erwartungswerte  $EX_n, n \geq 1$ , gleich ( $EX_n \equiv EX_1$ ), so konvergieren die Mittel  $(M_n, n \geq 1)$  stochastisch gegen  $EX_1$ :*

$$\lim_{n \rightarrow \infty} P(|M_n - EX_1| > \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

Beweis:

$$\begin{aligned}
 D^2(M_n) &= E(M_n - EM_n)^2 = \frac{1}{n^2} E\left(\sum_{k=1}^n (X_k - EX_k)\right)^2 = \\
 &= \frac{1}{n^2} \sum_{k=1}^n E(X_k - EX_k)^2 + \frac{1}{n^2} \sum_{k \neq \ell} E(X_k - EX_k)(X_\ell - EX_\ell) = \\
 &= \frac{1}{n^2} \left( \sum_{k=1}^n D^2 X_k + \sum_{k \neq \ell} \text{Kov}(X_k, X_\ell) \right) = \\
 &= \frac{1}{n^2} \sum_{k=1}^n D^2 X_k \leq \frac{C}{n}.
 \end{aligned}$$

Für jedes  $\varepsilon > 0$  gilt folglich

$$P(|M_n - EM_n| > \varepsilon) \leq D^2 \frac{M_n}{\varepsilon^2} \leq \frac{C}{n \cdot \varepsilon^2}$$

(Tschebyschev'sche Ungleichung).

Daraus ergibt sich die Behauptung.  $\square$

Sind die  $(X_n, n \geq 1)$  nicht unkorreliert, so liegt im allgemeinen keine Konvergenz der zentrierten arithmetischen Mittel  $(M_n - EM_n, n \geq 1)$  vor oder die Konvergenz erfolgt nicht gegen eine Konstante. Als Beispiel betrachten wir für eine reellwertige Zufallsgröße  $X$  die Folge  $X_n = X, n \geq 1$ , und erhalten  $M_n = X, n \geq 1$ .

Wir geben ein weiteres schwaches Gesetz der großen Zahlen an, in dem auf die Endlichkeit der Varianzen verzichtet wird.

**Aussage 10.4 (2. Schwaches Gesetz der großen Zahlen)** *Es seien  $(X_n, n \geq 1)$  unabhängige, identisch verteilte Zufallsgrößen mit  $E|X_1| < \infty$ , wir setzen  $\mu = EX_1$ .*

Dann gilt

$$\lim_{n \rightarrow \infty} P(|M_n - \mu| > \varepsilon) = 0 \text{ für jedes } \varepsilon > 0.$$

Beweis: Es sei

$$\varphi(u) = \mathbb{E}e^{iuX_1} \text{ und } \varphi_{M_n}(u) = \mathbb{E}e^{iuM_n}, u \in R_1.$$

Dann gilt

$$\varphi_{M_n}(u) = \left[\varphi\left(\frac{u}{n}\right)\right]^n, u \in R_1, \text{ und}$$

$$\varphi(u) = 1 + iu\mu + o(u) \text{ für } u \rightarrow 0.$$

Folglich ist für jedes  $u \in R_1$

$$\varphi\left(\frac{u}{n}\right) = 1 + i\frac{u}{n} \cdot \mu + o\left(\frac{1}{n}\right) \text{ für } n \rightarrow \infty$$

und somit

$$\varphi_{M_n}(u) = \left[1 + \frac{i u}{n} \cdot \mu + o\left(\frac{1}{n}\right)\right]^n \rightarrow e^{iu\mu}, u \in R_1.$$

Die Funktion  $u \rightarrow e^{iu\mu}$  ist die charakteristische Funktion der in  $\mu$  ausgearteten Verteilung und deshalb gilt

$$M_n \xrightarrow{d} \mu,$$

woraus sich auf Grund der entsprechenden obigen Aussage die Behauptung ergibt.  $\square$

### 10.3 Starke Gesetze der großen Zahlen

Als starke Gesetze der großen Zahlen bezeichnet man Aussagen, die die  $P$ -fast sichere Konvergenz der arithmetischen Mittel  $M_n, n \geq 1$  betreffen.

Die  $P$ -fast sichere Konvergenz ist im allgemeinen sehr viel schwieriger zu beweisen, als die stochastische Konvergenz, liefert dafür aber auch für alle  $\omega$  außerhalb einer Nullmenge  $N$  die Konvergenz der  $M_n(\omega)$  gegen einen Grenzwert, wogegen bei der stochastischen Konvergenz nichts über die "individuellen"  $\omega$  bzw.  $M_n(\omega)$  ausgesagt wird.

Es ist interessant, dass es sich bei starken Gesetzen der großen Zahlen tatsächlich nur um eine Konvergenz  $P$ -fast sicher handelt. Das heißt, dass diese Konvergenz im Allgemeinen nicht für alle  $\omega$  aus  $\Omega$  vorliegt. Wir werden diese in 10.4.

an einem Beispiel demonstrieren.

Aus der Fülle der in der Literatur vorhandenen starken Gesetze der großen Zahlen greifen wir drei Beispiele zur Illustration heraus. Für den Beweis der ersten beiden verweisen wir wieder auf Siraev (1988), das dritte werden wir hier beweisen, um einen Einblick in die Technik des Arbeitens mit der fast sicheren Konvergenz zu geben.

**Aussage 10.5** *Es sei  $(X_n, n \geq 1)$  eine Folge reellwertiger, voneinander unabhängiger Zufallsgrößen mit  $EX_n = 0, n \geq 1$ .*

a) *Ist*

$$\sum_{k=1}^{\infty} EX_k^2 < \infty, \quad (10.3)$$

*so konvergiert die Reihe  $\sum_{n=1}^{\infty} X_n$   $P$ -fast sicher,*

b) *Sind die  $X_n$  überdies gleichmäßig beschränkt ( $P(|X_n| \leq c) = 1, n \geq 1$ , für ein  $c > 0$ ), so folgt aus der  $P$ -fast sicheren Konvergenz von  $\sum_{n=1}^{\infty} X_n$  die Eigenschaft (10.3).*

**Beispiel 10.6** Es seien  $(X_n, n \geq 1)$  eine Folge unabhängiger Zufallsgrößen mit

$$P(X_n = +1) = P(X_n = -1) = \frac{1}{2}, n \geq 1,$$

und  $(c_n, n \geq 1)$  eine beschränkte Folge positiver reeller Zahlen.

Genau dann konvergiert die Reihe  $P$ -fast sicher,

$$\sum_{n=1}^{\infty} c_n X_n,$$

wenn  $\sum c_n^2 < \infty$  erfüllt ist.

**Aussage 10.7** Die Folge  $(X_n, n \geq 1)$  bestehe aus voneinander unabhängigen Zufallsgrößen mit

$$\sum_{n=1}^{\infty} \frac{\text{Var}(X_n)}{b_n^2} < \infty$$

für eine Folge positiver Zahlen  $b_n$  mit  $\lim_{n \rightarrow \infty} b_n = \infty$ .

Dann gilt

$$\lim_{n \rightarrow \infty} \frac{S_n - ES_n}{b_n} = 0 \text{ } P\text{-fast sicher.}$$

**Beispiel 10.8** Aus der Bedingung

$$\sum_{n=1}^{\infty} \frac{\text{Var}(X_n)}{n^2} < \infty$$

folgt, dass gilt:

$$\lim_{n \rightarrow \infty} \frac{S_n - ES_n}{n} = 0 \text{ } P\text{-fast sicher.}$$

Zum Beweis dieser beiden Aussagen siehe z. B. Siraev (1988), Kap. IV, 2 und 3.

**Satz 10.9 (Kolmogorov'sches Starkes Gesetz der großen Zahlen)** Die  $(X_n)$  seien voneinander unabhängig und identisch verteilt. Dann gilt:

- a) Existiert der Erwartungswert  $EX_1$  und ist er endlich (d.h.  $E|X_1| < \infty$ ), so konvergiert  $(M_n)$   $P$ -fast sicher mit

$$\lim_{n \rightarrow \infty} M_n = \lim_{n \rightarrow \infty} \frac{S_n}{n} = EX_1 \quad (P\text{-f.s.})$$

- b) Ist  $E|X_1| = \infty$ , so konvergiert  $(M_n)$   $P$ -fast sicher nicht gegen einen endlichen Grenzwert.

c) Ist  $EX_1^+ = \infty$  und  $EX_1^- < \infty$ , so gilt

$$\lim_{n \rightarrow \infty} M_n = \lim_{n \rightarrow \infty} \frac{S_n}{n} = \infty \quad (P - f.s.)$$

Bemerkung: Aus dem Beweis wird deutlich werden, dass es im Fall a) bereits ausreicht voraus zu setzen, dass die  $X_n$  paarweise unabhängig sind.

Beweis: Zu a)

Die  $X_n, n \geq 1$  seien paarweise unabhängig, identisch verteilt und es gelte  $E|X_1| < \infty$ .

Wir gliedern den Beweis in sechs Schritte.

1. Vorbereitungen: Wegen  $X_n = X_n^+ - X_n^-$  und  $E|X_n| < \infty$  folgt  $EX_n^\pm < \infty$  und wir können  $X_n^+$  und  $X_n^-$  einzeln betrachten. O.B.d.A. sei also  $X_n \geq 0, n \geq 1$ . Wir setzen

$$Y_n := X_n \cdot \mathbb{1}_{\{X_n < n\}}, \quad n \geq 1.$$

Die Zufallsgrößen  $Y_n, n \geq 1$  sind ebenfalls paarweise unabhängig (Beweis?), allerdings nicht notwendig identisch verteilt. Es gilt aber  $D^2 Y_n < \infty, n \geq 1$ .

Nun sei

$$\tilde{S}_n := \sum_{k=1}^n Y_k, \text{ woraus sich}$$

$$E\tilde{S}_n = \sum_{k=1}^n EY_k, \quad n \geq 1, \text{ ergibt.}$$

Es seien  $\varepsilon > 0, \alpha > 1$  beliebig, aber fest gewählt.

Wir führen ein:

$$k_n := [\alpha^n] = \max\{k \geq 1 : k \leq \alpha^n\}, \quad n \geq 1,$$

(offenbar gilt  $k_n \leq k_{n+1} \uparrow \infty$  und  $\alpha^n - 1 < k_n \leq \alpha^n$ ,  $n \geq 1$ ),

$$n_i := \min\{n : k_n \geq i\}, \quad i \geq 1.$$

Aus diesen Definitionen folgt sofort

$$\alpha^{n_i} \geq [\alpha^{n_i}] = k_{n_i} \geq i, \quad i \geq 1. \quad (10.4)$$

2. Der nächste Schritt ist der Beweis des folgenden Lemmas.

**Lemma 10.10** *Für eine positive Konstante  $C$  gilt*

$$\sum_{n=1}^{\infty} P\left(\frac{|\tilde{S}_{k_n} - E\tilde{S}_{k_n}|}{k_n} \geq \varepsilon\right) \leq C \cdot EX_1 < \infty$$

Beweis: Mittels der Tschebyschev'schen Ungleichung erhält man für eine von  $\varepsilon$  abhängende Konstante  $C_1$

$$\begin{aligned} \sum_{n=1}^{\infty} P\left(\frac{|\tilde{S}_{k_n} - E\tilde{S}_{k_n}|}{k_n} \geq \varepsilon\right) &\leq C_1 \sum_{n=1}^{\infty} \frac{D^2 \tilde{S}_{k_n}}{k_n^2} = C_1 \\ &\sum_{n=1}^{\infty} \frac{1}{k_n^2} \sum_{i=1}^{k_n} D^2 Y_i. \end{aligned} \quad (10.5)$$

(An dieser Stelle wurde benutzt, dass die  $(X_n)$  und folglich auch die  $(Y_n)$  paarweise unabhängig sind. Die paarweise Unkorreliertheit der  $(X_n)$  würde noch nicht die der  $(Y_n)$  nach sich ziehen.)

Die rechte Seite von (10.5) wird weiter vergrößert:

$$\begin{aligned} C_1 \sum_{n=1}^{\infty} \frac{1}{k_n^2} \sum_{i=1}^{k_n} D^2 Y_i &\leq C_1 \sum_{n=1}^{\infty} \frac{1}{k_n^2} \sum_{i=1}^{k_n} E(Y_i^2) = \\ &C_1 \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} \mathbb{1}_{[1, k_n]}(i) \frac{E(Y_i^2)}{k_n^2} = \\ C_1 \sum_{n=1}^{\infty} \left( \sum_{n: k_n \geq i} \frac{1}{k_n^2} \right) E(Y_i^2) &= C_1 \sum_{i=1}^{\infty} \left( \sum_{n=n_i}^{\infty} \frac{1}{k_n^2} \right) E(Y_i^2) \end{aligned} \quad (10.6)$$

Nun ist aber:

$$\begin{aligned}
 \sum_{n=n_i}^{\infty} \frac{1}{k_n^2} &= \frac{1}{i^2} \left[ \frac{i^2}{[\alpha^{n_i}]^2} + \frac{i^2}{[\alpha^{n_i+1}]^2} + \dots \right] \leq \\
 &\leq \frac{1}{i^2} \left[ \frac{i^2}{(\alpha^{n_i} - 1)^2} + \frac{i^2}{(\alpha^{n_i+1} - 1)^2} + \frac{i^2}{(\alpha^{n_i+2} - 1)^2} + \dots \right] = \\
 &\frac{1}{i^2} \left[ \frac{1}{\left(\frac{\alpha^{n_i}}{i} - \frac{1}{i}\right)^2} + \frac{1}{\left(\frac{\alpha^{n_i+1}}{i} - \frac{1}{i}\right)^2} \dots \right] \leq \\
 &\frac{1}{i^2} \left[ \frac{1}{\left(1 - \frac{1}{2}\right)^2} + \frac{1}{\left(\alpha - \frac{1}{2}\right)^2} + \dots \right] = \frac{C_2}{i^2} \quad i \geq 2,
 \end{aligned}$$

für eine Konstante  $C_2$ , die von  $\alpha$  abhängt. Dabei wurde (10.4), nämlich  $\alpha^{n_i} \geq i$ , benutzt.

Wir nutzen diese Zwischenrechnung zur Fortsetzung der Ungleichung (10.6) und vergrößern deren rechte Seite durch ( $C_3 := C_1 \cdot C_2$ )

$$\leq C_3 \sum_{i=1}^{\infty} \frac{EY_i^2}{i^2} = C_3 \cdot \sum_{i=1}^{\infty} \frac{1}{i^2} \sum_{k=0}^{i-1} \int_{[k, k+1)} x^2 dP^{X_1}. \quad (10.7)$$

(In der letzten Gleichung wurde benutzt, dass alle  $X_n$  identisch wie  $X_1$  verteilt sind und dass gilt

$$\begin{aligned}
 E(Y_i^2) &= EX_i^2 \cdot \mathbb{1}_{\{X_i < i\}} = EX_1^2 \mathbb{1}_{\{X_1 < i\}} = \\
 \sum_{k=0}^{i-1} E(X_1^2 \cdot \mathbb{1}_{\{k \leq X_1 < k+1\}}) &= \sum_{k=0}^{i-1} \int_{[k, k+1)} x^2 P^{X_1}(dx).
 \end{aligned}$$

Die rechte Seite von (10.7) ist gleich folgendem Wert, den wir wiederum nach oben abschätzen:

$$C_3 \sum_{k=0}^{\infty} \left( \sum_{i=k+1}^{\infty} \frac{1}{i^2} \right) \int_{[k,k+1)} x^2 P^{X_1}(dx) \leq$$

$$C_4 \sum_{k=0}^{\infty} k = 0 \frac{1}{k+1} \int_{[k,k+1)} x^2 P^{X_1}(dx) \leq C_4 \sum_{k=0}^{\infty} \int_{[k,k+1)} x P^{X_1}(dx) =$$

$$C_4 EX_1 < \infty.$$

Hier wurde benutzt:

$$\begin{aligned} \sum_{i=k+1}^{\infty} \frac{1}{i^2} &= \frac{1}{(k+1)^2} + \frac{1}{(k+1)^2} + \dots \\ &\leq \frac{1}{k(k+1)} + \frac{1}{(k+1)(k+2)} + \dots \\ &= \frac{1}{k} - \frac{1}{k+1} + \frac{1}{k+1} - \frac{1}{k+2} + \dots = \frac{1}{k} = \frac{k+1}{k} \cdot \frac{1}{k+1} \leq \frac{2}{k+1}. \end{aligned}$$

Damit ist Lemma 10.10 bewiesen.

3. Aus dem eben bewiesenen Lemma folgt auf Grund des ersten Lemmas von Borel-Cantelli

$$\left( \frac{|\tilde{S}_{k_n} - E\tilde{S}_{k_n}|}{k_n} < \varepsilon \quad \forall n \geq n_0. \right)$$

(hier tritt zum ersten Mal eine Eigenschaft  $P$ -fast sicher auf.)

Da  $\varepsilon > 0$  beliebig gewählt war, folgt

$$\lim_{n \rightarrow \infty} \frac{\tilde{S}_{k_n} - E\tilde{S}_{k_n}}{k_n} = 0 \quad P\text{-f.s.} \quad (10.8)$$

4. Jetzt zeigen wir: Die  $\frac{\tilde{S}_{k_n}}{k_n}$  konvergieren fast sicher für  $n \rightarrow \infty$ .

Wegen

$$EY_n = \int_{[0,n)} x dP^{X_1} \uparrow EX_1, \text{ gilt folglich}$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_1^n EY_n = EX_1 \text{ und damit } \lim_{n \rightarrow \infty} \frac{E\tilde{S}_{k_n}}{k_n} = EX_1.$$

Das bedeutet mit :

$$\lim_{n \rightarrow \infty} \frac{\tilde{S}_{k_n}}{k_n} = EX_1 \quad P - \text{fast sicher.}$$

5. Die  $\frac{S_{k_n}}{k_n}$  konvergieren fast sicher: (Wir benutzen wieder, dass alle  $X_n$  dieselbe Verteilung wie  $X_1$  haben.)

Wegen

$$\begin{aligned} \sum_{n=1}^{\infty} P(Y_n \neq X_n) &= \sum_{n=1}^{\infty} P(X_n \geq n) = \sum_{n=1}^{\infty} \int_{[n,\infty)} x P^{X_1}(dx) = \\ &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \int_{[k,k+1)} P^{X_1}(dx) = \sum_{k=1}^{\infty} \sum_{n=1}^k \int_{[k,k+1)} P^{X_1}(dx) = \\ &= \sum_{k=1}^{\infty} k \cdot \int_{[k,k+1)} P^{X_1}(dx) \leq \sum_{k=1}^{\infty} \int_{[k,k+1)} x P^{X_1}(dx) \leq EX_1 < \infty \end{aligned}$$

folgt wiederum aus dem 1. Lemma von Borel-Cantelli:

$$P\{X_n \neq Y_n \text{ für unendlich viele } n\} = 0.$$

Das bedeutet aber wegen  $k_n \uparrow \infty$ , dass gilt

$$\lim_{n \rightarrow \infty} \frac{S_{k_n}}{k_n} = EX_1 \quad P - \text{f.s.}$$

6. Die  $\frac{S_n}{n}$  konvergieren fast sicher:  
 Jedes  $n \geq 1$  liegt zwischen zwei der  $k_m$ .  
 Es sei  $m = m(n)$  derart, dass  $k_{m(n)} \leq n < k_{m(n)+1}$ .  
 Offenbar gilt  $m(n) \rightarrow \infty$  für  $n \rightarrow \infty$ .  
 Weil  $n \rightarrow S_n = \sum_1^n X_k$  eine nichtfallende Folge ist, gilt

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{S_n}{n} &\geq \liminf_{n \rightarrow \infty} \frac{S_{k_{m(n)}}}{k_{m(n)}} \cdot \frac{k_{m(n)}}{k_{m(n)+1}} \\ &\geq \frac{1}{\alpha} \lim_{n \rightarrow \infty} \frac{S_{k_{m(n)}}}{k_{m(n)}} = \frac{1}{\alpha} EX_1 \end{aligned}$$

Hierbei wurde benutzt:

$$\begin{aligned} \frac{k_{m(n)}}{k_{m(n)+1}} &= \frac{[\alpha^{m(n)}]}{[\alpha^{m(n)+1}]} = \frac{\alpha^{m(n)} - (\alpha^{m(n)} - [\alpha^{m(n)}])}{\alpha^{m(n)+1} - (\alpha^{m(n)+1} - [\alpha^{m(n)+1}])} = \\ &= \frac{1}{\alpha} \frac{\left(1 - \frac{(\alpha^{m(n)} - [\alpha^{m(n)}])}{\alpha^{m(n)}}\right)}{1 - \frac{(\alpha^{m(n)+1} - [\alpha^{m(n)+1}])}{\alpha \cdot \alpha^{m(n)}}} \xrightarrow{n \rightarrow \infty} \frac{1}{\alpha}. \\ \text{da } \frac{x - [x]}{x} &\leq \frac{1}{x} \xrightarrow{x \uparrow \infty} 0. \end{aligned}$$

Analog zeigt man:

$$\limsup_{n \rightarrow \infty} \frac{S_n}{n} \leq \alpha EX_1 \quad P - \text{f.s.}$$

Da  $\alpha > 1$  beliebig gewählt war, folgt

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = EX_1 \quad P - \text{f.s.}$$

Damit ist a) bewiesen.

b) Es sei  $E|X_1| = \infty$ . Angenommen, es gilt nicht, dass  $(M_n)$   $P$ -fast sicher gegen keinen endlichen Grenzwert konvergiert, dann haben wir:

$$P(\underbrace{(M_n) \text{ konvergiert gegen einen endlichen Grenzwert}}_{=:A}) > 0.$$

Da das Ereignis  $A$  zur Tail- $\sigma$ -Algebra  $\tau = \bigcap_n \bigvee_{k \geq n} \sigma(X_k)$  gehört, folgt aus dem Kolmogorov'schen 0-1-Gesetz  $P(M_\infty) = 1$ , wobei  $M_\infty(\omega) := \lim_{n \rightarrow \infty} M_n(\omega)$ ,  $\omega \in A$ , gesetzt wurde. Hier wird benutzt, dass die  $X_n$ ,  $n \geq 1$ , nicht nur paarweise, sondern insgesamt voneinander unabhängig sind. Also haben wir

$$\lim_{n \rightarrow \infty} M_n =: M_\infty \quad P\text{-f.s. mit } P(|M_\infty| < \infty) = 1.$$

Daraus folgt  $\frac{X_n}{n} = \frac{S_n}{n} - \frac{S_{n-1}}{n-1} \cdot \frac{n-1}{n} \rightarrow 0$ .

Also hat das Ereignis  $\left\{ \frac{|X_n|}{n} \geq 1 \text{ unendlich oft} \right\}$  die Wahrscheinlichkeit 0.

Aus dem 2. Lemma von Borel-Cantelli folgt  $\sum P\left(\frac{|X_n|}{n} \geq 1\right) < \infty$ .

Wegen  $\infty = E|X_1| = \sum_{k=0}^{\infty} P(|X_1| \geq k) = \sum_{k=0}^{\infty} P(|X_k| \geq k)$  dies ist ein Widerspruch.

c) Es sei  $C > 0$  beliebig. Mit

$S_n^C := \sum_{k=1}^n X_k \mathbb{1}_{\{X_k \geq C\}}$  gilt  $EX_1 \mathbb{1}_{\{X_1 \leq C\}} < \infty$  und die  $(X_n^C)$  sind unabhängig.

Es ist

$$\liminf \frac{S_n}{n} \geq \liminf \frac{S_n^C}{n} = EX_1^C \uparrow \infty \text{ für } C \uparrow \infty.$$

Daraus folgt:

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \infty. \quad P\text{-f.s.}$$

Damit ist das starke Gesetz der großen Zahlen bewiesen.  $\square$

## 10.4 Anwendungen des starken Gesetzes der großen Zahlen

### 10.4.1 Die Monte-Carlo-Methode

Es seien  $(X_n, n \geq 1)$  und  $(Y_n, n \geq 1)$  zwei Folgen von Zufallsgrößen über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$ , die alle dieselbe Verteilung besitzen, und zwar die gleichmäßige Verteilung auf  $([0, 1), \mathcal{B}_{[0,1)})$ .

Außerdem seien alle  $X_n, Y_m, n, m \geq 1$ , voneinander unabhängig. Folglich ist  $(X_n, Y_n), n \geq 1$ , eine Folge unabhängiger, gleichmäßig auf  $([0, 1)^2, \mathcal{B}_{[0,1)^2})$  verteilter zweidimensionaler zufälliger Vektoren.

Es sei  $B$  eine Borelmenge aus  $[0, 1)^2$  mit  $p := \lambda_2(B) \in (0, 1)$ .

Dann bildet  $(Z_n, n \geq 1)$  mit

$$Z_n := \mathbb{1}_B(X_n, Y_n), \quad n \geq 1$$

ein Bernoullischema mit dem Parameter  $p$ . Offenbar gilt  $p = EZ_1$ .

Aus dem Gesetz der großen Zahlen folgt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n Z_k = EZ_1 \quad P - \text{fast sicher.}$$

Man kann also den Flächeninhalt  $\lambda_2(B)$  der Menge  $B$  näherungsweise bestimmen, indem man das Bernoullischema  $(Z_n, n \geq 1)$  sehr oft, sagen wir  $n$ -mal, ausführt (d. h., indem man nacheinander und unabhängig voneinander Punkte aus  $[0, 1)^2$  rein zufällig auswählt) und die relative Häufigkeit bestimmt, mit der die Punkte  $Z_k, k = 1, \dots$ , in  $B$  fallen.

Beispiel:  $B = \{(x, y) \in [0, 1)^2 : x^2 + y^2 \leq 1\}, p = \lambda_2(B) = \frac{\pi}{4}$ .

### 10.4.2 "Normale Zahlen" aus $[0, 1)$

Es seien  $\Omega = [0, 1), \mathfrak{A} = \mathcal{B}_{[0,1)}$  und  $P = \lambda_{[0,1)}$  (Lebesguemaß auf  $[0, 1)$ ). Jedes  $\omega \in \Omega$  hat genau eine dyadische Darstellung

$$\omega = 0, \omega_1, \omega_2, \dots \quad \text{mit} \quad \omega_i \in \{0, 1\}, i \geq 1,$$

in der  $\omega_i = 0$  unendlich oft vorkommt.

Wir setzen

$$X_n(\omega) := \omega_n, \quad n \geq 1, \omega \in \Omega.$$

**Lemma 10.11**  $(X_n, n \geq 1)$  bildet ein Bernoullischema mit dem Parameter  $p = \frac{1}{2}$ .

Beweis: Es gilt  $P(X_1 = i_1, \dots, X_n = i_n) =$

$$P\left(\left\{\omega : \sum_{k=0}^n \frac{i_k}{2^k} \leq \omega < \sum_{k=0}^n \frac{i_k}{2^k} + \frac{1}{2^n}\right\}\right) = \frac{1}{2^n}.$$

Insbesondere ist  $P(X_k = i_k) = \frac{1}{2}; k \geq 1$ .

Also sind die  $(X_n, n \geq 1)$  voneinander unabhängig und identisch verteilt mit  $P(X_n = 1) = P(X_n = 0) = \frac{1}{2}, n \geq 1$ .

Aus dem starken Gesetz der großen Zahlen folgt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_n(\omega) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \omega_k = \frac{1}{2} \quad \lambda_{[0,1)} \text{ - fast sicher.}$$

**Definition 10.12** Eine reelle Zahl  $x$  aus  $[0, 1)$  heißt normal, falls in ihrer Dualdarstellung  $x = 0, i_1, i_2, \dots$  mit unendlich vielen Nullen gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n i_k = \frac{1}{2}.$$

Das starke Gesetz der großen Zahlen impliziert also, dass Lebesgue-fast alle Zahlen aus  $[0, 1)$  normal sind. Die dyadischen Zahlen  $k \cdot 2^{-n} (0 \leq k < 2^n, n \geq 1)$  sind nicht normal, da für sie  $i_k = 0$  für alle hinreichend große  $k$  gilt. Es ist

unbekannt, ob  $\sqrt{2}$ ,  $\log 2$ ,  $e$ ,  $\pi$  normal sind.

An diesem Beispiel wird deutlich, dass die Konvergenz im starken Gesetz der großen Zahlen tatsächlich nicht für alle  $\omega \in \Omega$ , sondern nur für alle Punkte  $\omega$  außerhalb einer  $P$ -Nullmenge vorzuliegen braucht.

# Kapitel 11

## Zentrale Grenzwertsätze

Viele zufällige Größen in der Natur, Wirtschaft und Gesellschaft sind das Ergebnis einer Überlagerung zahlreicher kleiner zufälliger Einflüsse, die weitgehend unabhängig voneinander wirken. So ist der tägliche Schlusskurs einer Aktie das Ergebnis einer i. Allg. großen Zahl von Käufen und Verkäufen, Messergebnisse werden häufig durch zahlreiche Einwirkungen zufälliger Art beeinflusst (Temperatur, Ablesefehler u. a.). Die Wahrscheinlichkeitstheorie widmet sich diesen Fragen, indem sie die Wahrscheinlichkeitsverteilungen der Summe einer großen Anzahl  $n$  von einzelnen Zufallsgrößen studiert.

Wie oft in der Mathematik üblich, geht man dabei zum Grenzwert für  $n \rightarrow \infty$  über, um übersichtliche Resultate zu erzielen. Eine Gruppe entsprechender Sätze, die sogenannten *zentralen Grenzwertsätze*, befasst sich mit Bedingungen an die zugrunde liegenden Zufallsgrößen, unter denen eine *Normalverteilung* im Limes erscheint.

### 11.1 Lokaler Grenzwertsatz von Moivre-Laplace

Es sei  $(X_n, n \geq 1)$  ein Bernoullischema mit dem Parameter  $p \in (0, 1)$ . Dann besitzt  $S_n = \sum_{k=1}^n X_k$  bekanntlich (siehe Aussage 6.3) eine Binomialverteilung mit den Parametern  $n$  und  $p$ :

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} =: b(n, p; k), \quad k = 0, 1, \dots, n. \quad (11.1)$$

Es gilt (vgl. Beispiele 4.13 c) und 4.21 c))

$$ES_n = np, \text{ und } D^2S_n = npq \text{ mit } q = 1 - p.$$

Wir untersuchen, wie sich die Verteilung von  $S_n$  bei unbegrenzt wachsendem  $n$  verändert. Offenbar wachsen  $ES_n$  und  $D^2S_n$  unbeschränkt falls  $n$  nach unendlich strebt, und  $b(n, p; k)$  konvergiert für  $n \rightarrow \infty$  bei festen  $p$  und  $k$  gegen Null. (Beachten Sie  $b(n, p; k) \leq \left(\frac{p}{1-p}\right)^k \cdot \frac{1}{k!} \cdot n^k (1-p)^n$ .)

Um dennoch etwas über die asymptotischen Eigenschaften der Binomialverteilung für  $n \rightarrow \infty$  aussagen zu können, gehen wir zur standardisierten Zufallsgröße  $S_n^*$  über:

$$S_n^* = \frac{S_n - ES_n}{\sqrt{D^2S_n}} = \frac{S_n - np}{\sqrt{npq}}.$$

Diese Zufallsgröße hat die möglichen Werte

$$x_k^{(n)} := \frac{k - np}{\sqrt{npq}},$$

die sie jeweils mit der Wahrscheinlichkeit  $b(n, p; k)$  annimmt,  $k = 0, 1, \dots, n$ . Die  $x_k^{(n)}$  ( $k = 0, 1, \dots, n$ ) bilden ein Gitter mit dem Gitterabstand  $\Delta_n := (npq)^{-\frac{1}{2}}$ , dem kleinsten Gitterpunkt  $x_0^{(n)} = -\sqrt{\frac{np}{q}}$  und dem größten  $x_n^{(n)} = \sqrt{\frac{nq}{p}}$ . Wir führen eine Funktion  $\varphi_n(\cdot)$  auf folgende Weise ein:

$$\varphi_n(x) = \frac{b(n, p; k)}{\Delta_n} \quad \text{falls } x \in \left[ x_k^{(n)} - \frac{\Delta_n}{2}, x_k^{(n)} + \frac{\Delta_n}{2} \right)$$

$$(k = 0, 1, \dots, n).$$

$$\varphi_n(x) = 0, \quad \text{falls } x < x_0^{(n)} \quad \text{oder falls } x \geq x_n^{(n)}.$$

$\varphi_n$  beschreibt ein Säulendiagramm mit  $(n+1)$  senkrechten Säulen der Höhe  $\varphi_n(x_k^{(n)})$ , der Breite  $\Delta_n$  und den Säulenmitten  $x_k^{(n)}$ ,  $k = 0, 1, \dots, n$ .

Die Fläche der  $k$ -ten Säule beträgt  $b(n, p; k)$  und die Gesamtfläche unter der Oberkante des Säulendiagramms ist gleich Eins.

**Satz 11.1 (Lokaler zentraler Grenzwertsatz von Moivre-Laplace)** Für alle  $a > 0$  gilt

$$\lim_{n \rightarrow \infty} \sup_{|x| \leq a} |\varphi_n(x) - \varphi(x)| = 0,$$

$$\text{wobei } \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right]$$

die Dichte der Standard-Normalverteilung auf  $\mathbb{R}_1$  ist:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{s^2}{2}} ds, \quad x \in \mathbb{R}_1.$$

Beweis: siehe z. B. Siraev, I, § 6.

Der Beweis stützt sich im Wesentlichen auf die Stirling'sche Formel der Approximation von Fakultäten  $n!$ .

Der lokale Grenzwertsatz von Moivre-Laplace wird häufig benutzt, um Wahrscheinlichkeiten der Form  $P(k \leq S_n \leq l)$  näherungsweise zu bestimmen. Es gilt nämlich wegen  $S_n^* = \frac{S_n - np}{\sqrt{npq}}$  die Beziehung

$$\begin{aligned} P(k \leq S_n \leq l) &= \sum_{m=k}^l b(n, p; m) = \\ &P\left(\frac{k - np}{\sqrt{npq}} \leq S_n^* \leq \frac{l - np}{\sqrt{npq}}\right) = \\ &P\left(x_k^{(n)} \leq S_n^* \leq x_l^{(n)}\right) = \end{aligned}$$

$$\int_{x_k^{(n)} - \frac{\Delta_n}{2}}^{x_k^{(n)} + \frac{\Delta_n}{2}} \varphi_n(s) ds \sim \int_{x_k^{(n)} - \frac{\Delta_n}{2}}^{x_k^{(n)} + \frac{\Delta_n}{2}} \varphi(s) ds = \Phi\left(x_l^{(n)} + \frac{\Delta_n}{2}\right) - \Phi\left(x_k^{(n)} - \frac{\Delta_n}{2}\right) \quad (11.2a)$$

Analog erhält man

$$P(k \leq S_n < l) \approx \Phi\left(x_l^{(n)} - \frac{\Delta_n}{2}\right) - \Phi\left(x_k^{(n)} - \frac{\Delta_n}{2}\right) \quad (11.2b)$$

$$P(k < S_n \leq l) \approx \Phi\left(x_l^{(n)} + \frac{\Delta_n}{2}\right) - \Phi\left(x_k^{(m)} + \frac{\Delta_n}{2}\right) \quad (11.2c)$$

$$P(k < S_n < l) \approx \Phi\left(x_l^{(n)} + \frac{\Delta_n}{2}\right) - \Phi\left(x_k^{(n)} - \frac{\Delta_n}{2}\right). \quad (11.2d)$$

Häufig trifft man auf die folgenden etwas ungenaueren Approximationen, die wir als "grobe Approximation" bezeichnen wollen, im Gegensatz zu der vorhergehenden "feinen Approximation".

$$P(k \leq (<)S_n \leq (<)l) \approx \Phi(x_l^{(n)}) - \Phi(x_k^{(n)}),$$

wobei auf der linken Seite jeweils entweder  $\leq$  oder  $<$  steht. Sie liefert für größere  $n$  ebenfalls brauchbare Werte.

**Beispiel 11.2** 16-maliges Werfen einer regulären Münze. Wie groß ist die Wahrscheinlichkeit, dass mindestens sechs und höchstens zehnmal die Zahl oben liegt?

$$n = 16, p = \frac{1}{2}, k = 6, l = 10, np = 8, npq = 4$$

1. Exaktes Resultat:

$$P(6 \leq S_{16} \leq 10) = \left[2 \binom{16}{6} + 2 \binom{16}{7} + \binom{16}{8}\right] 2^{-16} =$$

$$0,244 + 0,349 + 0,196 = 0,789$$

2. "Grobe" Approximation:

$$P(6 \leq S_{16} \leq 10) = P(-1 \leq S_{16}^* \leq 1) \approx$$

$$\Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0,6826$$

3. "Feine" Approximation:

$$P(6 \leq S_{16} \leq 10) = P\left(-1 - \frac{1}{4} \leq S_{16}^* \leq 1 + \frac{1}{4}\right) \approx$$

$$\Phi(1, 25) - \Phi(-1, 25) = 2\Phi(1, 25) - 1 = 2 \cdot 0,8944 - 1 = 0,788.$$

Die Approximation ist nicht so gut für  $p \neq \frac{1}{2}$ . Man berechne sie für  $n = 16, p = 0,2$ .

## 11.2 Der zentrale Grenzwertsatz von Feller-Lévy

**Voraussetzung 11.3** Es seien  $(X_n, n \geq 1)$  eine Folge unabhängiger, identisch verteilter Zufallsgrößen mit  $\sigma^2 := D^2X_1 \in (0, \infty)$  und  $S_n := \sum_{k=1}^n X_k$ .

Insbesondere gilt

$$ES_n = nEX_1, D^2S_n = n\sigma^2. \quad (11.3)$$

Das Gesetz der großen Zahlen besagt, dass die arithmetischen Mittel  $M_n = \frac{1}{n}S_n$   $P$ -fast sicher gegen  $EX_1$  konvergieren. Insbesondere streben im vorliegenden Fall auch die Streuungen  $D^2M_n = \frac{\sigma^2}{n}$  gegen Null. Daraus folgt, dass die Verteilungen von  $M_n$  gegen die ausgeartete Verteilung, die in  $EX_1$  konzentriert ist, konvergieren. Wenn man dagegen  $S_n$  zentriert und normiert zu

$$S_n^* = \frac{S_n - ES_n}{\sqrt{D^2S_n}} \quad (\text{Standardisierung})$$

so hat  $S_n^*$  den Erwartungswert Null und die Streuung Eins, und zwar für jedes  $n \geq 1$ .

Der folgende Grenzwert stellt fest, dass die Verteilungen der  $S_n^*$  gegen die Standard-Normalverteilung konvergieren.

**Satz 11.4 (Zentraler Grenzwertsatz von Feller-Lévy)** Für die standardisierten Zufallsgrößen  $S_n^*$ ,  $n \geq 1$  gilt

$$\lim_{n \rightarrow \infty} \sup_{-\infty \leq a < b \leq \infty} |P(a < S_n^* \leq b) - (\Phi(b) - \Phi(a))| = 0, \quad (2) \quad (11.4)$$

wobei  $\Phi$  die Verteilungsfunktion der Standard-Normalverteilung ist.

Der Beweis erfolgt mittels des Faltungsgesetzes und des Stetigkeitssatzes für charakteristische Funktionen, vgl. auch Übung 12.6.

Die Beweisidee lässt sich folgendermaßen skizzieren:

$$\begin{aligned} \varphi_{S_n^*}(u) &= E e^{iuS_n^*} = E e^{iu \left( \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \right)} = \\ &= E \exp \left[ \frac{iu}{\sigma\sqrt{n}} ((X_1 - \mu) + \dots + (X_n - \mu)) \right] = \\ &= \left[ E \exp \left[ \frac{iu}{\sigma\sqrt{n}} (X_1 - \mu) \right] \right]^n = \\ &= \left[ \varphi_{X_1} \left( \frac{u}{\sigma\sqrt{n}} \right) e^{-\frac{i u \mu}{\sigma\sqrt{n}}} \right]^n, \quad u \in \mathbb{R}_1. \end{aligned}$$

Da nach Voraussetzung  $EX_1^2 < \infty$  gilt, ist die charakteristische Funktion  $\varphi_{X_1}$  zweimal stetig differenzierbar, und es gilt

$$\varphi_{X_1}(v) = 1 + i\mu v - \frac{v^2}{2} EX_1^2 + o(v^2)$$

und

$$e^{iw} = 1 + iw - \frac{w^2}{2} + o(w^2)$$

und man erhält mit  $v = \frac{u}{\sigma\sqrt{n}}$  bzw.  $w = -\frac{u\mu}{\sigma\sqrt{n}}$  die Beziehung

$$\varphi_{S_n^*}(u) = \left[ 1 - \frac{u^2}{2n} + o\left(\frac{1}{n}\right) \right]^n \xrightarrow{n \rightarrow \infty} e^{-\frac{u^2}{2}}, \quad u \in \mathbb{R}_1.$$

Der Grenzwert ist aber gerade die charakteristische Funktion der Standard-Normalverteilung. Nun ergibt sich die Aussage des Satzes aus dem Stetigkeitssatz für charakteristische Funktionen und der Tatsache, dass die  $w$ -Konvergenz

von Verteilungsfunktionen  $F_n$  gegen eine Verteilungsfunktion  $F$  (siehe Bemerkung nach Stetigkeitssatz 9.6) im Falle, dass  $F$  stetig ist, gleichmäßig erfolgt (Übung 11.1).

Der eben angegebene Zentrale Grenzwertsatz ist ein geeignetes Hilfsmittel, um mit guter Näherung Wahrscheinlichkeiten bestimmen zu können, die im Zusammenhang mit arithmetischen Mitteln unabhängiger, identisch verteilter Zufallsgrößen stehen.

Wir werden dafür einige Beispiele angeben. Sie stützen sich alle auf folgende Näherungsgleichung:

Auf Grund des Zentralen Grenzwertsatzes gilt

$$F_{S_n^*}(x) := P(S_n^* \leq x) \approx \Phi(x), \quad x \in R_1 \quad (11.5)$$

Wir werden im Folgenden diese Näherung verwenden, die in Anwendungsfällen meist für nicht allzu große  $n$  genügend genau erfüllt ist. (Zur genauen Konvergenzgeschwindigkeit siehe die Ungleichung 11.4.)

Insbesondere folgen die häufig nützlichen Formeln:

$$P(S_n \leq x\sigma\sqrt{n} + n\mu) \approx \Phi(x), \quad x \in R_1, \quad (11.6)$$

$$P(S_n \leq y) \approx \Phi\left(\frac{y - n\mu}{\sigma\sqrt{n}}\right), \quad y \in R_1, \quad (11.7)$$

$$P\left(\left|\frac{S_n}{n} - \mu\right| > c\right) \approx 2\left(1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right)\right), \quad c > 0, \quad (11.8)$$

$$P\left(\left|\frac{S_n}{n} - \mu\right| \leq c\right) \approx 2\Phi\left(\frac{c\sqrt{n}}{\sigma}\right) - 1, \quad c > 0. \quad (11.9)$$

Die Werte der Standard-Normalverteilungsfunktion  $\Phi$  entnimmt man einer entsprechenden Tabelle. (Erinnert sei an die Voraussetzung (11.3).)

Im Folgenden geben wir einige Anwendungen dieser Formeln an. Dabei setzen wir voraus, dass  $(X_n, n \geq 1)$  den Voraussetzungen 11.3 des Satzes von Feller-

Lévy genügt und definieren wie gehabt  $S_n = \sum_{k=1}^n X_k$ .

- a) Mit welcher Wahrscheinlichkeit weicht das arithmetische Mittel  $\frac{S_n}{n}$  um mehr als  $c$  vom Erwartungswert  $\mu$  ab?

Antwort: Wegen (11.7) mit der Wahrscheinlichkeit

$$P\left(\left|\frac{S_n}{n} - \mu\right| > c\right) \approx 2\left(1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right)\right).$$

Mit welcher Wahrscheinlichkeit überdeckt das (zufällige) Intervall  $\left(\frac{S_n}{n} - c, \frac{S_n}{n} + c\right)$  den Erwartungswert  $\mu$ ?

Antwort: Mit der Wahrscheinlichkeit (siehe (11.8))

$$P\left(\frac{S_n}{n} - c \leq \mu \leq \frac{S_n}{n} + c\right) = P\left(\left|\frac{S_n}{n} - \mu\right| \leq c\right) =$$

$$1 - P\left(\left|\frac{S_n}{n} - \mu\right| > c\right) \approx 2\Phi\left(\frac{c\sqrt{n}}{\sigma}\right) - 1$$

- b) Es seien  $\alpha \in (0, 1)$  und  $n$  vorgegeben. Wie groß muss man  $c$  mindestens wählen, damit gilt

$$P\left(\left|\frac{S_n}{n} - \mu\right| > c\right) \leq \alpha?$$

Antwort: Wegen (11.7) wählt man  $c$  mindestens so groß, dass  $2\left(1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right)\right) \leq \alpha$  erfüllt ist. Das bedeutet

$$c \geq q_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}},$$

wobei  $q_p$  das  $p$ -Quantil der Standard-Normalverteilung bezeichnet. Siehe Definition 3.35 und Aussage 3.38 sowie die Normalverteilungstabelle.

Mit der Wahrscheinlichkeit  $1 - \alpha$  gilt dann für die Beziehung  $c = q_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

$$n\mu - q_{1-\frac{\alpha}{2}} \cdot \sigma \sqrt{n} \leq S_n \leq n\mu + q_{1-\frac{\alpha}{2}} \cdot \sigma \cdot \sqrt{n}, \quad \text{d.h.}$$

$$\mu - q_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \frac{S_n}{n} \leq \mu + q_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

- c)  $\alpha \in (0, 1)$  und  $c > 0$  seien gegeben. Wie groß sollte man  $n$  mindestens wählen, damit gilt:

$$P\left(\left|\frac{S_n}{n} - \mu\right| \leq c\right) \geq 1 - \alpha$$

Antwort:  $P\left(\left|\frac{S_n}{n} - \mu\right| \leq c\right) = 1 - P\left(\left|\frac{S_n}{n} - \mu\right| > c\right) \approx$

$$1 - 2\left(1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right)\right) \geq 1 - \alpha$$

Also sollte man  $n$  mindestens so groß wählen, dass

$$\Phi\left(\frac{c\sqrt{n}}{\sigma}\right) \leq 1 - \frac{\alpha}{2}$$

gilt, d. h.

$$n \geq \frac{\sigma^2}{c^2} \cdot q_{1-\frac{\alpha}{2}}^2 \quad \text{also}$$

$$n \geq n_0 = \left\lceil \frac{\sigma^2}{c^2} \cdot q_{1-\frac{\alpha}{2}}^2 \right\rceil + 1,$$

wobei  $[z] = \max\{k \geq 0 \mid k \in \mathbb{N}, k \leq z\}$ , ( $z > 0$ ) gesetzt wird.

Um die Konvergenzgeschwindigkeit im zentralen Grenzwertsatz von Lévy-Feller abschätzen zu können, ist folgende Ungleichung nützlich.

**Ungleichung 11.5 von Berry-Essen**

Unter den Voraussetzungen des Satzes von Feller-Lévy und der Annahme  $E|X_1|^3 < \infty$  gilt:

$$\sup_x |F_{S_n^*}(x) - \Phi(x)| \leq C \cdot \frac{E|X_1 - \mu|^3}{\sigma^3 \sqrt{n}} \quad (11.10)$$

mit einer Konstanten  $C$ , für die  $(2\pi)^{-\frac{1}{2}} < C < 0,8$  gilt.

Die Konvergenzordnung  $n^{-\frac{1}{2}}$  kann im Allgemeinen nicht verbessert werden. (Siehe z. B. Siraev (1988), Kap. III, §4.)

**Der Spezialfall der Binomialverteilung**

Für den Fall, dass die  $(X_n, n \geq 1)$  ein Bernoullischema mit dem Parameter  $p$  bilden, gilt natürlich der zentrale Grenzwertsatz von Feller-Lévy und wird aus historischen Gründen als globaler zentraler Grenzwertsatz von Moivre-Laplace bezeichnet.

Mit

$$S_n = \sum_{k=1}^n X_k, S_n^* = \frac{S_n - np}{\sqrt{npq}}, n \geq 1$$

gilt also in diesem Fall

$$\lim_{n \rightarrow \infty} \sup_{-\infty \leq a < b \leq \infty} |P(a < S_n^* \leq b) - (\Phi(b) - \Phi(a))| = 0.$$

Eine ähnliche Ungleichung wie die von Berry-Essen (11.9) lautet hier

$$\sup_{x \in R_1} |F_{S_n^*}(x) - \Phi(x)| \leq \frac{p^2 + q^2}{\sqrt{npq}} \quad (11.11)$$

(vgl. Siraev, Kap. I, §6).

**Bemerkung 11.6** Als "Praxiserfahrung" gibt Henze (2006) in seinem Kap. 27 die Faustregel  $npq \geq 9$  als "für praktische Zwecke ausreichend" an.

Ist  $npq \geq 9$  nicht erfüllt,  $n$  aber nicht zu klein, so ist evtl. der *Poisson'sche Grenzwertsatz* anwendbar (vgl. Übung 6.6):

Ist  $p_n \xrightarrow{n \rightarrow \infty} 0$ , mit  $np_n \xrightarrow{n \rightarrow \infty} \lambda > 0$  so gilt für jede  $k \geq 0$

$$\lim_{\substack{n \rightarrow \infty \\ p_n \rightarrow 0 \\ np_n \rightarrow \lambda}} \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda} =: \pi_k(\lambda).$$

Für  $p \ll 1, n \gg 1$  und  $\lambda := np < 9$  kann man mit der Näherung

$$\binom{n}{k} p^k (1 - p)^{n-k} \approx \pi_k(\lambda)$$

rechnen.

### Zahlenbeispiele 11.7

- a) In einem Computerprogramm wird nach jeder Operation auf die  $j$ -te Dezimale gerundet. Rundungsfehler addieren sich, sind unabhängig und gleichverteilt auf  $\left[-\frac{10^{-j}}{2}; \frac{+10^{-j}}{2}\right]$ ,  $n = 10^6$  Operationen werden ausgeführt. Wie groß ist die Wahrscheinlichkeit dafür, dass der absolute Rundungsfehler im Endresultat größer als  $c = 500 \cdot 10^{-j}$  ist?

Antwort: Hier sind  $X_1, \dots, X_n$  unabhängig, identisch verteilt,  $EX_i = 0$ ,  $D^2 X_i = \frac{10^{-2j}}{12}$ .

Auf Grund des Zentralen Grenzwertsatzes von Lévy-Feller ist

$S_n^* = S_n \cdot \sqrt{12} \cdot 10^j$  annähernd Standard-normalverteilt.

Für die gesuchte Wahrscheinlichkeit erhalten wir

$$P(|S_n| > 500 \cdot 10^{-j}) = P(|S_n^*| > \frac{\sqrt{12} \cdot 500}{10^3}) =$$

$$P(|S_n^*| > \sqrt{3}) \approx 2(1 - \Phi(\sqrt{3})) =$$

$$2(1 - 0,9584) = 0,083.$$

Will man dagegen eine Schranke, die mit Sicherheit gilt, rechnet man mit dem ungünstigsten Fall, dass alle Fehler gleiches Vorzeichen haben und sich summieren. Dann kann man nur sagen, dass mit Sicherheit

$$S_n \in \left[ -\frac{10^{6-j}}{2}, +\frac{10^{6-j}}{2} \right]$$

gilt. Das sind Schranken, die weit größer als die vorher bestimmten sind.

- b) Ein regulärer Spielwürfel wird 1000mal unabhängig voneinander geworfen. Der Erwartungswert der Augensumme beträgt 3500. In welchem (möglichst kleinem) Intervall  $[3500 - c, 3500 + c]$  wird die Augensumme mit der Wahrscheinlichkeit 0,95 (0,99 bzw. Eins) liegen?

Antwort: Die Wahrscheinlichkeit

$$P(|S_{1000} - 3500| \leq c) = P(|S_{1000}^*| \leq \frac{c}{\sigma\sqrt{10^3}}) \approx 2\Phi\left(\frac{c}{\sigma\sqrt{10^3}}\right) - 1$$

soll gleich 0,95 sein. ( $\sigma^2 =$  Streuung der Augenzahl eines Wurfes = 2,917.)

Daraus folgt  $2\Phi\left(\frac{c}{\sigma\sqrt{10^3}}\right) - 1 = 0,95$  also  $\Phi\left(\frac{c}{\sigma\sqrt{10^3}}\right) = 0,975$  und

$$c = 92,23 \cdot q_{0,975} = 180,8$$

.

Für 0,99 an Stelle 0,95 ergibt sich  $c = 237,5$

und für 1 an Stelle 0,95 erhalten wir  $c = 2500$ .

- c) Wie oft muss man einen Punkt rein zufällig aus dem Einheitsquadrat auswählen, um mit der in Abschnitt 10.4.1 beschriebenen Methode die Zahl  $\frac{\pi}{4}$  mit einer approximativen Wahrscheinlichkeit von 0,95 auf  $m$  Stellen genau zu bestimmen?

Antwort: Mit  $\alpha = 0,05$  gilt

$$n_0 = \left\lceil \frac{q_{1-\frac{\alpha}{2}}^2 \cdot \frac{\pi}{4} \left(1 - \frac{\pi}{4}\right)}{10^{-2m}} \right\rceil$$

$$n_0 = 0,65 \cdot 10^{2m}.$$

### 11.3 Der zentrale Grenzwertsatz von Lindeberg-Feller

Es seien im Weiteren  $(X_n, n \geq 1)$  eine Folge unabhängiger, aber nicht notwendig identisch verteilter Zufallsgrößen,  $S_n := \sum_{k=1}^n X_k$ . Die Verteilungsfunktion von  $X_n$  werde mit  $F_n$  bezeichnet.

Problem: Unter welchen Bedingungen gibt es Zahlenfolgen  $(a_n)$  und  $(b_n)$  mit  $b_n > 0$ , so dass die Verteilungen von  $\frac{S_n - a_n}{b_n}$  schwach (d. h. in Verteilung) gegen die Normalverteilung konvergieren?

Ohne weitere Voraussetzungen kann man Konvergenz gegen die Normalverteilung nicht erwarten.

**Beispiel 11.8** Alle  $X_n$  seien Cauchyverteilt mit dem Parameter  $a$ . Dann ist auch  $\frac{S_n}{n}$  Cauchyverteilt mit dem Parameter  $a$ . (Beweis mittels charakteristischer Funktionen.) Das heißt für  $a_n \equiv 0$  und  $b_n = n$  erhalten wir die Konvergenz von  $\frac{S_n - a_n}{b_n}$  für  $n \rightarrow \infty$ , aber nicht gegen die Normalverteilung.

Eine wesentliche Rolle bei der Lösung des oben gestellten Problems spielt der folgende Begriff.

**Definition 11.9** Man sagt, die Folge  $(X_n)$  erfüllt die Lindeberg-Bedingung (L), falls gilt  $D^2X_n < \infty, n \geq 1$  und falls

$$\lim_{n \rightarrow \infty} \frac{1}{D^2S_n} \sum_{k=1}^n \int_{\{x: |x-EX_k| \geq \varepsilon \sigma_n\}} |x-EX_k|^2 F_k(dx) = 0 \quad \forall \varepsilon > 0. \quad (L)$$

Dabei werde  $\sigma_n = \sqrt{D^2S_n}$  gesetzt.

Falls die Lindeberg-Bedingung (L) gilt, so folgt

$$\lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} \frac{D^2X_k}{D^2S_n} = 0. \quad (F)$$

Die Eigenschaft (F) wird auch als Feller-Bedingung bezeichnet.

Beweis: Es gilt

$$\frac{D^2X_k}{D^2S_n} \leq \varepsilon^2 + \frac{1}{D^2S_n} E \left[ (X_k - EX_k)^2 \mathbb{1}_{\{|X_k - EX_k| \geq \varepsilon \sigma_n\}} \right].$$

Daraus folgt für jedes  $\varepsilon > 0$ .

$$\max_{1 \leq k \leq n} \frac{D^2X_k}{D^2S_n} \leq \varepsilon^2 + \frac{1}{D^2S_n} \sum_{k=1}^n \left[ E(X_k - EX_k)^2 \mathbb{1}_{\{|X_k - EX_k| \geq \varepsilon \sigma_n\}} \right].$$

Aus (L) folgt nunmehr (F). □

Die Feller-Bedingung besagt anschaulich, dass jede der Streuungen  $D^2X_k, k = 1, \dots, n$ , für große  $n$  verschwindend klein ist im Vergleich zur Streuung  $D^2S_n$  der Summe  $X_1 + X_2 + \dots + X_n$ .

Aus der Feller-Eigenschaft (F) ergibt sich eine weitere Eigenschaft der Folge  $(X_n)$ , die man als "Asymptotische Kleinheit der  $X_{n,k} := \frac{X_k - EX_k}{\sigma_n}$ " bezeichnet:

$$\lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} P \left( \frac{|X_k - EX_k|}{\sigma_n} > \varepsilon \right) = 0. \quad (AK)$$

Der Beweis ergibt sich unmittelbar aus (F) mittels der Tschebyschev'schen Ungleichung:

$$P\left(\frac{|X_k - EX_k|}{\sigma_n} > \varepsilon\right) \leq \frac{D^2 X_k}{\sigma_n^2 \cdot \varepsilon^2}, \quad k = 1, \dots, n.$$

Nunmehr haben wir alle Begriffe, um folgenden Satz zu formulieren.

**Satz 11.10 (Zentraler Grenzwertsatz von Lindeberg-Feller)** *Es sei  $(X_n, n \geq 1)$  eine Folge unabhängiger Zufallsgrößen über  $(\Omega, \mathfrak{A}, P)$  mit  $0 < D^2 X_n < \infty$ . Dann sind folgende Aussagen äquivalent:*

- 1) *Die  $X_{n,k} = \frac{X_k - EX_k}{\sigma_n}, k = 1, \dots, n; n \geq 1$  mit  $\sigma_n^2 = D^2 S_n$  sind asymptotisch klein im Sinne von (AK) und es gilt*

$$\lim_{n \rightarrow \infty} \sup_{-\infty \leq a < b \leq \infty} \left| P\left(a < \frac{S_n - ES_n}{\sqrt{D^2 S_n}} \leq b\right) - (\Phi(b) - \Phi(a)) \right| = 0.$$

- b) *Die Lindeberg-Bedingung (L) gilt.*

Beweis: Siraev, (1988), Kap. III, 4.

### Beispiele 11.11

- a)  $(X_n)$  unabhängig,  $EX_n \equiv EX_1 = m, D^2 X_n \equiv D^2 X_1 = \sigma^2 \in (0, \infty)$ .

Dann ist die Lindeberg-Bedingung erfüllt, denn es gilt

$$\frac{1}{\sigma_n^2} \sum_1^n \int_{\{|x-m| \geq \sigma_n \cdot \varepsilon\}} |x-m|^2 dF_1(x) =$$

$$\frac{n}{n\sigma^2} \int_{\{|x-m| \geq \sqrt{n}\sigma\varepsilon\}} |x-m|^2 dF_1(x) \rightarrow 0$$

wegen  $P^{X_1}(\{|x-m| \geq \sqrt{n}\sigma\varepsilon\}) \leq \frac{D^2 X_1}{n\sigma^2 \cdot \varepsilon^2} \xrightarrow{n \rightarrow \infty} 0$

b) Für ein  $\delta > 0$  sei die folgende *Ljapunov-Bedingung* erfüllt:

$$\frac{1}{\sigma_n^{2+\delta}} \sum_1^n E|X_k - m_k|^{2+\delta} \xrightarrow{n \rightarrow \infty} 0 \quad (Ljap).$$

Dann gilt die Lindeberg-Bedingung (L).

Beweis: Für jedes  $\varepsilon > 0$  haben wir

$$\begin{aligned} |X_k - m_k|^{2+\delta} &= \int_{R_1} |x - m_k|^{2+\delta} dF_k(x) \geq \\ &\geq \int_{\{|x - m_k| \geq \varepsilon \sigma_n\}} |x - m_k|^{2+\delta} dF_k(x) \geq \varepsilon^\delta \sigma_n^\delta \int_{\{|x - m_k| \geq \varepsilon \sigma_n\}} |x - m_k|^2 dF_k(x) \\ &\implies \frac{1}{\sigma_n^2} \sum_1^n \int_{\{|x - m_k| \geq \varepsilon \sigma_n\}} (x - m_k)^2 dF_k(x) \leq \\ &\frac{1}{\varepsilon^\delta \sigma_n^{2+\delta}} \sum_{k=1}^n E|X_k - m_k|^{2+\delta} \rightarrow 0 \end{aligned}$$

Es gibt Folgen  $(X_n)$  unabhängiger Zufallsgrößen mit  $S_n^* = \frac{\sum_{k=1}^n (X_k - EX_k)}{\sqrt{D^2 S_n}} \xrightarrow{w} N(0, 1)$ , wo weder (L) gilt noch Asymptotische Kleinheit (AK) vorliegt:

Die Zufallsgrößen  $(X_n, n \geq 1)$  seien unabhängig und normalverteilt mit

$$EX_n \equiv 0, \quad D^2 X_1 = 1, \quad D^2 X_n = 2^{n-2}, \quad n \geq 2.$$

Dann ist die Streuung  $D^2 S_n$  von  $S_n = \sum_{k=1}^n X_k$  gleich  $\sum_{k=1}^n D^2 X_k = 2^{n-1}$ . Wir setzen wie üblich

$$S_n^* = \frac{1}{\sqrt{D^2 S_n}} \sum_{k=1}^n X_k, \quad n \geq 1.$$

Die Folge  $(X_n, n \geq 1)$  genügt nicht der Lindeberg-Bedingung, da insbesondere die Fellereigenschaft (F) nicht gilt:

$$\max_{k=1, \dots, n} \frac{D^2 X_k}{D^2 S_n} = \max_{k=1, \dots, n} \frac{2^{k-2}}{2^{n-1}} = \frac{1}{2}.$$

Außerdem sind die  $X_{n,k} := \frac{X_k}{\sqrt{D^2 S_n}}$ ;  $k = 1, \dots, n; n \geq 1$  nicht asymptotisch klein im Sinne von (AK), da für alle  $\varepsilon > 0$  und  $n \geq 1$  die Gleichung

$$\max_{k=1, \dots, n} P\left(\frac{|X_k|}{\sqrt{D^2 S_n}} > \varepsilon\right) = P\left(\frac{|X_n|}{\sqrt{2^{n-1}}} > \varepsilon\right) = 2(1 - \Phi(\varepsilon)) > 0$$

erfüllt ist.

Andererseits genügt  $(X_n, n \geq 1)$  trivialerweise dem zentralen Grenzwertsatz:

$S_n^*$  ist für jedes  $n \geq 1$  Standard-normalverteilt.



# Kapitel 12

## Elemente der Mathematischen Statistik

Beim Umgang mit zufälligen Erscheinungen ist es oft von Interesse, die Verteilungsfunktion  $F_X$  gewisser Zufallsgrößen  $X$  zu kennen. Daraus lassen sich Erwartungswert, Streuung, aber auch Wahrscheinlichkeiten der Form  $P(X > c)$  berechnen. Diese Verteilungsfunktion ist in vielen Fällen jedoch nicht bekannt. Beispielsweise sind für ein Versicherungsunternehmen, das die Haftpflicht für Autofahrer versichert, die Wahrscheinlichkeitsverteilung der Anzahl der Unfälle pro Jahr und Versicherungsnehmer oder die Verteilung der Schadenssumme pro Jahr und Versicherungsbestand Grundlagen für die Berechnung der Versicherungsprämie, die jeder Versicherungsnehmer im Jahr zu bezahlen hat.

Bekannt sind in vielen Fällen jedoch Daten, die Auskunft über die unbekanntes Verteilungsfunktionen geben können. So verfügen Versicherungsunternehmen über umfangreiche Datensammlungen zeitlich zurück liegender Schadensfälle. Sie betreffen sowohl Schadenshäufigkeiten in einem Versicherungsbestand als auch Schadenshöhen.

In der klassischen Statistik geht man meist davon aus, dass der zugrunde liegende Datensatz die mehrfache voneinander unabhängige Realisierung einer Zufallsgröße  $X$  mit einer Verteilungsfunktion  $F_X$  ist, er bildet eine sogenannte "Stichprobe". Die Mathematische Statistik konstruiert und bewertet Verfahren, um aus Stichproben Rückschlüsse auf  $F_X$  oder Kenngrößen von  $F_X$  zu ziehen.

Zentrale Fragestellungen sind dabei das Schätzen von Parametern der zugrun-

de liegenden Verteilung und das Testen von Hypothesen über diese Parameter.

Eine prinzipielle Möglichkeit, wie man zu der Verteilungsfunktion  $F_X$  kommt, eröffnet der folgende Hauptsatz der Mathematischen Statistik. Er besagt, dass man  $F_X$  auf der Grundlage von Stichproben prinzipiell beliebig genau bestimmen kann.

## 12.1 Der Hauptsatz der mathematischen Statistik

Es seien  $F$  eine Verteilungsfunktion auf  $R_1$  und  $X^{(n)} := (X_1, \dots, X_n)$  eine Folge unabhängiger identisch verteilter Zufallsgrößen über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  mit der Verteilungsfunktion  $F$ :

$$F(x) = P(X_k \leq x), \quad x \in R_1, \quad k = 1, \dots, n.$$

**Definition 12.1** *Man bezeichnet  $X^{(n)}$  mit diesen Eigenschaften auch als mathematische Stichprobe vom Umfang  $n$  aus einer nach  $F$  verteilten Grundgesamtheit. Realisiert man die Zufallsgrößen  $X_k, k = 1, \dots, n$ , so erhält man eine konkrete Stichprobe  $x^{(n)} := (x_1, \dots, x_n)$  vom Umfang  $n$  aus einer nach  $F$  verteilten Grundgesamtheit.*

**Beispiel 12.2** Es sei  $X^{(n)} = (X_1, X_2, \dots, X_n)$  ein Bernoullischema  $BS_n(p)$  mit  $p \in (0, 1)$ . Der konkrete Wert von  $p$  sei unbekannt. Dann ist  $X$  im obigen Sinne eine mathematische Stichprobe aus einer zweipunktverteilten Grundgesamtheit mit den möglichen Werten 1 und 0 und den entsprechenden Wahrscheinlichkeiten  $p$  bzw.  $1 - p$ . Jede Realisierung  $x^{(n)}$  von  $X^{(n)}$ , zum Beispiel für  $n = 5$

$$x^{(5)} = (0, 1, 1, 0, 1),$$

ist eine konkrete Stichprobe aus der erwähnten Grundgesamtheit.

Wir verbinden nun mit jeder Stichprobe eine neue Verteilungsfunktion.

Wir definieren

$$\hat{F}_n(x; X^{(n)}) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(-\infty, x]}(X_i), \quad x \in R_1. \quad (12.1)$$

Die Funktion  $\hat{F}_n(\cdot)$  (das Argument  $X^{(n)}$  wird meist weggelassen) ist eine vom Zufall abhängige Verteilungsfunktion, die sogenannte "empirische Verteilungsfunktion der mathematischen Stichprobe  $X^{(n)} = (X_1, \dots, X_n)$ ".

Da zu jeder Verteilungsfunktion  $F$  auf  $R_1$  ein Wahrscheinlichkeitsmaß  $Q_F$  auf  $\mathfrak{B}_1$  gehört, das  $F$  als seine Verteilungsfunktion besitzt, ist das auch für  $\hat{F}_n$  der Fall.  $Q_{\hat{F}_n}$  ist ein vom Zufall abhängiges diskretes Wahrscheinlichkeitsmaß und ordnet jedem Punkt  $\{X_i(\omega)\}, i = 1, \dots, n$ , die Wahrscheinlichkeit

$$Q_{\hat{F}_n}(\{X_i(\omega)\}) = \frac{1}{n} \times \# \{j \in \{1, 2, \dots, n\} \text{ mit } X_j(\omega) = X_i(\omega)\}$$

zu.

Setzt man in (12.1) anstelle  $X^{(n)}$  eine Realisierung  $x^{(n)}$ , also eine konkrete Stichprobe, ein, so erhält man eine nichtzufällige Verteilungsfunktion

$$\hat{F}_n(x; x^{(n)}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(-\infty, x]}(x_i), \quad x \in R_1. \quad (12.2)$$

Die dazu gehörende Wahrscheinlichkeitsverteilung ist die diskrete gleichmäßige Verteilung  $Q_{\hat{F}_n}$  auf den Zahlen  $\{x_1, \dots, x_n\}$  mit

$$Q_{\hat{F}_n}(\{x_k\}) = \frac{1}{n} \times \#\{j \in \{1, \dots, n\} : x_j = x_k\}.$$

Für festes  $x \in R_1$  ist  $\hat{F}_n(x; X^{(n)})$  die (zufällige) relative Häufigkeit, mit der die  $\{X_k \leq x\}, k = 1, \dots, n$  eintreten. Es gilt

$$E\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n P(X_i \leq x) = F(x)$$

$$D^2(\hat{F}_n(x)) = \frac{F(x)(1 - F(x))}{n}.$$

Aus dem starken Gesetz der großen Zahlen folgt für jedes  $x \in R_1$

$$\lim_{n \rightarrow \infty} \hat{F}_n(x) = F(x) \quad P - f.s.$$

Darüber hinaus gilt der

**Satz 12.3 (Hauptsatz der mathematischen Statistik)**

Es seien  $F$  eine Verteilungsfunktion auf  $R_1$  und  $X^{(n)} = (X_1, X_2, \dots, X_n)$  eine mathematische Stichprobe aus einer nach  $F$  verteilten Grundgesamtheit.  $X^{(n)}$  sei definiert auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$ .

Für die Zufallsgrößen  $D_n, n \geq 1$ , definiert durch

$$D_n := \sup_{x \in R_1} |\hat{F}_n(x) - F(x)|, \quad (12.3)$$

gilt

$$\lim_{n \rightarrow \infty} D_n = 0 \quad P - f.s. \quad (12.4)$$

Beweis: Es seien  $N$  und  $j$  natürliche Zahlen mit  $0 \leq j \leq N$ ,

$$x_{j,N} := \inf\{x : F(x) \geq \frac{j}{N}\}, \quad x_{0,N} := -\infty, \quad \inf \emptyset := \infty.$$

Ist  $y < x_{j,N}$ , so folgt  $F(y) < \frac{j}{N}$ ,

und es gilt (wegen der Rechtsstetigkeit von  $F$ )

$$F(x_{j,N} - 0) \leq \frac{j}{N} \leq F(x_{j,N}).$$

Daraus ergibt sich für  $0 \leq j < N$ .

$$F(x_{j+1,N} - 0) \leq \frac{j+1}{N} \leq F(x_{j,N}) + \frac{1}{N} \quad (12.5)$$

Ist nun  $x \in [x_{j,N}, x_{j+1,N})$ , so erhalten wir wegen (12.5) und

$F(x) \leq F(x_{j+1,N} - 0)$  die Ungleichung

$$\hat{F}_n(x_{j,N}) - F(x_{j,N}) - \frac{1}{N} \leq \hat{F}_n(x) - F(x) \leq \hat{F}_n(x_{j+1,N} - 0) - F(x_{j,N}) \leq$$

$$\hat{F}_n(x_{j+1,N} - 0) - F(x_{j+1,N} - 0) + \frac{1}{N} \quad P - f.s.$$

Daraus ergibt sich für alle  $x \in \bigcup_{i=0}^{N-1} [x_{i,N}, x_{i+1,N}) = [-\infty, x_{N,N})$  und alle  $x$  mit  $x \geq x_{N,N}$

$$|\hat{F}_n(x) - F(x)| \leq \max_{0 \leq j < N} \{|\hat{F}_n(x_{j,N}) - F(x_{j,N})|, |\hat{F}_n(x_{j+1,N} - 0) - F(x_{j+1,N} - 0)|\}$$

$$+ \frac{1}{N} \quad P - f.s.$$

Aus dem starken Gesetz der großen Zahlen folgen für jedes  $j$  mit  $0 \leq j < N$  die Gleichungen  $\lim_{n \rightarrow \infty} \hat{F}_n(x_{j,N}) = F(x_{j,N})$  und  $\lim_{n \rightarrow \infty} \hat{F}_n(x_{j+1,N} - 0) = F(x_{j+1,N} - 0)$   $P$ -fast sicher.

Deshalb gilt:

$$D_n = \sup_{x \in R_1} |\hat{F}_n(x) - F(x)| \xrightarrow{n \rightarrow \infty} 0 \quad P - \text{fast sicher.}$$

□

Der Hauptsatz der mathematischen Statistik ist von grundlegender Bedeutung für die praktische Anwendung der Wahrscheinlichkeitstheorie. Er besagt, dass man eine unbekannte Verteilungsfunktion  $F$  grundsätzlich beliebig genau bestimmen kann, wenn man sich eine hinreichend große konkrete Stichprobe  $x^{(n)} = (x_1, \dots, x_n)$  aus einer nach  $F$  verteilten Grundgesamtheit verschafft und

$$\hat{F}_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{(-\infty, x]}(x_k), \quad x \in R_1$$

als Näherung für  $F(\cdot)$  verwendet.

Als eine Verfeinerung des Hauptsatzes im Falle, dass  $F$  stetig ist, geben wir noch folgende Aussage an.

**Aussage 12.4** *Ist  $F$  stetig, so gilt*

$$\lim_{n \rightarrow \infty} P(\sqrt{n}D_n \leq x) = K(x), \quad x \in R_1$$

mit

$$K(x) := \begin{cases} 0 & x \leq 0 \\ \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 x^2}, & x > 0. \end{cases}$$

( $K(\cdot)$  ist die Verteilungsfunktion der sogenannten Kolmogorov-Smirnov-Verteilung.)

Zum Beweis sei auf Winkler (1983) verwiesen. Für große  $n$  und für alle  $y > 0$  kann man also  $P(D_n \leq y)$  annähernd durch  $K(\sqrt{ny})$  ersetzen:

$$P(D_n \leq y) \approx K(\sqrt{ny}).$$

Wir haben gesehen, dass man prinzipiell auf der Grundlage von Stichproben die Verteilungsfunktion  $F_X$  einer Zufallsgröße  $X$  beliebig genau bestimmen kann. In praktischen Fällen wird dieses Verfahren jedoch selten angewandt. Vielfach hat man nämlich Vorabinformationen über  $F_X$  in dem Sinne, dass man weiß, dass  $F_X$  zu einer gewissen Klasse von Verteilungsfunktionen gehört. Zum Beispiel könnte aus inhaltlichen Gründen unter Verwendung eines zentralen Grenzwertsatzes geschlossen werden, dass  $F_X$  die Verteilungsfunktion einer Normalverteilung ist. Dann wären nur noch die Parameter  $\mu$  und  $\sigma^2$  zu bestimmen. Oder bei der Anzahl der Schäden, die ein Versicherungsnehmer pro Jahr verursacht, scheint in erster Näherung eine Poissonverteilung geeignet zu sein (Begründung?). Dann wäre nur noch ihr Parameter  $\lambda$  unbekannt. In vielen Fällen interessiert man sich auch nur für gewisse Kenngrößen der Verteilung, zum Beispiel für den Erwartungswert und/oder für die Streuung.

Die Konstruktion und Beurteilung von Verfahren zur näherungsweise Bestimmung von unbekanntem Parametern auf der Grundlage von Stichproben ist

Aufgabe der sogenannten statistischen Schätztheorie, aus der wir im folgenden Abschnitt einige grundlegende Begriffe und Aussagen kennen lernen.

## 12.2 Statistische Schätzungen

### 12.2.1 Definitionen

**Definition 12.5** *Es sei  $\mathfrak{P} = (P_\vartheta, \vartheta \in \Theta)$ ,  $\Theta \subseteq R_k$ ,  $k \geq 1$ , eine Familie von Wahrscheinlichkeitsmaßen auf  $(\Omega, \mathfrak{A})$ . Dann heißt  $(\Omega, \mathfrak{A}, \mathfrak{P})$  ein statistisches Modell.*

Für  $\Theta$  wählt man irgendeine nichtleere Menge, meist eine offene oder abgeschlossene Menge. Angenommen,  $X$  ist eine reellwertige Zufallsgröße über  $(\Omega, \mathfrak{A})$  und  $P_\vartheta^X$  die zu  $X$  gehörende Verteilung unter  $P_\vartheta$ :

$$P_\vartheta^X(B) := P_\vartheta(X \in B), \quad B \in \mathfrak{B}_1, \quad \vartheta \in \Theta.$$

Offenbar ist dann  $(R_1, \mathfrak{B}_1, \mathfrak{P}^X)$  mit  $\mathfrak{P}^X = (P_\vartheta^X, \vartheta \in \Theta)$  ebenfalls ein statistisches Modell. Den Erwartungswert von  $X$  oder irgendeiner anderen Zufallsgröße  $Y$  bezüglich der Verteilung  $P_\vartheta$  bezeichnen wir mit  $E_\vartheta X$  bzw.  $E_\vartheta Y$ .

Anschaulicher Hintergrund: Wir nehmen an, dass die Verteilung von  $X$  zu  $\mathfrak{P}^X$  gehört, kennen aber den wahren Wert  $\vartheta_0$  des Parameters  $\vartheta$  nicht.

Es sei  $X^{(n)} = (X_1, X_2, \dots, X_n)$  eine mathematische Stichprobe aus einer nach  $P_\vartheta, \vartheta \in \Theta$ , verteilten Grundgesamtheit.

Aufgabe: Man konstruiere auf der Grundlage einer Stichprobe eine Schätzung für den wahren Wert  $\vartheta_0$ .

Häufig ist man gar nicht an  $\vartheta$  selbst, sondern an einer gewissen Funktion von  $\vartheta$  interessiert, zum Beispiel am Erwartungswert  $\frac{1}{\lambda}$  einer  $Exp(\lambda)$ -verteilten Zufallsgröße.

Wir formulieren den Begriff der Schätzung deshalb zunächst einmal sehr allgemein. Auf Gütekriterien für Schätzungen gehen wir anschließend ein.

**Definition 12.6** *Es seien  $g$  und  $\hat{G}_n$  Borelmessbare Funktionen von  $\Theta$  bzw. von  $R_n$  in  $R_m$ . Überdies sei  $\vartheta \in \Theta$ . Dann heißt  $\hat{G}_n(X_1, \dots, X_n)$  eine Schätzung für*

$g(\vartheta)$ .

Durch Einsatz einer konkreten Stichprobe  $x^{(n)} = (x_1, \dots, x_n)$  in  $\hat{G}_n$  erhält man einen Schätzwert  $\hat{G}_n(x_1, x_2, \dots, x_n)$  für  $g(\vartheta)$ .

**Beispiel 12.7** Es sei  $X$  eine Zufallsgröße mit den möglichen Werten  $1, 2, \dots, N$  mit

$$P_N(X = k) = \frac{1}{N}, k = 1, 2, \dots, N.$$

Der Parameter  $N$  sei unbekannt. Als Schätzung für  $N$  auf der Grundlage von  $X^{(n)} = (X_1, \dots, X_n)$  hat man zum Beispiel

$$\hat{N}_n = \max_{k=1,2,\dots,n} X_k \quad \text{und} \quad \tilde{N}_n = \left[ \frac{2}{n} \sum_{k=1}^n X_k \right].$$

## 12.2.2 Güteeigenschaften von Schätzungen

Wir verwenden die Terminologie des vorangegangenen Abschnittes.

Im Allgemeinen gibt es viele Schätzungen  $\hat{G}_n(X_1, \dots, X_n)$  für  $g(\vartheta)$ . Bei der Frage, welche Kriterien man bei der Auswahl anlegen sollte, bietet sich zuerst die Eigenschaft der *Erwartungstreue* an.

**Definition 12.8** Die Schätzung  $\hat{G}_n(X_1, \dots, X_n)$  für  $g(\vartheta)$  heißt erwartungstreu, falls gilt

$$E_{\vartheta} \hat{G}_n(X_1, \dots, X_n) = g(\vartheta) \quad \text{für alle } \vartheta \in \Theta.$$

Ist  $\hat{G}_n(X^{(n)})$  irgendeine Schätzung für  $g(\vartheta)$ ,  $\vartheta \in \Theta$ , so nennt man die Funktion

$$E_{\vartheta} \hat{G}_n(X^{(n)}) - g(\vartheta), \quad \vartheta \in \Theta$$

die Verzerrung der Schätzung, ihren systematischen Fehler oder ihren Bias. Eine erwartungstreue Schätzung heißt auch unverzerrt oder unbiased.

Erwartungstreue Schätzungen haben die Eigenschaft, dass sich ihre Werte bei häufiger (unabhängiger) Wiederholung der Schätzung um den Erwartungswert, also  $g(\vartheta)$ , gruppieren (Gesetz der großen Zahlen). Man kann also ein gewisses Vertrauen haben, dass die entsprechenden Schätzwerte in der Nähe des zu schätzenden Wertes  $g(\vartheta)$  liegen.

### Beispiele 12.9

1. Der Erwartungswert  $\mu(\vartheta) := E_{\vartheta}X_1$  sei unbekannt. Dann ist für jeden Vektor  $a = (a_1, \dots, a_n)$  mit  $a_k \geq 0, k = 1, \dots, n$ , und  $\sum_{k=1}^n a_k = 1$

die Schätzung

$$\hat{\mu}_{(a)}(X^{(n)}) := \sum_{k=1}^n a_k X_k$$

eine erwartungstreue Schätzung für  $\mu(\vartheta)$ .

Spezialfälle sind  $\hat{\mu}_n := \frac{1}{n} \sum_{k=1}^n X_k$  und  $\hat{\mu}_1 := X_1$ .

- 2.

$$\hat{\sigma}_n^2(X^{(n)}) = \frac{1}{n} \sum_{k=1}^n (X_k - \hat{\mu}_n)^2$$

ist keine erwartungstreue Schätzung für  $\sigma^2(\vartheta) = D_{\vartheta}^2 X_1$ .

Es gilt nämlich  $E_{\vartheta} \hat{\sigma}_n^2(X^{(n)}) = \frac{n-1}{n} \sigma^2(\vartheta)$ .

Ihr Bias ist

$$E_{\vartheta} \hat{\sigma}_n^2(X^{(n)}) - \sigma^2(\vartheta) = \frac{n-1}{n} \sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}.$$

Die Streuung  $\sigma^2$  wird also bei häufiger Schätzung durch  $\hat{\sigma}_n^2$  systematisch unterschätzt. Dagegen ist

$$\tilde{\sigma}_n^2(X^{(n)}) = \frac{1}{n-1} \sum_{k=1}^n (X_k - \hat{\mu}_n)^2$$

eine erwartungstreue Schätzung für  $\sigma^2$ .

Wie wir am Beispiel 12.9(1) gesehen haben, gibt es mitunter mehrere erwartungstreue Schätzungen für  $g(\vartheta)$ . Um unter ihnen eine Auswahl zu treffen, führen wir ein weiteres Gütekriterium ein.

**Definition 12.10** Sind  $\hat{G}(X^{(n)})$  und  $G^*(X^{(n)})$  zwei erwartungstreue Schätzungen für  $g(\vartheta)$ ,  $\vartheta \in \Theta$ , so heißt  $\hat{G}(X^{(n)})$  besser als  $G^*(X^{(n)})$ , falls

$$D_{\vartheta}^2 \hat{G}(X^{(n)}) \leq D_{\vartheta}^2 G^*(X^{(n)}) \quad \text{für alle } \vartheta \in \Theta \quad (12.6)$$

gilt.  $\hat{G}(X_n)$  heißt beste erwartungstreue Schätzung für  $g(\vartheta)$ ,  $\vartheta \in \Theta$ , oder erwartungstreue Schätzung mit minimaler Streuung, falls (6) für jede erwartungstreue Schätzung  $G^*(X^{(n)})$  für  $g(\vartheta)$ ,  $\vartheta \in \Theta$ , gilt.

**Beispiel 12.11** (Fortsetzung des Beispiels 12.9(1)):

Es gilt  $D_{\vartheta}^2(\hat{\mu}_{(a)}(X^{(n)})) = \sigma^2(\vartheta) \sum_{k=1}^n a_k^2$ , und dieser Ausdruck wird minimal (unter der Nebenbedingung  $a_k \geq 0$ ,  $\sum a_k = 1$ ) für  $a_k \equiv \frac{1}{n}$ . Das arithmetische Mittel  $\hat{\mu}_n(X^{(n)})$  ist also unter allen gewichteten Mitteln  $\hat{\mu}_{(a)}(X^{(n)})$  die beste erwartungstreue Schätzung für  $\mu(\vartheta)$ .

Die Definition bester erwartungstreuer Schätzungen wirft die Frage auf nach der Existenz solcher Schätzungen und gegebenenfalls nach der Größe ihrer Streuung.

Ein Ergebnis in dieser Richtung ist die sogenannte *Ungleichung von Cramer-Rao*. Bevor wir auf sie eingehen, stellen wir noch einige Begriffe bereit.

## Die Likelihoodfunktion

Es sei  $X^{(n)} = (X_1, X_2, \dots, X_n)$  eine mathematische Stichprobe aus einer nach  $P_{\vartheta}^X$ ,  $\vartheta \in \Theta$ , verteilten Grundgesamtheit, wobei  $X$  eine reellwertige Zufallsgröße

ist. Die Verteilung  $P_{\vartheta}^{X^{(n)}}$  ist also für jedes  $\vartheta \in \Theta$  eine Verteilung auf  $(R_n, \mathfrak{B}_n)$  mit

$$P_{\vartheta}^{X^{(n)}}(B_1 \times \dots \times B_n) = \prod_{k=1}^n P_{\vartheta}^{X}(B_k), \quad B_1, \dots, B_n \in \mathfrak{B}_1. \quad (12.7)$$

Um die sogenannte Likelihoodfunktion definieren zu können, unterscheiden wir zwei Fälle.

1. Fall:  $X$  besitzt für alle  $\vartheta \in \Theta$  eine Dichte  $f_{\vartheta}(\cdot)$  bezüglich des Lebesguemaßes.

In diesem Fall setzen wir

$$L^X(\vartheta, x) := f_{\vartheta}(x), \quad \vartheta \in \Theta, x \in R_1.$$

Es gilt nach Definition der Dichte

$$P_{\vartheta}^X(B) = \int_B L^X(\vartheta, x) dx, \quad \vartheta \in \Theta, B \in \mathfrak{B}_1.$$

2. Fall:  $X$  sei diskret verteilt unter  $P_{\vartheta}$  mit den möglichen Werten  $a_k, k \in N_0$ , die nicht von  $\vartheta$  abhängen. In diesem Fall sei

$$L^X(\vartheta, a_k) := P_{\vartheta}(X = a_k), \quad k \in N_0.$$

Es gilt dann

$$P_{\vartheta}^X(B) = \sum_{a_k \in B} L^X(\vartheta, a_k).$$

Offenbar gilt in beiden Fällen

$$L^X(\vartheta, \cdot) \geq 0 \text{ und}$$

$$P_{\vartheta}^X(\{x : L^X(\vartheta, x) = 0\}) = 0. \quad (12.8)$$

Ist  $H$  im ersten Fall eine messbare nichtnegative Funktion auf  $R_n$ , so gilt

$$E_{\vartheta}H(X) = \int_{\mathbb{R}_1} H(x)P_{\vartheta}^X(dx) = \int_{\mathbb{R}_1} H(x)L^X(\vartheta, x)dx, \quad (12.9)$$

und ist  $H$  im zweiten Fall eine nichtnegative Funktion auf  $A$ , so haben wir

$$E_{\vartheta}H(X) = \sum_{a_k \in A} H(a_k)p_{\vartheta}(a_k) = \sum_{a_k \in A} H(a_k)L^X(\vartheta, a_k). \quad (12.10)$$

**Definition 12.12** Wir setzen voraus, es liegt der 1. oder 2. der eben eingeführten Fälle vor.

Für jedes  $x^{(n)} = (x_1, x_2, \dots, x_n) \in R_n$  heisst die Funktion

$$\vartheta \rightarrow L_n(\vartheta; x^{(n)}) = \prod_{k=1}^n L^X(\vartheta, x_k), \quad \vartheta \in \Theta,$$

Likelihoodfunktion des statistischen Modells

$\mathfrak{P}^X = (P_{\vartheta}^X, \vartheta \in \Theta)$  (bei gegebener konkreter Stichprobe  $x^{(n)}$ ).

**Bemerkung 12.13** Mit Hilfe der Likelihoodfunktion kann man die gemeinsame Verteilung von  $X^{(n)} = (X_1, X_2, \dots, X_n)$  ausdrücken (beachte die Schreibweise  $x^{(n)} = (x_1, x_2, \dots, x_n)$ ):

$$P_{\vartheta}^{X^{(n)}}(B_1, \times \dots \times B_n) = \int_{B_1} \dots \int_{B_n} L_n(\vartheta, x^{(n)}) dx_1 \dots dx_n$$

im ersten Fall und

$$P_{\vartheta}^{X^{(n)}}(B_1, \times \dots \times B_n) = \sum_{x_1 \in A} \dots \sum_{x_n \in A} L_n(\vartheta, x^{(n)}).$$

im zweiten Fall.

Offenbar gilt im ersten Fall für alle nichtnegativen messbaren  $H$

$$E_{\vartheta}H(X^{(n)}) = \int \cdots \int_{R_n} H(x^{(n)})L_n(\vartheta, x^{(n)})dx_1 \dots, dx_n$$

und im zweiten Fall für alle nichtnegativen Funktionen  $H$

$$E_{\vartheta}H(X^{(n)}) = \sum_{x^{(n)} \in A^n} H(x^{(n)})L_n(\vartheta, x^{(n)}).$$

### Beispiele 12.14

a) Es sei  $X \sim N(\mu, \sigma^2)$ . Dann gilt mit  $\vartheta = (\mu, \sigma^2)^T \in R_1 \times R_+ =: \Theta$

$$L_n(\vartheta; x^{(n)}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - \mu)^2 \right] =$$

$$(\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{k=1}^n x_k^2 + \frac{\mu}{\sigma^2} \sum_{k=1}^n x_k - \frac{n\mu^2}{2\sigma^2} \right] (2\pi)^{-\frac{n}{2}}, x^{(n)} \in R_n.$$

b) Es sei  $X \sim \text{Bin}(m, p)$ . Dann ist mit  $\vartheta = p \in (0, 1) = \Theta$

$$L_n(\vartheta; x^{(n)}) = \prod_{k=1}^n \binom{m}{i_k} p^{i_k} (1-p)^{m-i_k}$$

$$x^{(n)} = (i_1, i_2, \dots, i_n), 0 \leq i_k \leq m, k = 1, \dots, n.$$

**Aussage 12.15 (Cramer-Rao-Ungleichung)** *Es sei vorausgesetzt:*

a) Die Likelihoodfunktion  $\vartheta \rightarrow L_n(\vartheta, x^{(n)})$  ist für jedes  $x^{(n)}$  differenzierbar bezüglich  $\vartheta$ ,  $\text{grad} \ln L_n(\vartheta, X^{(n)})$  ist ein zentrierter zufälliger Vektor und alle seine zweiten Momente bez.  $P_{\vartheta}$  sind endlich ( $\vartheta \in \Theta \subseteq R_k$ ).

$$(\text{grad} = \text{grad}_{\vartheta} = \left( \frac{\partial}{\partial \vartheta_1}, \frac{\partial}{\partial \vartheta_2}, \dots, \frac{\partial}{\partial \vartheta_k} \right)^T)$$

b) Für jede reellwertige Borelmeßbare Funktion  $h$  mit  $E_{\vartheta}|h(X^{(n)})|^2 < \infty$  gilt im ersten Fall

$$\text{grad} \int_{R_n} L_n(\vartheta; x^{(n)}) h(x^{(n)}) dx^{(n)} = \int_{R_n} \text{grad} L_n(\vartheta, x^{(n)}) h(x^{(n)}) dx^{(n)}$$

und im zweiten Fall

$$\text{grad} \sum_{x^{(n)} \in A^n} L_n(\vartheta, x^{(n)}) h(x^{(n)}) p_{\vartheta}(x^{(n)}) = \sum_{x^{(n)} \in A^n} \text{grad} L_n(\vartheta, x^{(n)}) h(x^{(n)}).$$

c) Die Matrix  $I_n(\vartheta)$ , definiert durch

$$(I_n(\vartheta))_{1 \leq i, j \leq k} = E_{\vartheta} \left( \frac{\partial}{\partial \vartheta_i} \ln L_n(\vartheta, X^{(n)}) \cdot \frac{\partial}{\partial \vartheta_j} \ln L_n(\vartheta, X^{(n)}) \right)_{1 \leq i, j \leq k}$$

ist invertierbar für jedes  $\vartheta \in \Theta$ .

(Es gilt  $I_n(\vartheta) = E_{\vartheta}(\text{grad} \ln L_n(\vartheta, X^{(n)}) \text{grad}^T \ln L_n(\vartheta; X^{(n)}))$ .)

Dann gilt für jede reellwertige Zufallsgröße  $Y$  der Form  $Y = h(X_1, X_2, \dots, X_n)$

$$E_{\vartheta}(Y - E_{\vartheta}Y)^2 \geq (\text{grad} E_{\vartheta}Y)^T [I_n(\vartheta)]^{-1} (\text{grad} E_{\vartheta}Y).$$

Ist insbesondere  $Y$  eine erwartungstreue Schätzung für  $g(\vartheta)$ , so gilt

$$E_{\vartheta}(Y - E_{\vartheta}Y)^2 \geq (\text{grad} g(\vartheta))^T [I_n(\vartheta)]^{-1} (\text{grad} g(\vartheta))$$

und für  $k = 1$  erhalten wir:

$$D_{\vartheta}^2 Y \geq \frac{[g'(\vartheta)]^2}{I_n(\vartheta)}.$$

**Definition 12.16** Die Matrix  $I_n(\vartheta)$  heißt Fisher'sche Informationsmatrix.

Sie ist nichtnegativ definit, da sie die Kovarianzmatrix des Vektors  $\text{grad } \ln L_n(\vartheta, X^{(n)})$  ist.

Die Matrix  $I_n(\vartheta)$  lässt sich durch  $I_1(\vartheta)$  ausdrücken.  
Es gilt nämlich wegen

$$\ln L_n(\vartheta, X^{(n)}) = \sum_{k,l=1}^n \ln L^X(\vartheta, X_k)$$

die Beziehung

$$\begin{aligned} (I_n(\vartheta))_{ij} &= E_\vartheta \left( \sum_{k,l=1}^n \frac{\partial}{\partial \vartheta_i} \ln L^X(\vartheta, X_k) \cdot \frac{\partial}{\partial \vartheta_j} \ln L^X(\vartheta, X_l) \right) = \\ &= E_\vartheta \left( \sum_{k=1}^n \frac{\partial}{\partial \vartheta_j} \ln L^X(\vartheta, X_k) \frac{\partial}{\partial \vartheta_i} \ln L^X(\vartheta, X_k) \right) \\ &= n(I_1(\vartheta))_{ij}. \end{aligned}$$

Beweis der Aussage 12.15: (Anstelle  $L_n$  schreiben wir hier auch kurz  $L$ .) Wir beschränken uns auf den ersten Fall. Der zweite wird völlig analog bewiesen. Aus der Voraussetzung b) folgt für  $h \equiv 1$ , dass

$$\int_{R_n} \text{grad } L(\vartheta, x^{(n)}) dx^{(n)} = 0$$

und damit haben wir

$$E_\vartheta [\text{grad } \ln L(\vartheta, X^{(n)})] = E_\vartheta \left[ \frac{\text{grad } L(\vartheta, X^{(n)})}{L(\vartheta, X^{(n)})} \right] = 0.$$

Weiterhin folgt damit aus b), falls  $\int_{R_n} h^2(x^{(n)}) dx^{(n)} < \infty$  gilt,

$$\text{grad } E_\vartheta h(X^{(n)}) = \text{grad } \int_{R_n} L(\vartheta, x^{(n)}) h(x^{(n)}) dx^{(n)} =$$

$$E_\vartheta [\text{grad } \ln L(\vartheta, X^{(n)}) \cdot h(X^{(n)})] =$$

$$E_{\vartheta} \left[ \text{grad } \ln L(\vartheta, X^{(n)}) (h(X^{(n)}) - E_{\vartheta} h(X^{(n)})) \right].$$

Es sei nun  $u \in R^k \setminus \{0\}$ . Dann gilt ( $\langle \cdot, \cdot \rangle$  bezeichnet das Skalarprodukt):

$$\langle u, \text{grad } E_{\vartheta} h(X^{(n)}) \rangle =$$

$$E_{\vartheta} [\langle u, \text{grad } \ln L(\vartheta, X^{(n)}) \rangle (h(X^{(n)}) - E_{\vartheta} h(X^{(n)}))].$$

Mittels der Schwarz'schen Ungleichung ergibt sich

$$E_{\vartheta} (h(X^{(n)}) - E_{\vartheta} h(X^{(n)}))^2 \geq \frac{\langle u, \text{grad } E_{\vartheta} h(X^{(n)}) \rangle^2}{E_{\vartheta} [\langle u, \text{grad } \ln L(\vartheta) \rangle^2]}$$

für alle  $u \in R_k \setminus \{0\}$ . Wir bestimmen den maximalen Wert der rechten Seite dieser Ungleichung für  $u \in R_k \setminus \{0\}$ .

Es sei  $\vartheta \in \Theta$  fest gewählt und  $u$  so normiert, dass gilt

$$\langle u, \text{grad } E_{\vartheta} h(X^{(n)}) \rangle = 1.$$

Man beachte in der Schreibweise des Skalarproduktes  $\langle u, v \rangle = u^T v$ :

$$\begin{aligned} \langle u, \text{grad } \ln L(\vartheta) \rangle^2 &= (u^T \text{grad } \ln L(\vartheta))^2 = \\ &= (u^T \text{grad } \ln L(\vartheta)) ((\text{grad } \ln L(\vartheta))^T u). \end{aligned}$$

Somit gilt

$$E_{\vartheta} \langle u, \text{grad } \ln L(\vartheta) \rangle^2 = u^T I_n(\vartheta) u.$$

Wir definieren:

$$\nu = \text{grad } E_{\vartheta} h(X^{(n)})$$

und haben folglich die quadratische Form

$$u^T I_n(\vartheta) u$$

unter der Nebenbedingung  $\langle u, \nu \rangle = 1$  zu minimieren.

Mittels der Methode des Lagrange'schen Multiplikators folgt als notwendige Bedingung

$$2I_n(\vartheta)u = \lambda\nu.$$

Nach Voraussetzung c) ergeben sich  $\langle u, \nu \rangle = 1$  und  $u = \frac{\lambda}{2} I_n^{-1}(\vartheta)\nu$  als notwendige Bedingungen. Daraus folgt

$$\begin{aligned} 1 = \langle u, \nu \rangle &= \frac{\lambda}{2} \nu^T I_n^{-1}(\vartheta)\nu \text{ und} \\ u^T I_n(\vartheta)u &= \frac{\lambda^2}{4} \nu^T I_n^{-1}(\vartheta)I(\vartheta)I_n^{-1}(\vartheta)\nu = \\ &= \frac{\lambda^2}{4} \nu^T I_n^{-1}(\vartheta)\nu = \frac{1}{\nu^T I_n^{-1}(\vartheta)\nu}. \end{aligned}$$

Somit ergibt sich für diese Wahl von  $u$

$$E_{\vartheta}(h(X^{(n)}) - E_{\vartheta}h(X^{(n)}))^2 \geq (\text{grad } E_{\vartheta}h(X^{(n)}))^T I_n^{-1}(\vartheta)(\text{grad } E_{\vartheta}h(X^{(n)})).$$

□

**Definition 12.17** Jede erwartungstreue Schätzung  $\hat{G}_n(X_n^{(n)})$  für  $g(\vartheta)$ , für die  $D_{(\vartheta)}^2 \hat{G}_n(X_n^{(n)})$  gleich der unteren Schranke in der Cramer-Rao-Ungleichung ist, heißt eine effiziente Schätzung für  $g(\vartheta)$ ,  $\vartheta \in \Theta$ .

Effiziente Schätzungen sind offenbar beste erwartungstreue Schätzungen. Die Umkehrung gilt im Allgemeinen nicht.

**Beispiel 12.18 (Effiziente Schätzung)** Ist  $X$  eine Zufallsgröße mit  $P_p(X = 1) = p$ ,  $P_p(X = 0) = 1 - p = q$ ,  $p \in (0, 1)$  unbekannt, und ist  $X^{(n)}$  eine mathematische Stichprobe aus einer wie  $X$  verteilten Grundgesamtheit, so gilt

$$L_n(\vartheta; x^{(n)}) = p^{\sum x_i} q^{n - \sum x_i}, \quad x^{(n)} = (x_1, \dots, x_n) \in \{0, 1\}^n,$$

und folglich ist

$$\begin{aligned} \ln L_n(\vartheta; x^{(n)}) &= \sum x_l \ln p + (n - \sum x_l) \ln(1 - p) = \\ &= s_n \ln p + (n - s_n) \ln(1 - p), \text{ mit } s_n = \sum_{l=1}^n x_l. \end{aligned}$$

Daraus folgt

$$E_p \left( \frac{d}{dp} \ln L_n(X^{(n)}) \right)^2 = \frac{n}{p(1-p)} = I_n(\vartheta), \quad I_1(\vartheta) = [p(1-p)]^{-1}.$$

Setzen wir  $g(p) = p$ , so erhalten wir mit  $S_n = \sum_{l=1}^n X_l$  für die erwartungstreue Schätzung  $\hat{p}_n(X^{(n)}) := \frac{S_n}{n}$  für den Parameter  $p$  die Streuung:

$$D_p^2 \left( \frac{S_n}{n} \right) = \frac{p(1-p)}{n} = I_n^{-1}(p).$$

Also ist  $\frac{S_n}{n}$  eine effiziente Schätzung für  $p$ .

Die gleichfalls erwartungstreue Schätzung  $\hat{G}_n(X^{(n)}) = X_1$  für  $p$  zum Beispiel hat dagegen eine wesentlich größere Streuung, nämlich  $p(1-p)$ .

### 12.2.3 Konstruktion von Schätzungen

Wir haben bisher Eigenschaften von Schätzungen angegeben und einige plausible Schätzungen kennen gelernt. Im Folgenden gehen wir auf zwei Methoden ein, Schätzungen zu konstruieren, die *Momentenmethode* und die *Maximum-Likelihood-Methode*. Keine dieser Methoden liefert universell beste Lösungen. Die mit ihrer Hilfe konstruierten Schätzungen müssen individuell auf ihre Eigenschaften untersucht werden. Einige allgemeine Aussagen lassen sich jedoch treffen.

#### 1. Momentenmethode

Es sei  $(\Omega, \mathfrak{A}, (P_\vartheta, \vartheta \in \Theta))$  ein statistisches Modell und  $X$  eine reellwertige Zufallsgröße über  $(\Omega, \mathfrak{A})$ . Für ein  $k \geq 1$  gelte  $E_\vartheta |X|^k < \infty, \vartheta \in \Theta$ .

Wir setzen  $\mu_l(\vartheta) := E_\vartheta X^l, \quad 1 \leq l \leq k, \vartheta \in \Theta$ .

Dann ist, falls  $X^{(n)} = (X_1, \dots, X_n)$  eine mathematische Stichprobe aus einer nach  $P_\vartheta^X$  verteilten Grundgesamtheit bildet,  $\hat{\mu}_l(X^{(n)}) := \frac{1}{n} \sum_{k=1}^n X_k^l$  eine Schätzung für  $\mu_l(\vartheta)$ . Das Prinzip besteht also darin, zur Schätzung des  $l$ -ten Momentes  $\mu_l(\vartheta)$  der Zufallsgröße  $X$  bez. der Verteilung  $P_\vartheta$  das  $l$ -te Moment der empirischen Verteilungsfunktion der mathematischen Stichprobe  $X^{(n)}$  zu verwenden.

Diese Methode lässt sich auch zur Konstruktion von Schätzungen für Größen der Form

$$g(\mu_1(\vartheta), \dots, \mu_m(\vartheta))$$

ausnutzen, wobei  $g$  irgendeine stetige Funktion auf  $R_k$  ist. Man wählt in diesem Fall

$$\hat{G}_n(X^{(n)}) := g(\hat{\mu}_1(X^{(n)}), \dots, \hat{\mu}_m(X^{(n)}))$$

als Schätzung für  $g(\mu_1(\vartheta), \dots, \mu_m(\vartheta))$ . Dieses Vorgehen zur Konstruktion von Schätzungen bezeichnet man als *Momentenmethode*.

Diese Methode der Gewinnung von Schätzungen bezieht ihre Rechtfertigung aus der Gültigkeit des starken Gesetzes der großen Zahlen. Es gilt nämlich  $P_\vartheta - f.s.$

$$\lim_{n \rightarrow \infty} \hat{\mu}_l(X^{(n)}) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k^l = E_\vartheta X^l = \mu_l(\vartheta), \vartheta \in \Theta \quad (12.11)$$

und

$$\lim_{n \rightarrow \infty} g(\hat{\mu}_1(X^{(n)}), \dots, \hat{\mu}_m(X^{(n)})) = g(\mu_1(\vartheta), \dots, \mu_m(\vartheta)), \vartheta \in \Theta. \quad (12.11')$$

Man geht also bei großem Stichprobenumfang davon aus, dass  $\hat{\mu}_l(X^{(n)})$  in der Nähe von  $\mu_l(\vartheta)$  liegt, wobei  $\vartheta$  der wahre Parameter ist.

Die Eigenschaft (12.11) bzw. (12.11') wird auch (*starke*) *Konsistenz der Schätzungen*  $\hat{\mu}_l(X^{(n)})$ ,  $n \geq 1$ , bzw.  $\hat{G}_n(X^{(n)})$ ,  $n \geq 1$ , genannt.

## 2. Maximum-Likelihood-Methode

Es sei  $(\Omega, \mathfrak{A}, (P_\vartheta, \vartheta \in \Theta \subseteq R_k))$  ein statistisches Modell und  $X$  eine reellwertige Zufallsgröße über  $(\Omega, \mathfrak{A})$ . Mit  $F_\vartheta$  werde die Verteilungsfunktion von  $X$  bez.  $P_\vartheta$  bezeichnet,  $\vartheta \in \Theta$ . Weiterhin sei  $X^{(n)} = (X_1, X_2, \dots, X_n)$  eine mathematische Stichprobe aus einer nach  $F_\vartheta, \vartheta \in \Theta$ , verteilten Grundgesamtheit und  $x^{(n)} = (x_1, \dots, x_n)$  eine Realisierung von  $X^{(n)}$  (konkrete Stichprobe). Wir nehmen der Einfachheit halber an, dass  $F_\vartheta$  für jedes  $\vartheta \in \Theta$  eine Dichte  $f_\vartheta$  besitzt (Fall 1) oder für jedes  $\vartheta \in \Theta$  eine diskrete Verteilung mit den Einzelwahrscheinlichkeiten  $p_\vartheta(a_j) := P_\vartheta(X_1 = a_j)$ ,  $j \in \mathfrak{J} \subseteq N$  (Fall 2) darstellt. Die Menge  $A = \{a_j | j \in \mathfrak{J}\}$  bildet im zweiten Fall die Menge der möglichen Werte von  $X$ .

Bei festem  $x^{(n)}$  ist durch die Funktionen

$$\vartheta \longrightarrow L_n(\vartheta; x^{(n)}) = \prod_{k=1}^n f_\vartheta(x_k) \quad , \quad \vartheta \in \Theta \quad (1. \text{ Fall}) \quad :$$

bzw.

$$\vartheta \longrightarrow L_n(\vartheta; x^{(n)}) = \prod_{k=1}^n p_\vartheta(x_k), \vartheta \in \Theta \quad (2. \text{ Fall})$$

die *Likelihoodfunktion*  $L_n(\vartheta, x^{(n)})$  der Familie  $(P_\vartheta^X, \vartheta \in \Theta)$  gegeben.

**Definition 12.19** *Als Maximum-Likelihood-Schätzwert bezeichnet man jeden Wert  $\hat{\vartheta}_n(x^{(n)})$  mit*

$$L_n(x^{(n)}; \hat{\vartheta}_n(x^{(n)})) = \max_{\vartheta \in \Theta} L_n(x^{(n)}; \vartheta).$$

Man wählt den Parameter  $\vartheta \in \Theta$  also so, dass die beobachtete Stichprobe  $x^{(n)}$  im Fall 1. Ort der maximalen Dichte von  $X^{(n)}$  bzw. im Fall 2. der Parameter ist, für den  $X^{(n)}$  die maximale Wahrscheinlichkeit besitzt. Setzt man die mathematische Stichprobe  $X^{(n)}$  anstelle  $x^{(n)}$  ein, so erhält man eine *Maximum-Likelihood-Schätzung*  $\hat{\vartheta}_n(X^{(n)})$ . Dabei handelt es sich um eine Zufallsgröße mit Werten in  $\Theta$ , deren Wert von der Stichprobe  $X^{(n)}$  abhängt.

Das Prinzip der Maximum-Likelihood-Methode ist ein sehr allgemeines. Man könnte es so formulieren:

Kann eine Erscheinung mehrere Ursachen haben, so nimmt man diejenige als die wahre Ursache an, für die die Wahrscheinlichkeit dafür, dass sie die Erscheinung nach sich zieht, am größten ist.

R.A. Fisher: "Finde diejenigen Voraussetzungen, die das Beobachtete mit großer Wahrscheinlichkeit nach sich ziehen und fasse Zutrauen, dass diese Voraussetzungen die wirksamen sind."

Anstelle  $L_n$  kann man auch  $\ln L_n$  bez.  $\vartheta$  maximieren. Das führt häufig zu rechnerischen Vorteilen, da  $\ln L_n$  eine Summe,  $L_n$  dagegen ein Produkt von Funktionen von  $\vartheta$  ist.

In vielen Fällen ist die Likelihoodfunktion stetig differenzierbar bzw.  $\vartheta$  und das Maximum bez.  $\vartheta$  liegt nicht auf dem Rand von  $\Theta$ . Dann sind die Gleichungen

$$\frac{\partial}{\partial \vartheta_m} L_n(x^{(n)}; \vartheta) = 0, \quad m = 1, 2, \dots, k \quad (12.12)$$

notwendige Bedingung für  $\vartheta = \hat{\vartheta}_n(x^{(n)})$  und liefern häufig bereits eine Lösung  $\hat{\vartheta}_n(x^{(n)})$ .  
(*Maximum-Likelihood-Gleichungen*)

Äquivalent zu (12.12) sind folgende häufig besser zu behandelnde Gleichungen, die man ebenfalls als Maximum-Likelihood-Gleichungen be-

zeichnet.

$$\frac{\partial}{\partial \vartheta_m} \ln L_n(x^{(n)}, \vartheta) = 0, \quad m = 1, 2, \dots, k. \quad (12.13)$$

### Beispiele 12.20:

1)  $\vartheta = (\mu, \sigma^2)^T \in R_1 \times (0, \infty)$ ,  $F_\vartheta = N(\mu, \sigma^2)$

$$\ln L_n(x^{(n)}; \vartheta) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - \mu)^2$$

Aus den Maximum-Likelihood-Gleichungen (12.13) ergibt sich die eindeutige Lösung

$$\hat{\mu}_n = \frac{1}{n} \sum_{k=1}^n x_k, \quad \hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \hat{\mu}_n)^2, \quad \hat{\vartheta}_n = (\hat{\mu}_n, \hat{\sigma}_n^2)^T.$$

2) Poissonverteilung:

$$L_n(x^{(n)}, \lambda) = C \cdot \exp\left(\sum_{k=1}^n x_k \cdot \ln \lambda - n\lambda\right)$$

mit einer nicht von  $\lambda$  abhängenden Konstanten  $C$ .

$$\frac{d}{d\lambda} \ln L_n(x^{(n)}, \lambda) = 0$$

liefert  $\hat{\lambda}_n = \frac{1}{n} \sum_{k=1}^n x_k$ .

3) Gleichmäßige Verteilung auf  $[0, \vartheta]$ :

$$L_n(\vartheta; x^{(n)}) = \frac{1}{\vartheta^n} \prod_{k=1}^n \mathbb{1}_{[0, \vartheta]}(x_k) = \frac{1}{\vartheta^n} \mathbb{1}_{[0, \vartheta]}(\max(x_1, \dots, x_n)).$$

In diesem Fall ist  $L_n$  bez.  $\vartheta$  nicht differenzierbar und wird maximal für  $\vartheta = \max(x_1, \dots, x_n)$ .

Folglich lautet die Maximum-Likelihood-Schätzung hier

$$\hat{\vartheta}_n = \max(X_1, X_2, \dots, X_n).$$

Maximum-Likelihood-Schätzungen sind i. Allg. nicht erwartungstreu, aber (*schwach*) *konsistent*, d. h., es gilt

$$\hat{\vartheta}_n(X_1, \dots, X_n) \xrightarrow{P_\vartheta} \vartheta, \vartheta \in \Theta.$$

Außerdem ist unter gewissen Regularitätsbedingungen an die zugrundeliegenden Verteilungen  $P_\vartheta$  (der Einfachheit halber sei  $\Theta \subseteq R_1$ )

$$\sqrt{n}(\hat{\vartheta}_n(X_1, \dots, X_n) - \vartheta) \xrightarrow{d} N\left(0, \frac{1}{I_1(\vartheta)}\right) \quad (12.14)$$

mit  $I_1(\vartheta) = E_\vartheta\left(\frac{d}{d\vartheta} \ln f_\vartheta(X)\right)^2 = \int \frac{(f'_\vartheta(x))^2}{f_\vartheta(x)} dx$ , falls  $F_\vartheta$  die Dichte  $f_\vartheta$  hat

bzw.  $E_\vartheta\left(\frac{d}{d\vartheta} \ln p_\vartheta(X)\right)^2 = \sum_x \left(\frac{dp_\vartheta(x)}{d\vartheta}\right)^2 / p_\vartheta(x)$ , falls  $F_\vartheta$

Verteilungsfunktion einer diskreten Verteilung mit den Einzelwahrscheinlichkeiten

$p_\vartheta(x), x \in A, \vartheta \in \Theta$  ist.

Das bedeutet insbesondere, Maximum-Likelihood-Schätzungen sind *asymptotisch effizient*. Für große  $n$  hat dann  $\hat{\vartheta}_n$  nämlich annähernd die Varianz  $(nI_1(\vartheta))^{-1}$ .

Maximum-Likelihood-Schätzungen sind häufig einfach auszurechnen, existieren aber nicht immer bzw. sind eventuell nicht eindeutig bestimmt. Weitere Details und Beweise findet man z.B. in Winkler (1983) und Dacunha-Castelle, Band I, (1999).

Die Eigenschaft (12.14) kann man nutzen, sogenannte *Vertrauensintervalle* für die Schätzungen von  $\vartheta$  zu konstruieren. Es gilt wegen (12.14) nämlich für  $\alpha \in (0, 1)$

$$P_\vartheta\left(\sqrt{n} I_1^{\frac{1}{2}}(\vartheta)(\hat{\vartheta}_n - \vartheta) \leq x\right) \approx \Phi(x)$$

und somit

$$P_{\vartheta} \left( \hat{\vartheta}_n - \frac{x}{\sqrt{n}} I_1^{-\frac{1}{2}}(\vartheta) \leq \vartheta \leq \hat{\vartheta}_n + \frac{x}{\sqrt{n}} I_1^{-\frac{1}{2}}(\vartheta) \right) \approx$$

$$1 - 2(1 - \Phi(x)) = 2\Phi(x) - 1.$$

Das bedeutet,

mit der Wahrscheinlichkeit  $1 - \alpha$  überdeckt das Intervall

$$K_{\alpha,n} := \left( \hat{\vartheta}_n - \frac{1}{\sqrt{n}} q_{1-\frac{\alpha}{2}} I_1^{-\frac{1}{2}}(\vartheta), \hat{\vartheta}_n + \frac{1}{\sqrt{n}} q_{1-\frac{\alpha}{2}} I_1^{-\frac{1}{2}}(\vartheta) \right)$$

den unbekanntem Parameter  $\vartheta$ .

Hat man eine positive untere Schranke  $I_0$  für  $I_1^{\frac{1}{2}}(\vartheta)$ ,  $\vartheta \in \Theta$ , so überdeckt auch

$$\tilde{K}_{\alpha,n} := \left( \hat{\vartheta}_n - \frac{1}{\sqrt{n}} q_{1-\frac{\alpha}{2}} I_0^{-1}, \hat{\vartheta}_n + \frac{1}{\sqrt{n}} q_{1-\frac{\alpha}{2}} I_0^{-1} \right)$$

den unbekanntem wahren Parameter  $\vartheta$  mit mindestens der  $P_{\vartheta}$ -Wahrscheinlichkeit  $1 - \alpha$ .

## 12.3 Elemente der Testtheorie

Wir gehen in diesem Punkt auf einige Grundbegriffe der statistischen Testtheorie ein und beschränken uns auf beispielhafte Ausführungen.

Gegeben sei ein zufälliges Experiment  $(\Omega, \mathfrak{A}, P)$  mit einer Zufallsgröße  $X$ , die nur zwei mögliche Werte annehmen kann:

$$P(X = 1) = p, P(X = 0) = 1 - p =: q, \quad p \in (0, 1).$$

Die Wahrscheinlichkeit  $p$  sei unbekannt.

**Beispiel 12.21:** Zufälliges Werfen einer Münze.

Erscheint im  $k$ -ten Wurf das Wappen, so wird  $X_k = 1$  gesetzt, anderenfalls  $X_k = 0$ .

Beim Münzenwurf liegt die Vermutung  $p = \frac{1}{2}$  nahe. Man spricht von einer Hypothese  $H_0 : p = \frac{1}{2}$ , oder im Allgemeinen  $H_0 : p = p_0$  für ein gegebenes  $p_0$ .

Zur Verfügung stehe eine konkrete Stichprobe  $x^{(n)}$  vom Umfang  $n$  aus einer wie  $X$  verteilten Grundgesamtheit:

$x^{(n)} = (x_1, x_2, \dots, x_n)$  mit  $x_k \in \{0, 1\}, k = 1, \dots, n$ .

Anhand der Stichprobe soll geprüft werden, ob die Hypothese  $H_0 : p = p_0$  zutrifft.

Grundidee: Wenn  $H_0$  richtig ist, so sollte die relative Häufigkeit des Auftretens von Eins in  $x^{(n)}$  auf Grund des Gesetzes der großen Zahlen etwa gleich  $p_0$  sein.

Sollte diese relative Häufigkeit stark von  $p_0$  abweichen, so sind Zweifel an der Richtigkeit der Hypothese angebracht, wir werden  $H_0$  ablehnen.

### 12.3.1 Beispiel eines Alternativtests

”Tea tasting person” (siehe Krengel (2002))

Eine Person behauptet, anhand des Geschmacks bei jeder mit Zitrone und Zucker versehenen Tasse Tee in durchschnittlich 8 von 10 Fällen entscheiden zu können, ob zuerst die Zitrone oder zuerst der Zucker hinzu getan wurde. Wir bezweifeln diese Fähigkeit und vertreten die Hypothese, dass die Person ihre Aussage jedesmal rein zufällig trifft. Bezeichnet  $p$  die Wahrscheinlichkeit, mit der die Person die richtige Entscheidung trifft, so lautet unsere Hypothese  $H_0 : p = \frac{1}{2}$ , die der Person  $H_1 : p = 0,8$ .

Um zu einer Entscheidung zu kommen welcher Hypothese Glauben zu schenken ist, werden  $n = 20$  Tassen verkostet. Ist die Entscheidung der Person bei der  $k$ -ten Tasse richtig, so setzen wir  $x_k = 1$ , sonst  $x_k = 0$ . Im Ergebnis erhalten wir eine konkrete Stichprobe  $x^{(n)} = (x_1, x_2, \dots, x_n)$  aus Nullen und Einsen.

Als Entscheidungsgröße berechnen wir die Anzahl  $s_n = \sum_{k=1}^n x_k$  der Erfolge der

Person beim  $n$ -maligen Prüfen. Ist  $\frac{s_n}{n}$  wesentlich größer als  $\frac{1}{2}$ , etwa in der Nähe von 0,8, würde man der Behauptung der Person Glauben schenken und unsere Hypothese  $H_0 : p = \frac{1}{2}$  verwerfen. Ist dagegen  $\frac{s_n}{n}$  in der Nähe von  $\frac{1}{2}$  (oder sogar kleiner), so würde man  $H_0$  annehmen und die Behauptung der Person

zurückweisen.

Um diese Vorgehensweise präzisieren zu können, gehen wir dazu über, die Situation vorab zu betrachten, bevor die Verkostung stattfindet. Dann wird das zukünftige Ergebnis der Verkostung durch einen zufälligen Vektor  $X^{(n)} = (X_1, \dots, X_n)$  mit  $X_k = 1$ , falls die Person im  $k$ -ten Versuch recht hat, andernfalls  $X_k = 0$ , modelliert. Wir nehmen an,  $X^{(n)}$  bestehe aus unabhängigen Zufallsgrößen  $X_k$  mit  $P^{X_k}(\{1\}) = p, P^{X_k}(\{0\}) = 1 - p, k = 1, \dots, n$ , und  $p$  sei unbekannt. Das heißt,  $X^{(n)}$  bildet eine mathematische Stichprobe aus einer wie  $X_1$  verteilten Grundgesamtheit. Unsere Hypothese ist  $H_0 : p = \frac{1}{2}$ , die der Person  $H_1 : p = 0,8$ .  $H_0$  wird auch als *Nullhypothese*,  $H_1$ , als *Alternativhypothese* bezeichnet.

Es sei zunächst vermerkt, dass eine absolut sichere Entscheidung auf der Grundlage der Kenntnis von  $X^{(n)}$  nicht möglich ist, da jede der  $2^n$  Möglichkeiten für  $X^{(n)}$  unter beiden Hypothesen mit positiver Wahrscheinlichkeit eintreten kann. Allerdings ist unter  $H_1$  eine größere Anzahl richtiger Antworten wahrscheinlicher als unter  $H_0$ .

Entscheidungsvorschrift: Wenn die Anzahl  $S_n = \sum_{k=1}^n X_k$  richtiger Antworten größer oder gleich einer noch festzulegenden Zahl  $n_0$  ist, so wird  $H_0$  abgelehnt und  $H_1$  angenommen. Ist  $S_n$  kleiner als  $n_0$ , so wird  $H_1$  abgelehnt und  $H_0$  angenommen.

Die Zufallsgröße  $S_n$  heißt in diesem Zusammenhang die *Testgröße* und  $K := \{n_0, n_0 + 1, \dots, n\}$  der *kritische Bereich*: Die Zahl  $n_0$  nennt man *kritischen Wert*. Im Fall  $S_n \in K$  wird  $H_0$  abgelehnt.

Es gibt bei dem geschilderten Vorgehen zwei mögliche Fehlerarten:

*Fehler erster Art:*  $H_0$  ist richtig und wird abgelehnt (d. h. in unserem Fall,  $H_1$  wird angenommen).

*Fehler zweiter Art:*  $H_0$  ist nicht richtig und wird nicht abgelehnt (in unserem Fall:  $H_1$  ist richtig und  $H_0$  wird angenommen).

Die durch die Wahl des kritischen Bereiches, hier also durch den "kritischen Wert"  $n_0$ , lassen sich die Wahrscheinlichkeiten der Fehler erster und zweiter Art beeinflussen. Je umfangreicher  $K$  (d.h. je kleiner  $n_0$ ) ist, umso größer wird die Wahrscheinlichkeit des Fehlers erster Art und umso kleiner die Wahrscheinlichkeit des Fehlers zweiter Art.

In der Praxis legt man Wert darauf, dass die Wahrscheinlichkeit des Fehlers erster Art kleiner oder gleich einer vor dem Test festzulegenden *Irrtumswahrscheinlichkeit*  $\alpha$  ist. Für  $\alpha$  wählt man üblicherweise 0,05 oder, falls ein Fehler erster Art gravierende Schäden verursachen kann, 0,01, eventuell sogar noch kleiner. Häufig ist dadurch der kritische Bereich  $K$  und somit das Testverfahren schon festgelegt. Der Fehler zweiter Art ist dann bereits bestimmt und kann u. U. relativ groß sein. Es ist aber zunächst einmal von Interesse, die Wahrscheinlichkeit des Fehlers erster Art kleiner oder gleich  $\alpha$  zu haben. Gemeinhin wählt man dabei als Nullhypothese diejenige Hypothese, deren Ablehnung, obwohl sie richtig ist, die schädlicheren Konsequenzen hat.

Angenommen  $H_0$  in unserem Test ist richtig. Dann beträgt die Wahrscheinlichkeit eines Fehlers erster Art

$$p_1(n_0) := P_{\frac{1}{2}}(S_n \in K) = 2^{-n} \sum_{k=n_0}^n \binom{n}{k}.$$

Je kleiner  $n_0$  ist, umso größer wird  $p_1(n_0)$ .

Der kritische Wert  $n_0$  wird nun so groß gewählt, dass  $p_1(n_0) \leq \alpha$  gilt. Allerdings vergrößert sich mit  $n_0$  auch der Fehler zweiter Art:

$$p_2(n_0) := P_{0,8}(S_n \notin K) = \sum_{k=0}^{n_0-1} \binom{n}{k} 0,8^k 0,2^{n-k}.$$

Man wird also  $n_0$  unter Einhaltung von  $p_1(n_0) \leq \alpha$  möglichst klein wählen:

$$n_0 := \min\{m \in \{1, 2, \dots, n\} : \sum_{k=m}^n \binom{n}{k} 2^{-n} \leq \alpha\}.$$

Als Wahrscheinlichkeit  $\beta$  des Fehlers zweiter Art ergibt sich dann  $\beta = p_2(n_0)$ . Die Zahl  $1 - \beta$  bezeichnet man auch als *Macht des Testes*.

Zahlenbeispiel:

$$n = 20, p_0 = \frac{1}{2}$$

$m$	14	15	16	17
$P_{\frac{1}{2}}(S_n \geq m)$	0,0577	0,0476	0,0207	0,0059

$H_0$  wird abgelehnt und  $H_1$  angenommen (mit der Irrtumswahrscheinlichkeit  $\alpha \leq 0,05$ ), falls mindestens bei  $n_0 = 15$  Tassen richtig entschieden wird.

Die Wahrscheinlichkeit des Fehlers zweiter Art beträgt in diesem Fall  $P_{0,8}(S_n < 15) = p_2(n_0) = 0,196$ . Sie ist also wesentlich größer als die Wahrscheinlichkeit des Fehlers erster Art.

Um die Wahrscheinlichkeiten der Fehler erster und zweiter Art in ihrer Abhängigkeit von  $\alpha$  und  $n$  zu studieren, untersucht man die *Gütefunktion des Testes*:

$$g_{n_0}(p) := P_p(S_n \geq n_0) = \sum_{k=n_0}^n \binom{n}{k} p^k (1-p)^{n-k}, \quad p \in (0,1).$$

Für jedes  $p \in (0,1)$  ist der Wert  $g_{n_0}(p)$  die Wahrscheinlichkeit, bei dem oben konstruierten Test die Hypothese,  $H_0 : p = \frac{1}{2}$  abzulehnen, falls die tatsächliche Wahrscheinlichkeit gleich  $p$  ist. Nach Konstruktion gilt in dem von uns betrachteten Fall

$$g_{n_0}(0) = 0, \quad g_{n_0}(1) = 1,$$

$$g_{n_0}\left(\frac{1}{2}\right) = p_1(n_0) \leq \alpha,$$

$$g_{n_0}(0,9) = 1 - p_2(n_0) = 1 - \beta.$$

Liegt  $p$  zwischen  $\frac{1}{2}$  und  $0,8$ , so ist die Wahrscheinlichkeit des Fehlers zweiter Art noch größer als bei  $p = 0,8$ . Wenn in unserem Fall die Person gesagt hätte, sie rät durchschnittlich in sechs von zehn Fällen richtig, also  $H_1 : p = 0,6$ , so wäre die Wahrscheinlichkeit des Fehlers zweiter Art recht groß, nämlich  $P_{0,6}(S_{20} < 15) = 0,874$ . Wir würden also, falls  $H_1$  richtig ist, trotzdem  $H_0$  mit hoher Wahrscheinlichkeit annehmen. In einem solchen Fall sagt man, falls  $S_n \notin K$  eintritt, nicht, dass  $H_1$  falsch und  $H_0$  richtig ist, sondern etwas zurück-

haltender, dass auf Grund der Stichprobe gegen  $H_0$  nichts einzuwenden ist. Gegebenenfalls zieht man weitere Entscheidungskriterien heran. Insbesondere wäre das der Fall, wenn die Person nur behauptet, dass sie mit einer Wahrscheinlichkeit  $p$ , die größer als  $\frac{1}{2}$  ist, die richtige Entscheidung trifft.

### 12.3.2 Signifikanztests

Wir betrachten erneut die Situation, dass eine Zufallsgröße  $X$  gegeben ist, deren Verteilung  $P^X$  unbekannt ist, von der man aber weiß, dass sie zu einer Familie  $\mathfrak{P}^X = (P_\vartheta^X, \vartheta \in \Theta)$  mit  $\Theta \subseteq R_k$  gehört. Wir formulieren eine Hypothese  $H_0 : \vartheta = \vartheta_0$ , d. h., wir unterstellen, dass der wahre Parameter  $\vartheta_0$  ist, mit anderen Worten, dass  $P^X = P_{\vartheta_0}^X$  gilt. Diese Hypothese soll an Hand einer Stichprobe  $x^{(n)} = (x_1, x_2, \dots, x_n)$  aus einer nach  $P^X$  verteilten Grundgesamtheit geprüft werden. Wie im vorigen Abschnitt bezeichnet man  $H_0$  als Nullhypothese. Allerdings formulieren wir jetzt keine Alternativhypothese. Häufig ist nämlich die Alternative zur Hypothese  $H_0$  nicht einmal genau festlegbar. Solche Tests nennt man Signifikanztests.

Mitunter setzt man Signifikanztests auch dazu ein, allgemeinere Hypothesen zu testen, zum Beispiel  $H_0 : \vartheta \in \Theta_0$  bei vorgegebenem  $\Theta_0 \subset \Theta$ . Mittels der Stichprobe  $x^{(n)}$  soll also entschieden werden, ob  $H_0$  abzulehnen ist. Dabei soll die Wahrscheinlichkeit einer Fehlentscheidung, wenn  $H_0$  richtig ist (Fehler erster Art) nicht größer als eine vorgegebene Zahl  $\alpha \in (0, 1)$  sein. Die Zahl  $\alpha$  heißt *Irrtumswahrscheinlichkeit*, die Zahl  $1 - \alpha$  nennt man das *Signifikanzniveau*. (Ein Fehler zweiter Art ist hier mangels Alternativhypothese nicht vorhanden.)

Dazu konstruieren wir wie folgt einen statistischen Test.

1. Wir wählen eine Stichprobenfunktion  $T_n = T_n(x_1, x_2, \dots, x_n)$ , wobei  $T_n$  eine Borelmeßbare Funktion sein möge, die wir hier überdies als reellwertig annehmen.
2. Wir wählen einen *kritischen Bereich*  $K$ , d. h. eine Borelmeßbare Teilmenge des Wertbereiches von  $T_n$ , so dass

$$P_{\vartheta_0}^X(T_n \in K) \leq \alpha$$

erfüllt ist. (Hat  $T_n$  eine stetige Verteilung, wird man  $K$  so wählen, dass  $P_{\vartheta_0}^X(T_n \in K) = \alpha$  gilt.)

3. Sodann vereinbaren wir die *Entscheidungsregel*:

Die Hypothese  $H_0 : \vartheta = \vartheta_0$  wird auf Grund der Stichprobe  $x^{(n)}$  abgelehnt, falls  $T_n(x_1, \dots, x_n) \in K$ . Anderenfalls, also wenn  $T_n(x_1, \dots, x_n) \notin K$  gilt, ist gegen  $H_0$  auf Grund der Stichprobe nichts einzuwenden.

Man sagt im Fall der Ablehnung, dass sie zum *Signifikanzniveau*  $1 - \alpha$  erfolge und bezeichnet den so konstruierten Test als *Signifikanztest der Hypothese  $H_0$  zum Signifikanzniveau  $1 - \alpha$* .

In der Wahl des kritischen Bereiches  $K$  steckt noch eine gewisse Willkür. Häufig ist er durch die konkreten Rahmenbedingungen nahegelegt. Allgemein sollte er so konstruiert werden, dass das Ereignis  $\{T_n \in K\}$  unter  $H_0$  eine derart kleine Wahrscheinlichkeit hat ( $\leq \alpha$ ), dass man das Eintreten von  $\{T_n \in K\}$  nicht als Zufall ansieht, sondern eher daran zweifelt, dass die Hypothese  $H_0$  stimmt. Das wird umso mehr berechtigt sein, wenn das Ereignis  $\{T_n \in K\}$  für den Fall, dass  $H_0$  nicht stimmt, eine große Wahrscheinlichkeit besitzt.

### Beispiele 12.22:

1. *Test des Mittelwertes einer  $N(\mu, \sigma^2)$ -verteilten Grundgesamtheit bei bekannter Streuung  $\sigma^2$*

Es sei  $X \sim N(\mu, \sigma^2)$  und  $X^{(n)} = (X_1, X_2, \dots, X_n)$  eine mathematische Stichprobe aus einer wie  $X$  verteilten Grundgesamtheit. Die Varianz  $\sigma^2$  sei bekannt,  $\alpha \in (0, 1)$  sei vorgegeben. Wir konstruieren einen Signifikanztest der Hypothese  $H_0 : \mu = \mu_0$  zum Niveau  $1 - \alpha$ .

Als Testgröße wählen wir

$$T_n(X^{(n)}) = \frac{(\bar{X}_n - \mu_0)\sqrt{n}}{\sigma},$$

wobei  $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$  gesetzt wurde.

Offenbar besitzt  $T_n = T_n(X^{(n)})$  eine  $N(0, 1)$ -Verteilung, falls  $H_0$  richtig ist. Stimmt  $H_0$ , so wird die Zufallsgröße  $T_n(X^{(n)})$  bei ihrer Realisierung mit hoher Wahrscheinlichkeit einen Wert in der Nähe von Null annehmen. Stimmt  $H_0$  nicht, ist also  $\mu \neq \mu_0$ , so hat  $T_n = \frac{(\bar{X}_n - \mu)\sqrt{n}}{\sigma} + \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}$  eine  $N\left(\frac{(\mu - \mu_0)\sqrt{n}}{\sigma}, 1\right)$ -Verteilung. Ihre Realisierung würde stark von Null abweichen (falls  $\mu$  sich stark von  $\mu_0$  unterscheidet). Deshalb wählen wir den kritischen Bereich  $K$  in der Form

$$K = \{t \mid |t| > z_{\alpha, n}\}$$

und bestimmen  $z_{\alpha, n}$  so, dass unter  $H_0$  gilt

$$P(|T_n(X^{(n)})| > z_{\alpha, n}) = \alpha.$$

Das ergibt wegen

$$P(|T_n(X^{(n)})| > z_{\alpha, n}) = 2(1 - \Phi(z_{\alpha, n}))$$

die Beziehung

$$z_{\alpha, n} = q_{1 - \frac{\alpha}{2}}$$

( $q_p$  bezeichnet das  $p$ -Quantil der Standard Normalverteilungsfunktion  $\Phi$ ).

*Entscheidungsregel:*  $H_0 : \mu = \mu_0$  wird abgelehnt, falls für die konkrete Stichprobe  $x^{(n)} = (x_1, x_2, \dots, x_n)$  gilt

$$|T_n(x^{(n)})| > q_{1 - \frac{\alpha}{2}}, \text{ d. h., falls gilt:}$$

$$|\bar{X}_n - \mu_0| > \frac{\sigma q_{1 - \frac{\alpha}{2}}}{\sqrt{n}}.$$

Anderenfalls ist gegen  $H_0$  auf Grund der Stichprobe  $x^{(n)}$  nichts einzuwenden.

**Bemerkung 12.23:** Ist der Stichprobenumfang  $n$  groß, so nimmt man für  $\sigma^2$ , falls nicht anders verfügbar, die Schätzung  $\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x}_n)^2$ . Das ist auf Grund des Gesetzes der großen Zahlen gerechtfertigt, da  $E\hat{\sigma}_n^2 = \sigma^2$  gilt. Wie man im Fall kleiner Stichprobenumfänge verfährt, wird im folgenden Beispiel erläutert.

2. *Test des Mittelwertes einer  $N(\mu, \sigma^2)$ -verteilten Grundgesamtheit bei unbekannter Streuung*

Wir behandeln das gleiche Problem wie im vorangegangenen Beispiel, nehmen aber an,  $\sigma^2$  ist nicht bekannt und der Stichprobenumfang  $n$  ist nicht allzu groß, so dass man bezweifeln kann, dass

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

bereits eine gute Näherung für  $\sigma^2$  ist. In diesem Fall verwendet man

$$T'_n(X^{(n)}) = \frac{(\bar{X}_n - \mu_0)}{\hat{\sigma}_n} \cdot \sqrt{n}$$

als Testgröße. Wir benötigen die Verteilung von  $T'_n(X^{(n)})$  unter der Nullhypothese  $H_0$ , um den kritischen Bereich bestimmen zu können.

**Lemma 12.24:**  $\bar{X}_n$  und  $\hat{\sigma}_n^2$  sind unter  $H_0$  voneinander unabhängige Zufallsgrößen.  $\bar{X}_n$  ist  $N(\mu_0, \frac{\sigma^2}{n})$ -verteilt und  $\frac{\hat{\sigma}_n^2}{\sigma^2} \cdot (n-1)$  besitzt eine  $\chi^2$ -Verteilung mit  $n-1$  Freiheitsgraden, d. h. eine Gammaverteilung  $\Gamma(\alpha, \lambda)$  mit den Parametern  $\alpha = \frac{n-1}{2}$   $\lambda = \frac{1}{2}$ . (Siehe auch Abschnitt 12.3.3)

Der Beweis soll in Übung 12.4 geführt werden (siehe auch Krenzel (2002), Kap. II, § 14). Als Verteilung von  $T'_n$  ergibt sich damit die Verteilung mit der Dichte

$$f_{n-1}(x) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\Gamma\left(\frac{n-1}{2}\right)\left(\frac{x^2}{n-1} + 1\right)^{\frac{n}{2}}}, \quad x \in R_1.$$

Diese Verteilung trägt die Bezeichnung *t-Verteilung* (oder *Studentverteilung*) mit  $n - 1$  Freiheitsgraden, die Werte ihrer Verteilungsfunktion  $F_{n-1}$  bzw. ihre Quantile sind vertafelt und in vielen Büchern über Mathematische Statistik zu finden, siehe zum Beispiel Krengel (2002), Tabellen Seite 247.

Auch hier wählen wir den kritischen Bereich  $K$  in der Form

$$K = \{t \mid |t| > z_{\alpha,n}\}$$

und bestimmen  $z_{\alpha,n}$  derart, dass unter  $H_0$  gilt

$$P(|T'_n(X^{(n)})| > z_{\alpha,n}) = \alpha.$$

Das ergibt

$$2(1 - F_{n-1}(z_{\alpha,n})) = \alpha, \text{ also}$$

$$z_{\alpha,n} = t_{n-1, 1-\frac{\alpha}{2}}$$

wobei  $t_{n-1, 1-\frac{\alpha}{2}}$  das  $(1-\frac{\alpha}{2})$ -Quantil der  $t$ -Verteilung mit  $n-1$  Freiheitsgraden ist.

### 12.3.3 Der $\chi^2$ -Test

Unter den Signifikanztests hat sich der sogenannte  $\chi^2$ -Test als ein sehr flexibles statistisches Werkzeug seinen festen Platz erobert.

Es sei  $X$  eine diskret verteilte Zufallsgröße mit den endlich vielen möglichen Werten  $a_k$  aus der Menge  $A = \{a_k : k \in \{1, 2, \dots, r\}\}$  reeller Zahlen. Weiterhin sei  $\{p_k : k \in \{1, 2, \dots, r\}\}$  eine Wahrscheinlichkeitsverteilung auf  $A$ . Anhand einer Stichprobe  $x^{(n)}$  aus einer nach  $P^X$  verteilten Grundgesamtheit soll die Hypothese  $H_0$  geprüft werden, dass  $X$  die Verteilung  $(a_k, p_k), k \in K$ , besitzt, d. h.

$$H_0 : P(X = a_k) = p_k, \quad 1 \leq k \leq r.$$

Zu diesem Zweck bildet man die Testgröße

$$\chi^2 = \sum_{k=1}^r \frac{(n_k - np_k)^2}{np_k},$$

wobei  $n_k$  die Anzahl derjenigen  $x_j$  aus der Stichprobe  $x^{(n)} = (x_1, x_2, \dots, x_n)$  mit  $x_j = a_k$  bezeichne,  $1 \leq k \leq r, 1 \leq j \leq n$ .

Die Größe  $\chi^2$  ist eine gewichtete Summe der quadratischen Abweichungen zwischen den Anzahlen  $n_k$  und ihren bei richtiger Hypothese  $H_0$  "zu erwartenden" Werte  $np_k$ .

Um die wahrscheinlichkeitstheoretischen Eigenschaften dieser Testgröße zu studieren, setzen wir in  $\chi^2$  an Stelle von  $n_k$  die Zufallsgrößen  $N_k$  ein, die sich aus der entsprechenden mathematischen Stichprobe  $X^{(n)}$  genauso berechnen, wie die  $n_k$  aus der konkreten Stichprobe  $x^{(n)}$ .

**Satz 12.25:** (*R. A. Fisher*) Die Wahrscheinlichkeiten  $p_k, k = 1, 2, \dots, r$ , seien gegeben. Dann konvergieren die Verteilungsfunktionen  $F_n$  der Zufallsgrößen  $\chi^2$ , falls die Hypothese  $H_0$  richtig ist, mit wachsendem Stichprobenumfang  $n$  gegen eine Gamma-Verteilung  $\Gamma(\alpha, \lambda)$  mit den Parametern

$$\alpha = \frac{r-1}{2} \text{ und } \lambda = \frac{1}{2}:$$

$$\begin{aligned} \lim_{n \rightarrow \infty} F_n(x) &= \frac{1}{2^{\frac{r-1}{2}} \Gamma\left(\frac{r-1}{2}\right)} \int_0^x y^{\frac{r-3}{2}} e^{-\frac{y}{2}} dy, \quad x > 0 \\ &= 0, \quad x \leq 0. \end{aligned}$$

Die Verteilung  $\Gamma\left(\frac{r-1}{2}, \frac{1}{2}\right)$  trägt einen eigenen Namen und heißt  $\chi^2$ -Verteilung mit  $r - 1$  Freiheitsgraden ( $r \geq 1$ ).

Den Beweis findet man z. B. in Krenzel (2002), Kap. II, § 14.

Seine Grundidee besteht in der Beobachtung, dass der Vektor  $(N_1, N_2, \dots, N_r)$  eine Multinomialverteilung mit den Parametern  $n, p_1, p_2, \dots, p_r$  besitzt. Dann

ist  $(N_1 - np_1, \dots, N_r - np_r)$  ein zentrierter zufälliger Vektor, mit  $\sum_{k=1}^r N_k = n$ , also

$$\sum_{k=1}^r (N_k - np_k) = 0. \quad (12.15)$$

Die  $r$ -dimensionale Multinomialverteilung von  $(N_1, N_2, \dots, N_r)$  konvergiert für  $n \rightarrow \infty$  ebenso wie die Binomialverteilung im globalen Grenzwertsatz von Moivre-Laplace gegen eine Normalverteilung, die wegen (12.15) auf einem  $(r - 1)$ -dimensionalen Teilraum von  $R_r$  konzentriert ist.

Die Zufallsgröße

$$\chi^2 = \sum_{k=1}^r \frac{(N_k - np_k)^2}{np_k} \quad (12.16)$$

lässt sich damit durch Grenzübergang  $n \rightarrow \infty$  zurückführen auf die Quadratsumme von  $(r - 1)$  Standard normalverteilten Zufallsgrößen. Dann erhält man mit folgendem Lemma die Aussage des Satzes.

**Lemma 12.26:** *Es seien  $Y_1, Y_2, \dots, Y_m$  ( $m \geq 1$ ) voneinander unabhängige  $N(0, 1)$ -verteilte Zufallsgrößen. Dann besitzt*

$$S^2 = \sum_{k=1}^m Y_k^2$$

eine  $\chi^2$ -Verteilung mit  $m$  Freiheitsgraden.

Beweis:

$P(Y_1^2 \leq y) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\Phi(\sqrt{y}) - 1$ , woraus sich die Dichte  $f_{Y_1^2}$  ergibt:

$$f_{Y_1^2}(y) = \frac{2\varphi(\sqrt{y})}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}y}, \quad y > 0.$$

Also besitzt jedes  $Y_k^2$  eine  $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$ -Verteilung:

$$Y_k^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right), \quad k = 1, 2, \dots, m,$$

und auf Grund der Unabhängigkeit der  $Y_1, \dots, Y_m$  folgt

$$S^2 \sim \Gamma\left(\frac{m}{2}, \frac{1}{2}\right).$$

Die Verteilungsfunktion der  $\chi^2$ -Verteilungen ist nicht explizit berechenbar, sie bzw. ihre Quantile sind vertafelt, und man findet sie, wie oben bereits erwähnt zum Beispiel in Krenzel (2002), Tabellen, Seite 249.

Für jede mit  $m$  Freiheitsgraden  $\chi^2$ -verteilte Zufallsgröße  $Y$  gilt

$$EY = m, \quad D^2Y = 2m, \quad \text{Modalwert}(Y) = \max(0, m - 2).$$

Die Testgröße  $T_n(X^{(n)})$  wird also mit hoher Wahrscheinlichkeit (hier:  $1 - \alpha$ ) Werte annehmen, die in einem Intervall um den Modalwert liegen, z. B. in

$$\left(\chi_{r-1, \frac{\alpha}{2}}^2, \chi_{r-1, 1-\frac{\alpha}{2}}^2\right).$$

Dabei bezeichnet  $\chi_{r-1, p}^2$  das  $p$ -Quantil der  $\chi^2$ -Verteilung mit  $r - 1$  Freiheitsgraden.

Eine erste Anwendung des  $\chi^2$ -Test enthält das folgende Beispiel.

**Beispiel 12.27** ( $\chi^2$ -Anpassungstest): Es seien  $F$  eine Verteilungsfunktion auf  $R_1$  und  $X$  eine reellwertige Zufallsgröße über einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  mit

$$P(X \leq x) = F_X(x), \quad x \in R_1.$$

Wir wollen die Hypothese

$$H_0 : F_X = F$$

testen.

Zu diesem Zweck unterteilen wir  $R_1$  in  $r$  Intervalle ( $r \geq 2$ )

$$I_1 = (-\infty, a_1], I_2 = (a_1, a_2], \dots, I_{r-1} = (a_{r-2}, a_{r-1}], I_r = (a_{r-1}, \infty)$$

und setzen

$$p_k = F(a_k) - F(a_{k-1}), \quad k = 1, \dots, r$$

mit  $a_0 = -\infty$ ,  $F(a_0) = 0$ ,  $a_r = +\infty$ ,  $F(a_r) = 1$ .

Ist  $H_0$  richtig, so gilt

$$P(X \in I_k) = p_k, \quad k = 1, 2, \dots, r.$$

Wir verwenden die Testgröße

$$\chi^2 = \sum_{k=1}^r \frac{(n_k - np_k)^2}{np_k}$$

und den kritischen Bereich

$$K = R_1 \setminus (\chi_{r-1, \frac{\alpha}{2}}^2, \chi_{r-1, 1-\frac{\alpha}{2}}^2)$$

(zweiseitiger Test) bzw.

$$K = (\chi_{r-1, 1-\alpha}^2, \infty)$$

(einseitiger Test).

Der so konstruierte Signifikanztest heißt  $\chi^2$ -Anpassungstest zum Signifikanzniveau  $1 - \alpha$ .

Wir illustrieren diesen Test durch zwei Beispiele:

*Zufallszahlen aus  $[0, 1)$*

Angenommen,  $x^{(n)} = (x_1, x_2, \dots, x_n)$  ist eine  $n$ -elementige Folge reeller Zahlen aus  $[0, 1)$ . Wir wollen die Hypothese prüfen, dass sie aus einer gleichmäßig auf  $[0, 1)$  verteilten Grundgesamtheit stammen, und zwar zum Signifikanzniveau 0,95. Dazu nehmen wir an, die konkrete Stichprobe  $x^{(n)}$  ist Realisierung einer mathematischen Stichprobe  $X^{(n)} = (X_1, X_2, \dots, X_n)$ , jedes  $X_k$  habe die Ver-

teilungsfunktion  $F$  und formulieren die Hypothese

$$H_0 : F = F_0$$

mit  $F_0(x) = x$  für  $x \in [0, 1]$ ,  $= 0$  für  $x < 0$ ,  $= 1$  für  $x > 1$ .

Es sei  $n = 100$  und  $\alpha = 0,05$ .

Wir teilen  $[0, 1)$  in 10 Klassen

$$I_k = \left[ \frac{k-1}{10}, \frac{k}{10} \right), \quad k = 1, 2, \dots, 10$$

ein. Dann gilt

$$p_k = F_0\left(\frac{k}{10}\right) - F_0\left(\frac{k-1}{10}\right) = 0,1 \quad \text{für } k = 1, 2, \dots, 10$$

und die Testgröße  $\chi^2$  ergibt sich zu

$$\chi^2 = \sum_{k=1}^{10} \frac{(N_k - 10)^2}{10} = 0,1 \cdot \sum_{k=1}^{10} (N_k - 10)^2.$$

Der kritische Bereich  $K$  wird für einen zweiseitigen Test wie folgt festgelegt:

$$\begin{aligned} K &= (0, \chi_{0,025,9}^2) \cup (\chi_{0,975,9}^2, \infty) \\ &= (0; 2,70) \cup (19,02, \infty). \end{aligned}$$

Bei dieser Konstruktion wird die Hypothese  $H_0$  abgelehnt, wenn die empirische Verteilung  $\hat{F}_{100}$  zu weit von  $F_0$  entfernt ist (d. h., wenn die Testgröße  $\chi^2$  groß ist), oder wenn  $\hat{F}_{100}$  zu nahe an  $F_0$  liegt (wenn  $\chi^2$  zu klein ist). Empfindet man sehr kleine  $\chi^2$  nicht als Mangel, so kann man  $K$  auch in der Form

$$K = (\chi_{1-\alpha,9}^2, \infty) = (16,92; \infty)$$

wählen (einseitiger Test).

*Geburtenzahlen*

Im Landkreis Teltow-Fläming wurden 1996 insgesamt 360 Kinder geboren, davon 175 Mädchen und 185 Jungen. Widerspricht diese Zahl der Hypothese, dass das Geschlecht von Neugeborenen mit gleicher Wahrscheinlichkeit weiblich bzw. männlich ist, zum Signifikanzniveau von 0,95?

Bezeichnen wir mit  $p$  die Wahrscheinlichkeit, dass ein Neugeborenes ein Junge wird. Dann lautet die Hypothese

$$H_0 : p = \frac{1}{2}.$$

Die Testgröße berechnet sich zu

$$\chi^2 = \frac{(185-180)^2}{180} + \frac{(175-180)^2}{180} = \frac{50}{180} = 0,2778.$$

Da der kritische Bereich

$$K = (\chi_{1-\alpha,1}^2, \infty) = (3,84; \infty)$$

lautet, ist gegen  $H_0$  auf Grund der Stichprobe nichts einzuwenden.

In Deutschland wurde 1991 insgesamt 911 600 Kinder geboren, davon 442 400 Mädchen und 468 000 Jungen.

Wendet man den gleichen Test wie eben auf

$$H_0 : p = \frac{1}{2}$$

an, so ergibt sich

$$\chi^2 = 520,68 \gg 3,84.$$

Die Überschreitung des kritischen Wertes 3,84 durch die Testgröße ist hochsignifikant,  $H_0$  wird auf Grund dieser Stichprobe abgelehnt.

Wir kehren noch einmal zurück zum eingangs behandelten  $\chi^2$ -Test einer diskreten Verteilung  $P = (p_k, k = 1, \dots, r)$  auf  $A = \{a_1, a_2, \dots, a_r\}$ .

In manchen Fällen ist die Verteilung  $P$  nicht völlig festgelegt, sondern hängt noch von einem Parameter  $\vartheta$  ab:

$P \in \mathfrak{P} = (p_k(\vartheta), k = 1, \dots, r; \vartheta \in \Theta \subseteq R^l)$  für ein  $l \geq 1$ . Dadurch ist die Verteilung noch nicht eindeutig bestimmt und wir können den oben angegebenen Satz von Fisher nicht anwenden. Es gibt vielmehr die folgende allgemeinere Fassung.

Wir setzen voraus:

**Satz 12.28:** *Die Ableitungen*

$$\frac{\partial p_k}{\partial \vartheta_i}, \frac{\partial^2 p_k}{\partial \vartheta_i \partial \vartheta_j}, \quad k = 1, 2, \dots, r; \quad i, j \in \{1, 2, \dots, l\}$$

*existieren und sind stetig bzgl.  $\vartheta$ .*

*Die Matrix  $\left(\frac{\partial p_k}{\partial \vartheta_i}\right)_{k,i}$  habe den Rang  $l$ .*

*Werden die unbekannt Parameter  $\vartheta_1, \vartheta_2, \dots, \vartheta_l$  mit Hilfe der Stichprobe  $x^{(n)}$  nach der Maximum-Likelihood-Methode geschätzt, so konvergieren die Verteilungsfunktionen  $F_n$  der Stichprobenfunktion  $\chi^2$  aus Formel (12.16) gegen eine  $\chi^2$ -Verteilung mit  $r - l - 1$  Freiheitsgraden.*

Für einen Beweis siehe z. B. Dacunha-Castelle, Dufflo, Vol. II (1986).

**Beispiel 12.29:**

*Test auf Unabhängigkeit in Kontingenztafeln*

Gegeben seien zwei Zufallsgrößen  $X$  und  $Y$ , beide diskret verteilt mit  $r$  bzw.  $s$  möglichen Werten und den (unbekannten) Wahrscheinlichkeiten der gemeinsamen Verteilung

$$p_{ij} = P(X = x_i, Y = y_j), \quad i = 1, 2, \dots, r; \quad j = 1, 2, \dots, s.$$

Es werde eine konkrete Stichprobe vom Umfang  $n$  realisiert:

$n_{ij}$  = Häufigkeit des Auftretens des Paares  $(x_i, y_j)$ .

$i \setminus j$	1	2	...	s	
1	$n_{11}$	$n_{12}$		$n_{1s}$	$n_{1\cdot}$
$\vdots$					
r	$n_{r1}$	$n_{r2}$		$n_{rs}$	$n_{r\cdot}$
	$n_{\cdot 1}$	$n_{\cdot 2}$		$n_{\cdot s}$	$n$

Kontingenztafel

$H_0$  : "Die Merkmale  $X$  und  $Y$  sind voneinander unabhängig."

Das bedeutet mit den Bezeichnungen  $p_{i\cdot} = \sum_j p_{ij}$ ,  $p_{\cdot j} = \sum_i p_{ij}$

$H_0 : p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$ .

Durch diese Hypothese ist die Wahrscheinlichkeitsverteilung ( $p_{ij}$ ) noch nicht festgelegt. Die Größen  $p_{i\cdot}$  und  $p_{\cdot j}$  ( $1 \leq i \leq r, 1 \leq j \leq s$ ) müssen geschätzt werden.

Die Maximum-Likelihood-Methode liefert  $\hat{p}_{i\cdot} = \frac{n_{i\cdot}}{n}$  (siehe unten).

Wegen  $\sum_i p_{i\cdot} = \sum_j p_{\cdot j} = 1$  sind dies  $(r - 1) + (s - 1)$  geschätzte Parameter.

Testgröße:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(N_{ij} - p_{ij} \cdot n)^2}{n \cdot p_{ij}} = n \sum_{i=1}^r \sum_{j=1}^s \frac{(N_{ij} - \frac{n_{i\cdot} \cdot n_{\cdot j}}{n})^2}{N_{i\cdot} \cdot N_{\cdot j}}$$

Diese Testgröße besitzt für  $n \rightarrow \infty$  eine  $\chi^2$ -Verteilung mit  $r \cdot s - r - s + 2 - 1 = (r - 1)(s - 1)$  Freiheitsgraden.

*Maximum-Likelihood-Schätzung der  $p_{i\cdot}, p_{\cdot j}$ :*

Die Likelihoodfunktion ist unter der Hypothese  $H_0$  gegeben durch

$$L(\vartheta, X^{(n)}) = \prod_{i=1}^r \prod_{j=1}^s p_{ij}^{N_{ij}} = \prod_{i=1}^r \prod_{j=1}^s p_{i\cdot}^{N_{ij}} p_{\cdot j}^{N_{ij}} = \prod_{i=1}^r p_{i\cdot}^{N_{i\cdot}} \prod_{j=1}^s p_{\cdot j}^{N_{\cdot j}}$$
 mit

$$N_{ij} = \#\{k \leq n : X_k = x_i, Y_k = y_j\},$$

$$N_{i\cdot} = \sum_{j=1}^s N_{ij}, \quad N_{\cdot j} = \sum_{i=1}^r N_{ij} \quad 1 \leq i \leq r, 1 \leq j \leq s$$

und  $\vartheta = (p_{1\cdot}, \dots, p_{r-1\cdot}, p_{\cdot 1}, \dots, p_{\cdot s-1})$ .

Dabei wird gesetzt

$$p_{r\cdot} = 1 - p_{1\cdot} - \dots - p_{r-1\cdot} \text{ und}$$

$$p_{\cdot s} = 1 - p_{\cdot 1} - \dots - p_{\cdot s-1}.$$

Für die Maximum-Likelihood-Gleichungen ergibt sich

$$\frac{\partial}{\partial p_{i\cdot}} \ln L = \frac{N_{i\cdot}}{p_{i\cdot}} - \frac{N_{r\cdot}}{1 - p_{1\cdot} - \dots - p_{r-1\cdot}} = 0, \quad (12.17)$$

$i = 1, \dots, r - 1$ , also

$$\frac{N_{i\cdot}}{\hat{p}_{i\cdot}} = \frac{N_{r\cdot}}{\hat{p}_{r\cdot}}, \text{ mithin}$$

$$N_{i\cdot} = \hat{p}_{i\cdot} \cdot \frac{N_{r\cdot}}{\hat{p}_{r\cdot}}, i = 1, 2, \dots, r.$$

Summation über  $i$  liefert

$$n = 1 \cdot \frac{N_{r\cdot}}{\hat{p}_{r\cdot}}, \text{ also } \hat{p}_{r\cdot} = \frac{N_{r\cdot}}{n}.$$

Daraus ergibt sich  $\hat{p}_{i\cdot} = \frac{N_{i\cdot}}{n}$ .

Die Schätzungen  $\hat{p}_{\cdot j} = \frac{N_{\cdot j}}{n}$  ergeben sich analog aus

$$\frac{\partial}{\partial p_{\cdot j}} \ln L = 0, j = 1, \dots, s - 1.$$

# Index

- $\chi^2$ -Anpassungstest, 290, 291
- $\chi^2$ -Verteilung, 288
- $\sigma$ -Algebra, 15
- $\sigma$ -Stetigkeit von Wahrscheinlichkeitsverteilungen, 27
  
- Algebra, 15
- Anfangsverteilung, 111
- Ausgeartete Verteilung
  - charakteristische Funktion, 211
- Axiomensystem der Wahrscheinlichkeitstheorie, 24
  
- Bayes'sche Formel, 122
- bedingte Wahrscheinlichkeit, 118
  - im Laplace-Modell, 123
  - im mehrstufigen Versuch, 124
- Bernoulli-Schema, 141
- Binomialverteilung, 77
  - charakteristische Funktion, 211
  - Erwartungswert, 85
  - erzeugende Funktion, 106
  - Varianz, 90
- Bonferroni-Ungleichungen, 30
- Borel-Cantelli
  - 1. Lemma von, 29
  - 2. Lemma von, 130
  
- Cauchyverteilung
  - charakteristische Funktion, 211
  - Dichte, 183
  - Erwartungswert, 183
  - Varianz, 183
- charakteristische Funktion, 209
  - Eindeutigkeitssatz, 215
  - Faltungssatz, 216
  - Stetigkeitssatz, 216
  - Umkehrformel, 215
- Cramer-Rao-Ungleichung, 267
  
- Dichte
  - bedingte, 192
  - der Zufallsgröße  $X$ , 176
  - des Maßes  $Q$ , 176
  - Lebesgue- und Riemannintegrale, 178
  - mehrdimensionaler Verteilungen, 190
  
- effizient
  - asymptotisch, 277
- Ein- und Ausschlussformel, 29
- Eindeutigkeitssatz, *siehe* charakteristische Funktion
- einfache Zufallsgröße, 163
- Einpunktverteilung, 76
  - Erwartungswert, 85
  - erzeugende Funktion, 106
  - Varianz, 90
- Entscheidungsregel, 284
- Ereignis
  - fast sicheres, 49

- fast unmögliches, 49
- zufälliges, 11
- Erwartungswert, 167
  - diskret, 84, 86
  - Eigenschaften, 168
  - einfacher Zufallsgrößen, 164
  - endlicher, 167
  - nichtnegativer Zufallsgrößen, 166
- Erwartungswertregel, 190
- Erwartungswertvektor, 186
- erzeugende Funktion, 102
- Exponentialverteilung
  - Dichte, 182
  - Erwartungswert, 182
  - Varianz, 182
  - Verteilungsfunktion, 65
- fairer Preis, 84
- Faltungsformel, 201
- Faltungssatz, *siehe* charakteristische Funktion
- faltungsstabil, 205
- Fatou, Lemma von, 172
- Fehler
  - erster Art, 280
  - zweiter Art, 280
- Fisher'sche Informationsmatrix, 269
- Fubini, Satz von, 199
- Gütefunktion des Testes, 282
- Gammaverteilung
  - charakteristische Funktion, 211
  - Dichte, 182
  - Erwartungswert, 182
  - Varianz, 182
- geometrische Verteilung, 77
  - Erwartungswert, 85
  - erzeugende Funktion, 106
- Varianz, 90
- Gesetz der großen Zahlen
  - schwaches, 222
  - starkes, Kolmogorov, 226
- gleichmäßige Verteilung, diskret, 33, 76
  - Erwartungswert, 85
  - erzeugende Funktion, 106
  - Varianz, 90
- gleichmäßige Verteilung, stetig
  - Dichte, 182
  - Erwartungswert, 182
  - Varianz, 182
- Hauptsatz der mathematischen Statistik, 258
- hypergeometrische Verteilung, 81–83
  - Erwartungswert, 85
  - erzeugende Funktion, 106
  - Varianz, 90
- Hypothese, 279
  - Alternativhypothese, 280
  - Nullhypothese, 280
- Irrfahrt, 152
  - symmetrische, 152
  - Zeit des ersten Erreichens, 155
- Irrtumswahrscheinlichkeit, 281, 283
- Kolmogorov-Smirnov Verteilung, 260
- konsistent (schwach), 277
- Konvergenz
  - fast sichere, 172
  - majorisierte, 172
  - monotone, 172
  - stochastische, 220
- Korrelationskoeffizient, 98
- Kovarianz, 98
- Kovarianzmatrix, 187

- linearer Transformationen, 188
- kritischer
  - Bereich, 280, 283
  - Wert, 280
- Laplace
  - Experiment, 32
- Likelihoodfunktion, 266, 274
- Lindeberg-Bedingung, 249
- Münzenwurf, 16, 34
- Macht des Testes, 281
- majorisierte Konvergenz, *siehe* Konvergenz, majorisierte
- Maximum-Likelihood
  - Gleichungen, 275
  - Methode, 274
  - Schätzung, 275
  - Schätzwert, 274
- Median, 58
- Modell
  - statistisches, 261
- Moment
  - diskret, 87, 88
  - diskret, zentriert, 87, 88
- Momentenmethode, 272
- monotone Konvergenz, *siehe* Konvergenz, monotone
- Monte-Carlo-Methode, 234
- Multiplikationssatz für Wahrscheinlichkeiten, 120
- negative Binomialverteilung, 78
  - charakteristische Funktion, 211
  - Erwartungswert, 85
  - erzeugende Funktion, 106
  - Varianz, 90
- normale Zahl, 235
- Normalverteilung
  - $n$ -dimensional, 73
  - charakteristische Funktion, 211
  - Dichte, 183
  - eindimensional, 186
  - Erwartungswert, 183
  - Varianz, 183
  - zweidimensionale, 73, 194
- Pfadregel
  - erste, 110
  - zweite, 110
- Poissonverteilung, 77
  - charakteristische Funktion, 211
  - Erwartungswert, 85
  - erzeugende Funktion, 106
  - Varianz, 90
- Polya'sches Urnenschema, 111
- Produktmaß, 198
- Quantil, 58
  - unteres, oberes, 58
- Randverteilung
  - $r$ -dimensional, 62
  - diskret, 93
  - stetig, 192
- Regressionsgerade, 101, 189
- Schätzung, 262
  - beste erwartungstreue, 264
  - effiziente, 271
  - erwartungstreue, 262
  - Maximum-Likelihood, 275
  - Verzerrung der, Bias, 262
- Schätzwert, 262
  - Maximum-Likelihood, 274
- Signifikanzniveau, 283, 284
- Signifikanztest, 284
- Standardabweichung, 89

- Stetigkeitssatz, *siehe* charakteristische Funktion
- Stichprobe  
 konkrete, 256  
 mathematische, 256
- Streuung, *siehe* Varianz
- Studentverteilung, 287
- Substitutionsformel, 173
- t-Verteilung, 287
- Testgröße, 280
- Tonelli, Hobson, Folgerung von, 200
- totale Wahrscheinlichkeit, Satz von, 121
- Transformationsatz für Dichten  
 eindimensional, 184  
 n-dimensional, 74
- Uebergangsverteilung, 111
- Umkehrformel, *siehe* charakteristische Funktion
- Unabhängigkeit  
 in mehrstufigen Experimenten, 133  
 von  $\sigma$ -Algebren, 131  
 von Ereignissen, 127, 128  
 von Ereignissen, paarweise, 128  
 von Mengensystemen, 131  
 von Zufallsgrößen, 134, 136
- Ungleichung  
 von Cauchy-Schwarz, 97, 169  
 von Hölder, 170  
 von Jensen, 170  
 von Ljapunov, 171  
 von Minkowski, 171  
 von Tschebychev, 169, 175  
 diskret, 90  
 von Berry-Essen, 245
- unkorreliert, 99
- Urnenmodelle, 43–45
- Varianz, 174  
 diskret, 89–90  
 Eigenschaften, 174
- Verteilung  
 diskrete, 75  
 gemeinsame von  $U$  und  $V$ , diskret, 92  
 Wahrscheinlichkeits-,  $P^X$ , 52
- Verteilungsdichte, 63
- Verteilungsfunktion  
 der Zufallsgröße  $X$ , 55  
 diskret, 79  
 empirische, 257
- Vertrauensintervalle, 277
- Wahrscheinlichkeitsmaß, 24
- Wahrscheinlichkeitsraum, 25
- zentraler Grenzwertsatz  
 von Feller-Lévy, 241  
 von Lindeberg-Feller, 251  
 von Moivre-Laplace, lokal, 238
- zufälliger Vektor, 53  
 diskret, zweidimensional, 91  
 Funktionen diskreter, 94
- zufälliger Versuch, 9  
 mehrstufig, 106–115
- Zufallsgröße, 51  
 diskret, 78  
 reellwertige, 53  
 standardisiert, 175