TU Clausthal
Institut für Mathematik

Thomas Hanschke

Stochastik I

Prof. Dr. Thomas Hanschke
Institut für Mathematik
Technische Universität Clausthal
Erzstraße 1
38678 Clausthal-Zellerfeld

2. Auflage 2006 (7. November 2006)

Autor:	Prof. Dr. Thomas Hanschke
	(E-Mail: Hanschke@math.tu-clausthal.de)
Assistenz:	Dr. Michael Frank
	(E-Mail: Michael.Frank@dlh.de)
	Alexander Herzog
	(E-Mail: Herzog@math.tu-clausthal.de)
	Sylvia Arns
	(E-Mail: Sylvia.Arns@math.tu-clausthal.de)

Inhaltsverzeichnis

Vorwort **7**

Vorwort zur ersten Auflage . . . 7

Vorwort zur zweiten Auflage . . . 8

Auf Stochastik I aufbauende Vorlesungen . . . 9

1 Einführung und grundlegende Begriffe **11**

1.1 Zufallsexperiment . . . 12

1.2 Ergebnisraum . . . 13

1.3 Ereignisse . . . 13

1.4 Wahrscheinlichkeit . . . 15

Literaturverzeichnis . . . 18

2 Diskrete Wahrscheinlichkeitsverteilungen **21**

2.1 Konstruktion diskreter Wahrscheinlichkeitsmaße . . . 22

2.2 Laplacescher Wahrscheinlichkeitsraum . . . 23

2.3 Das Urnenmodell . . . 23

2.4 Beziehungen zwischen den Verteilungen . . . 28

2.5 Diskrete Zufallsvariable . . . 32

2.6 Kenngrößen einer diskreten Zufallsvariablen . . . 34

2.7 Erzeugende Funktion . . . 38

Literaturverzeichnis . . . 41

3 Bedingte Wahrscheinlichkeiten **43**

3.1 Begriffe und Zusammenhänge . . . 44

3.2 Stochastische Unabhängigkeit . . . 53

Literaturverzeichnis . . . 56

4 Statistische Methoden der Qualitätssicherung **59**

4.1 Hypothesentest . . . 60

4.2 Konstruktion von (n-c)-Stichprobenplänen . . . 62

4.3 Maximaler mittlerer Durchschlupf . . . 70

4.4 Mittlerer Prüfaufwand . . . 73

Literaturverzeichnis . . . 74

5 Mengensysteme **75**
5.1 Begriffe und Zusammenhänge 76
5.2 Erzeugendensysteme 82
5.3 Die σ–Algebra der Borelschen Mengen 85
Literaturverzeichnis 86

6 Mengenfunktionen **89**
6.1 Grundbegriffe 90
6.2 Erster Maß–Fortsetzungssatz 94
6.3 Zweiter Maß–Fortsetzungssatz 97
Literaturverzeichnis 103

7 Maßdefinierende Funktionen **105**
7.1 Korrespondenzsatz 106
7.2 Rechenregeln für maßdefinierende Funktionen 109
7.3 Beispiele für maßdefinierende Funktionen 111
Literaturverzeichnis 122

8 Messbare Abbildungen **125**
8.1 Messbare Abbildungen und Bildmaße 126
8.2 Kriterien für Messbarkeit 129
Literaturverzeichnis 133

9 Integration **135**
9.1 Vorbemerkungen 136
9.2 Das μ–Integral von Elementarfunktionen 137
9.3 Das μ–Integral nichtnegativer messbarer numerischer Funktionen 140
9.4 Das μ–Integral allgemeiner messbarer numerischer Funktionen 145
9.5 Erwartungswert und Varianz einer reellwertigen Zufallsvariable 149
9.6 Tabelle mit Kenngrößen verschiedener Verteilungen 165
9.7 Weitere Hilfsmittel aus der Maß– und Integrationstheorie 167
Literaturverzeichnis 171

10 Zuverlässigkeit **173**
10.1 Einführung und Grundbegriffe 174
10.2 Zuverlässigkeit von zusammengesetzten Systemen 179
Literaturverzeichnis 183

11 Produkträume und mehrdimensionale Zufallsvariable **185**
11.1 Zufällige Vektoren 186
11.2 Unabhängigkeit von Zufallsvariablen 198
11.3 Transformation von Zufallsvariablen (von stetigen Verteilungen) 206
11.4 Der Satz von Fubini und seine Anwendungen 211
Literaturverzeichnis 217

12 Schwaches und starkes Gesetz der großen Zahlen **219**
12.1 Das schwache Gesetz der großen Zahlen 220
12.2 Das starke Gesetz der großen Zahlen 223
Literaturverzeichnis 228

A Tabelle der χ^2–Verteilung **231**

B Zeichenerklärungen **233**

C Literatur **235**

D Historie **243**

Vorwort

Vorwort zur ersten Auflage

Die Wahrscheinlichkeitstheorie hat sich seit ihrer axiomatischen Begründung durch A. N. Kolmogorov im Jahre 1936 zu einem bedeutenden mathematischen Teilgebiet entwickelt und ist heute integraler Bestandteil der universitäten Ausbildung von Mathematikern. Die Wahrscheinlichkeitstheorie befasst sich mit der Abstraktion und der Modellierung von Zufallsvorgängen. Ihre Anfänge reichen weit in das 16. Jahrhundert zurück. Damals waren an den königlichen Höfen Glücksspiele in Mode und Gelehrte wie Pascal, Fermat und Huygens fragten nach den Gewinnchancen in diesen Spielen unter Zugrundelegung unterschiedlichster Spielregeln. Die Untersuchungen wurden fortgesetzt von J. Bernoulli, der das berühmte schwache Gesetz der großen Zahlen entdeckte, mit dem sich die asymptotische Stabilisierung der relativen Häufigkeit eines Zufallsereignisses mathematisch begründen lässt. Dieses Prinzip bildet die mathematische Vorlage für den axiomatischen Aufbau der Wahrscheinlichkeitstheorie. Das Konstruktionsprinzip für allgemeine Wahrscheinlichkeitsmaße und die damit verbundenen Existenzsätze verdanken wir C. Caratheodory. Die moderne Wahrscheinlichkeitstheorie ist durch einen starken Anwendungsbezug geprägt. Ob es sich um Fragen der Spracherkennung, der Produktionsplanung, der Bewertung von Optionspreismodellen, der Ausbreitung infektiöser Krankheiten oder der Entschlüsselung genetischer Codes handelt, überall ist Stochastik im Spiel. Der besondere Reiz der Stochastik liegt darin, dass in ihr viele mathematische Disziplinen wie Maßtheorie, Funktionalanalysis, Operatortheorie, Funktionentheorie und diskrete Mathematik zusammengeführt werden.

Das Online–Skriptum zur Stochastik ist aus Vorlesungen hervorgegangen, die ich seit 1993 regelmäßig an der Technischen Universität Clausthal abhalte. Eine Besonderheit in Clausthal ist, dass Mathematik– und Informatik–Studenten ein gemeinsames Grundstudium durchlaufen und die Wirtschaftsmathematiker und Wirtschaftsinformatiker diesselben Stochastik–Vorlesungen hören. Ich halte es für eine gute Idee, die Zusammenarbeit von Mathematikern und Informatikern bereits in der Ausbildung zu fördern. Denn Mathematik und Informatik bedingen einander in besonderer Weise. Die gemeinsame Ausbildung setzt natürlich die Bereitschaft voraus, sich für das jeweils andere Fach zu begeistern. Indem wir in besonderem Maße auf Anwendungen und innermathematische Zusammenhänge eingehen, versuchen wir, beiden Gruppen mit der Vorlesung gerecht zu werden. Da wir am eigentlichen und vorwiegend Maßtheorie–orientierten Vorlesungsstoff keine Abstriche vornehmen wollen, bieten wir in Form verschiedener Applets zusätzliche Erklärungs– und Lernhilfen an. Hierin sehen wir einen besonderen Vorteil der eLearning–Technologien.

Die Technologie für das Online–Skriptum haben mein langjähriger Mitarbeiter Dr. Michael

Frank und meine ehemaligen Diplomanden Torsten Hiddessen und Thomas Rosenau entwickelt. Die technische Weiterentwicklung des Systems ist durch die finanzielle Förderung im Rahmen von ELAN (elearning academic network Niedersachsen) und durch das besondere Engagement meines Mitarbeiters Dipl. Math. Alexander Herzog sichergestellt. Korrektur gelesen und Verbesserungsvorschläge eingebracht haben Dipl. Math. Sylvia Arns, Dipl. Math. Alexander Herzog, Dipl. Math. Susanne Lühr und unsere Hilfsassistenten Thomas Riemer und Henning Schmidt. Die Applets wurden von Thorsten Hiddessen und Volker Hein programmiert. Die grafische Gestaltung stammt von Franzika Dannehl. Aber auch meinem Kollegen, Prof. Dr. Joachim Hilgert, der nach einem Prototypen des Online-Skriptums im WS 01/02 und im SS 02 in Clausthal die Stochastik gelesen hat, verdanke ich wichtige Korrekturhinweise und Anmerkungen. Ihnen allen sei hiermit mein herzlichster Dank ausgesprochen.

Thomas Hanschke Clausthal, Oktober 2004

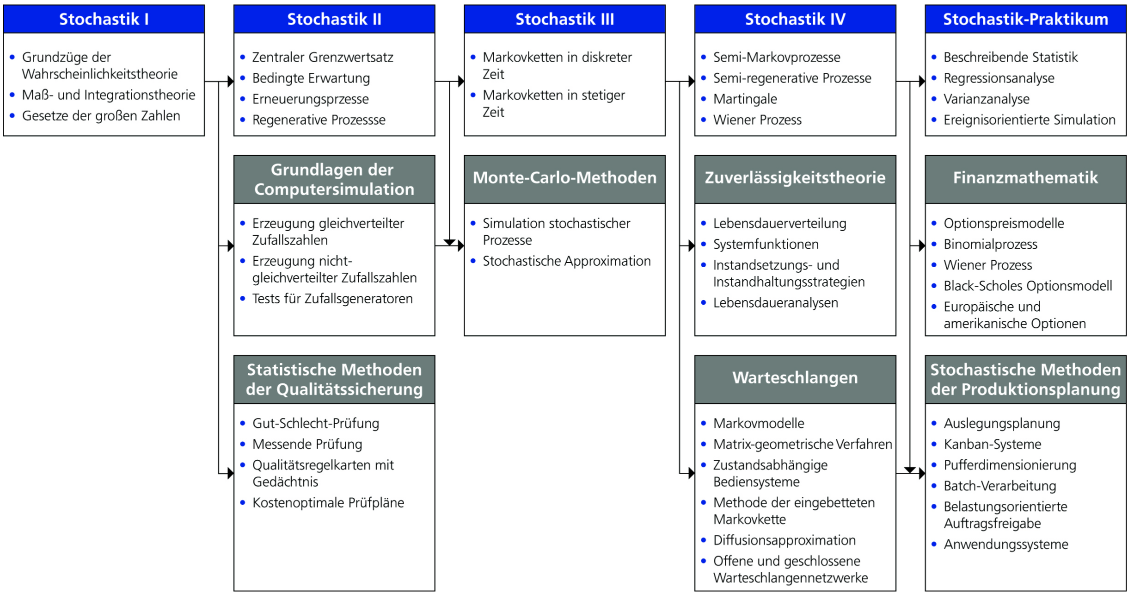
Vorwort zur zweiten Auflage

Die zweite Auflage unterscheidet sich von der ersten im Wesentlichen durch einige strukturelle Änderungen und die Hinzunahme weiterer Illustrationen. Bei der Durchführung halfen diesmal besonders Dipl.-Math. Sylvia Arns, Dipl.-Math. Alexander Herzog und unser Hilfsassistent Hendrik Baumann.

Thomas Hanschke Clausthal, Oktober 2006

Auf Stochastik I aufbauende Vorlesungen

Die Stochastik I Vorlesung stellt die Basis für eine Reihe weiterführender Veranstaltungen dar, die man der nachstehenden Grafik entnehmen kann.



(Siehe auch PowerPoint-Präsentation zum Stochastik–Vorlesungsplan.)

Kapitel 1

Einführung und grundlegende Begriffe

In diesem Kapitel werden die Objekte erläutert, mit denen Zufallsexperimente mathematisch beschrieben werden. Außerdem wird das von A.N. Kolmogorov stammende Axiomensystem der Wahrscheinlichkeitstheorie vorgestellt.

Schlüsselwörter: Zufallsexperiment, Elementarereignis, Stichprobenraum, Ereignis, Rechenregeln für Ereignisse, absolute und relative Häufigkeit, Wahrscheinlichkeit, Axiomensystem nach Kolmogorov.

1.1 Zufallsexperiment

Die Stochastik stellt Methoden und Verfahren zur Beschreibung und Analyse von Zufallsvorgängen zur Verfügung.

Stochastische Problemstellungen sind etwa:

- Aussagen über die Genauigkeit von Messergebnissen
- Vergleich der Verträglichkeit von Medikamenten
- Zuverlässigkeit technischer Systeme
- Gewinnchancen bei Glücksspielen
- Aussagekraft von Meinungsumfragen, Hochrechnungen
- Populationsprozesse, Ausbreitung infektiöser Krankheiten
- Vererbung von Eigenschaften

Der vermeintliche Zufall ist allgegenwärtig: In einem Produktionsprozess fallen unvorhergesehen Maschinen aus und führen zu Staus auf den Transportbändern. Die Fußball–Europa–Meisterschaft fiel anders aus, als von den Experten prognostiziert wurde. Trotz akribisch dokumentierter Permanenzen eines Roullette–Tisches bleibt der Ausgang des nächsten Spiels ungewiss. Doch auch eine algorithmisch erzeugte Zahlenfolge werden wir als zufällig erachten, solange wir ihr Bildungsgesetz nicht durchschaut haben. Angesichts dieser Beispiele mögen Zweifel aufkommen, ob Begriffe wie ,,Zufall” und ,,Wahrscheinlichkeit” überhaupt mathematisch objektivierbar sind oder mangels übergeordneter Erkenntnise lediglich Ausdruck einer subjektiv empfundenen Unsicherheit sind. Die Stochastik will und kann diesen eher philosophischen Sachverhalt nicht aufklären, sondern versucht durch geeignete Abstraktion und Modellbildung, Gesetzmäßigkeiten in Zufallserscheinungen zu erkennen und mathematisch zu erfassen. In der Regel beschränkt sich die Stochastik deshalb auch auf solche Zufallsexperimente, die analog einem physikalischen Experiment reproduzierbar und durch hinreichend lange Beobachtung geeignet überprüft werden können.

1.1 Definition (Zufallsexperiment):
Unter einem Zufallsexperiment versteht man einen, im Prinzip beliebig oft, wiederholbaren Vorgang mit ungewissem Ausgang.

1.2 Beispiel:
Zufallsexperimente in diesem Sinne sind

- der einmalige Wurf einer Münze,
- die Wartezeit am Postschalter oder
- die Gewinnausschüttung an einem Spielautomaten.

Keine Zufallsexperimente in diesem Sinne sind

- der Ausgang der nächsten Bundestagswahl,
- die Niederschlagsmenge in Clausthal–Zellerfeld am 20. Oktober 2009.

1.2 Ergebnisraum

Zu Beginn eines Zufallsexperiments muss festgelegt werden, welche Resultate (Elementarereignisse) für die Untersuchtung relevant sind und im mathematischen Modell berücksichtigt werden sollen. Der zugrundegelegte Ergebnisraum bestimmt gewissermaßen den Detaillierungsgrad der Modellbildung.

1.3 Definition (Elementarereignisse, Ergebnisraum, Stichprobenmenge):
Die konkreten Ergebnisse eines Zufallsexperimentes heißen Elementarereignisse. Die Menge aller Elementarereignisse wird Ergebnisraum, Ergebnismenge oder auch Stichprobenmenge genannt und mit Ω bezeichnet.

1.4 Beispiel:

- Das Werfen eines Würfels: $\Omega := \{1, 2, 3, 4, 5, 6\}$.
- Die Wartezeit an einer Ampel: $\Omega := \{\omega \in \mathbb{R} \mid \omega \geq 0\}$.
- Der Betriebszustand von n Maschinen, defekt ($\equiv 1$) oder intakt ($\equiv 0$):
  $\Omega := \{(\omega_1, \ldots, \omega_n) \mid \omega_i \in \{0, 1\}, i = 1, \ldots, n\}$.

1.3 Ereignisse

Aus den Elementarereignissen lassen sich kompliziertere Ereignisse zusammensetzen. Im Fall des Würfelspiels kann man z.B. auch das Ereignis A, eine ungerade Augenzahl zu würfeln, betrachten. Man schreibt dann $A = \{1, 3, 5\}$ und sagt, dass das Ereignis A eingetreten ist, wenn der beobachtete Versuchsausgang in A liegt. Deswegen definiert man:

1.5 Definition (Ereignis):
Wenn Ω diskret ist, d.h. endlich oder höchstens abzählbar unendlich ist, so heißen die Teilmengen der Stichprobenmenge Ω Ereignisse. Man sagt, dass das Ereignis A eingetreten ist, wenn der beobachtete Ausgang des Zufallsexperimentes in A liegt ($\omega \in A$). Ist Ω diskret, dann ist die Menge aller Ereignisse gerade die Potenzmenge $\mathfrak{P}(\Omega)$ von Ω.

1.6 Beispiel:

- Das Ereignis „gerade Augenzahl“ beim Würfeln: $A := \{2, 4, 6\}$.
- Das Ereignis, dass mindestens 2 von n Maschinen defekt sind:
  $A := \{\omega \in \Omega \mid \omega_1 + \cdots + \omega_n \geq 2\}$.

Mit Hilfe der mengentheoretischen Grundoperationen lassen sich weitere Ereignisse bilden.

1.7 Bemerkung (Interpretation von Ereignissen):
Es seien $A, B, A_1, A_2, \ldots$ Ereignisse, dann gilt:

$$
\begin{array}{ccl}
\Omega & := & \text{„sicheres Ereignis, das immer eintritt“}, \\
\varnothing & := & \text{„unmögliches Ereignis, das nie eintreten kann“}, \\
A \cup B & := & \text{„}A\text{ oder }B\text{ treten ein“}, \\
A \cap B & := & \text{„}A\text{ und }B\text{ treten ein“}, \\
A \setminus B & := & \text{„}A\text{, aber nicht }B\text{, tritt ein“; kurz: „}A\text{ ohne }B\text{“}, \\
\overline{A} & := & \text{„}A\text{ tritt nicht ein“}, \\
\bigcup_{n\in\mathbb{N}} A_n & := & \text{„mindestens ein }A_n\text{ tritt ein“}, \\
\bigcap_{n\in\mathbb{N}} A_n & := & \text{„alle }A_n\text{ treten ein“}.
\end{array}
$$

1.8 Bemerkung (Übersicht über die Eigenschaften von Mengen):
Für allgemeine Mengen gelten folgende Beziehungen:

1. Komplementbildung:
$\overline{\overline{A}} = A, \quad \overline{\Omega} = \varnothing, \quad \overline{\varnothing} = \Omega.$

2. Die Durchschnittsbildung ist kommutativ und assoziativ:
$A \cap B = B \cap A$, $A \cap (B \cap C) = (A \cap B) \cap C, \quad A \cap A = A, \quad A \cap \Omega = A, \quad A \cap \varnothing = \varnothing,$ $A \cap \overline{A} = \varnothing.$

3. Die Vereinigungsbildung ist kommutativ und assoziativ:
$A \cup B = B \cup A$, $A \cup (B \cup C) = (A \cup B) \cup C$, $A \cup \overline{A} = \Omega$, $A \cup A = A$, $A \cup \Omega = \Omega$, $A \cup \varnothing = A.$

4. Es gelten die Distributivgesetze:
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$

5. Absorptionsgesetze:
$A \cup (A \cap B) = A, \quad A \cap (A \cup B) = A.$

6. Reziprozitätsgesetze:
$\overline{A \cap B} = \overline{A} \cup \overline{B}, \quad \overline{A \cup B} = \overline{A} \cap \overline{B}.$

7. $A \subset B \Longleftrightarrow \overline{A} \supset \overline{B}.$

8. Die Regeln von de Morgan:
$\overline{\bigcup_{n \in \mathbb{N}} A_n} = \bigcap_{n \in \mathbb{N}} \overline{A_n}, \quad \overline{\bigcap_{n \in \mathbb{N}} A_n} = \bigcup_{n \in \mathbb{N}} \overline{A_n}.$

1.9 Bemerkung (*σ*–Algebra):
Im Falle eines beliebigen Stichprobenraumes können aus logischen Gründen nicht alle Teilmengen von Ω als Ereignisse zugelassen werden. Im Falle einer allgemeinen Ergebnismenge Ω ist es daher notwendig, sich auf kleinere Mengensysteme als $\mathfrak{P}(\Omega)$ zurückziehen, die aber hinsichtlich ihrer Verknüpfungsstruktur noch universell genug sind, um alle im Experiment enthaltenen Möglichkeiten reflektieren zu können. An die Stelle der Potenzmenge tritt ein Mengensystem $\mathfrak{F} \subseteq \mathfrak{P}(\Omega)$ mit folgenden Eigenschaften:

(i) $\Omega \in \mathfrak{F}$,

(ii) für $A \in \mathfrak{F}$ ist $\overline{A} \in \mathfrak{F}$,

(iii) für jede Folge $(A_n)_{n \in \mathbb{N}}$ von Elementen aus $\mathfrak{F}$ gilt: $\bigcup_{n \in \mathbb{N}} A_n \in \mathfrak{F}$.

$\mathfrak{F}$ nennt man σ–Algebra über Ω. $\mathfrak{F}$ wird sich später als der natürliche Definitionsbereich von Wahrscheinlichkeitsmaßen herausstellen.

Man erkennt sofort, dass aufgrund der Forderung (ii) mit Ω auch die leere Menge $\varnothing$ und aufgrund der Regeln von de Morgan mit $\cup_{n \in \mathbb{N}} A_n$ auch $\cap_{n \in \mathbb{N}} A_n$ zu $\mathfrak{F}$ gehören. Damit ergibt sich, dass $\mathfrak{F}$ bzgl. aller wesentlichen Mengenoperationen abgeschlossen ist und alle wesentlichen Ereignisse enthalten sollte.

Das Ziel ist es nun, jedem Ereignis $A \in \mathfrak{F}$ eine Maßzahl $P(A)$ zuzuordnen, die angibt, welche Chance A hat, bei einem Zufallsexperiment einzutreten.

1.4 Wahrscheinlichkeit

Ein naheliegender Weg, den Begriff der Wahrscheinlichkeit zu definieren, ist der folgende: Das Zufallsexperiment wird n–mal unter gleichen Bedingungen durchgeführt. Dabei wird beobachtet, wie oft das Ereignis A eingetreten ist.

1.10 Definition (absolute Häufigkeit, relative Häufigkeit):
Für jedes Ereignis A wird mit $H(A) := H_n(A)$ gezählt, wie oft es in einer Versuchsreihe der Länge n aufgetreten ist. $H(A)$ wird absolute Häufigkeit und

$$h(A) := h_n(A) := \frac{H(A)}{n}$$

relative Häufigkeit des Ereignisses A in einer Versuchsreihe der Länge n genannt.

Auf den ersten Blick nachteilig ist, dass $h(A)$ von n und der jeweiligen Versuchsreihe mit deren Bedingungen abhängt. Die Erfahrung zeigt jedoch, dass sich die relativen Häufigkeiten mit größer werdendem Stichprobenumfang stabilisieren.

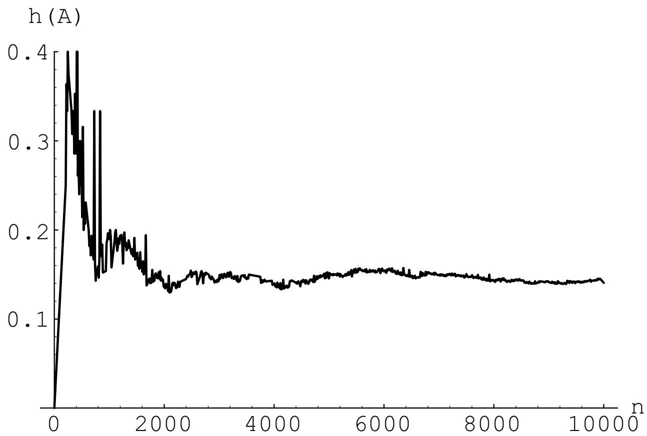


Abbildung 1.1: Stabilisierung von relativen Häufigkeiten

1.11 Beispiel:

- Als Zufallsexperiment wird das Werfen eines Würfels betrachtet. In Abhängigkeit von der Versuchslänge n ergibt sich für die relative Häufigkeit des Ereignisses A:=„Augenzahl 6 gewürfelt“, der Graph aus Abbildung 1.1. Es ist deutlich zu erkennen, dass sich $h(A)$ um den Wert 1/6 stabilisiert.

- Das Geschlecht von Neugeborenen: A:=„männlich“: $h(A) \xrightarrow[n\to\infty]{} 0.514$.

Man überprüft leicht folgende Eigenschaften absoluter und relativer Häufigkeiten:

a) Nichtnegativität: $H(A) \geq 0$, d.h. $h(A) \geq 0$,

b) Normiertheit: $H(\Omega) = n$, d.h. $h(\Omega) = 1$,

c) Additivität: Für $A \cap B = \varnothing$ gilt $H(A \cup B) = H(A) + H(B)$ d.h. $h(A \cup B) = h(A) + h(B)$.

Das auf A.N. Kolmogorov zurückgehende Axiomensystem der Wahrscheinlichkeitstheorie orientiert sich an diesen Eigenschaften relativer Häufigkeiten, wobei aus mathematischen Gründen die Eigenschaft der Additivität durch die Eigenschaft der sogenannten σ-Additivität ersetzt wird.

1.12 Definition (Axiomensystem nach Kolmogorov):
Gegeben sei eine σ–Algebra $\mathfrak{F}$ über einem Stichprobenraum $\Omega \neq \varnothing$. Jede Abbildung $P\colon \mathfrak{F} \to \mathbb{R}$ mit den Eigenschaften

1. *Nichtnegativität:* $\forall\ A \in \mathfrak{F} : P(A) \geq 0$
2. *Normiertheit:* $P(\Omega) = 1$
3. *P ist σ–additiv, d.h. sei $A_1, A_2, \ldots$ eine Folge von paarweise unvereinbaren Ereignissen ($A_i \cap A_j = \varnothing$ für $i \neq j$), dann gilt:*

$$P\left(\bigcup_{j\in\mathbb{N}} A_j\right) = \sum_{j\in\mathbb{N}} P(A_j)$$

heißt Wahrscheinlichkeitsmaß auf $\mathfrak{F}$ bzw. Ω.
Das Tupel $(\Omega, \mathfrak{F}, P)$ heißt Wahrscheinlichkeitsraum über Ω.
(Siehe auch Lebensdaten von Kolmogorov im Anhang D.)

Folgerung:
Aus dem Axiomensystem ergibt sich unmittelbar:

4. $P(\varnothing) = 0$
5. (Additivität) $A \cap B = \varnothing \Longrightarrow P(A \cup B) = P(A) + P(B)$
6. $P(\overline{A}) = 1 - P(A)$
7. (Monotonie) $A \subseteq B \Longrightarrow P(A) \leq P(B)$
   Aus dieser Eigenschaft folgt insbesondere $P(A) \leq 1$ für alle $A \in \mathfrak{F}$, da stets $A \subseteq \Omega$ gilt.
8. (Additionssatz) Für beliebige $A, B \in \mathfrak{F}$ gilt
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$
9. (Formel von Sylvester–Poincaré) In Erweiterung zum Additionssatz gilt für beliebige $A_1, \ldots, A_n \in \mathfrak{F}$:
$$\begin{aligned} P\left(\bigcup_{k=1}^{n} A_k\right) &= \sum_{k=1}^{n} P(A_k) - \sum_{\substack{k_1,k_2=1\\ k_1<k_2}}^{n} P\left(A_{k_1} \cap A_{k_2}\right) + \sum_{\substack{k_1,k_2,k_3=1\\ k_1<k_2<k_3}}^{n} P\left(A_{k_1} \cap A_{k_2} \cap A_{k_3}\right) \\ &\quad + \cdots + (-1)^{n-1} \cdot P(A_1 \cap \cdots \cap A_n). \end{aligned}$$

Beweis:

4. Es gilt: $1 = P(\Omega) = P(\Omega \cup \varnothing \cup \cdots \cup \varnothing \cup \ldots) \stackrel{3)}{=} P(\Omega) + P(\varnothing) + \cdots \Rightarrow P(\varnothing) = 0.$
5. Aussage 5 folgt unmittelbar aus Aussage 3.

6. Die Behauptung folgt unmittelbar aus der Additivität und der Zerlegung $\Omega = A \cup \overline{A}$.

7. Die Behauptung folgt unmittelbar aus der Additivität und der Zerlegung $B = A \cup (B \backslash A)$.

8. Es gilt $A \cup B = A \cup (B \backslash A)$, $A \cap (B \backslash A) = \varnothing$, $(A \cap B) \cup (B \backslash A) = B$ und $(A \cap B) \cap (B \backslash A) = \varnothing$, d.h.
$$P(A \cup B) = P(A \cup (B \backslash A)) = P(A) + P(B \backslash A).$$
Da $P(B) = P((A \cap B) \cup (B \backslash A)) = P(A \cap B) + P(B \backslash A)$ äquivalent ist zu $P(B \backslash A) = P(B) - P(A \cap B)$, folgt insgesamt
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

9. Es gilt zunächst
$$\begin{aligned} P(A_1 \setminus A_2) &= P(A_1 \setminus A_2) + P(A_1 \cap A_2) - P(A_1 \cap A_2) \\ &\overset{3.)}{=} P((A_1 \setminus A_2) \cup (A_1 \cap A_2)) - P(A_1 \cap A_2) \;=\; P(A_1) - P(A_1 \cap A_2). \end{aligned}$$
Der Beweis der Formel von Sylvester–Poincaré erfolgt durch Induktion. Für $n = 2$ ist die Aussage mit dem Additionssatz identisch. Für $n > 2$ gilt:
$$\begin{aligned} P\left(\cup_{k=1}^{n} A_k\right) &= P\left(\left(\left(\bigcup_{k=1}^{n-1} A_k\right) \setminus A_n\right) \bigcup \left(A_n \setminus \bigcup_{k=1}^{n-1} A_k\right) \bigcup \left(\left(\bigcup_{k=1}^{n-1} A_k\right) \cap A_n\right)\right) \\ &\overset{3.)}{=} P\left(\left(\left(\bigcup_{k=1}^{n-1} A_k\right) \setminus A_n\right) \bigcup \left(\left(\bigcup_{k=1}^{n-1} A_k\right) \cap A_n\right)\right) + P\left(A_n \setminus \bigcup_{k=1}^{n-1} A_k\right) \\ &= P\left(\bigcup_{k=1}^{n-1} A_k\right) + P\left(A_n \setminus \bigcup_{k=1}^{n-1} A_k\right) + P\left(\left(\bigcup_{k=1}^{n-1} A_k\right) \cap A_n\right) \\ &\quad -P\left(\left(\bigcup_{k=1}^{n-1} A_k\right) \cap A_n\right) \\ &= P\left(\bigcup_{k=1}^{n-1} A_k\right) + P(A_n) - P\left(\left(\bigcup_{k=1}^{n-1} A_k\right) \cap A_n\right) \\ &= P\left(\bigcup_{k=1}^{n-1} A_k\right) + P(A_n) - P\left(\bigcup_{k=1}^{n-1} (A_k \cap A_n)\right) \\ &\overset{IV}{=} \underbrace{\textstyle\sum_{k=1}^{n-1} P(A_k) - \sum_{\substack{k_1,k_2=1 \\ k_1<k_2}}^{n-1} P(A_{k_1} \cap A_{k_2}) + \ldots + (-1)^{n-2} P(A_1 \cap A_{n-1})}_{=P\left(\bigcup_{k=1}^{n-1} A_k\right)} + P(A_n) \\ &\quad - \underbrace{\left(\textstyle\sum_{k=1}^{n-1} P(A_k \cap A_n) - \sum_{\substack{k_1,k_2=1 \\ k_1<k_2}}^{n-1} P(A_{k_1} \cap A_{k_2} \cap A_n) + \ldots + (-1)^{n-2} P(A_1 \cap A_{n-1} \cap A_n)\right)}_{=P\left(\bigcup_{k=1}^{n-1} A_k \cap A_n\right)} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{n} P(A_k) \underbrace{- \sum_{\substack{k_1,k_2=1\\k_1<k_2}}^{n-1} P(A_{k_1}\cap A_{k_2}) - \sum_{k=1}^{n-1} P(A_k \cap A_n)}_{=-\sum_{\substack{k_1,k_2=1\\k_1<k_2}}^{n} P(A_{k_1}\cap A_{k_2})} + \ldots + \\
&\quad \underbrace{(-1)^{n-2}P(A_1\cap\ldots\cap A_{n-1}) - (-1)^{n-3}\sum_{k=1}^{n-1} P(A_1\cap\ldots\cap A_{k-1}\cap A_{k+1}\cap\ldots\cap A_{n-1}\cap A_n)}_{=(-1)^{n-2}\sum_{k=1}^{n} P(A_1\cap A_{k-1}\cap A_{k+1}\cap\ldots\cap A_n)} + \\
&\quad \underbrace{-(-1)^{n-2}P(A_1\cap\ldots\cap A_n)}_{=(-1)^{n-1}P(A_1\cap\ldots\cap A_n)} \\
&= \sum_{k=1}^{n} P(A_k) - \sum_{\substack{k_1,k_2=1\\k_1<k_2}}^{n} P\left(A_{k_1}\cap A_{k_2}\right) + \sum_{\substack{k_1,k_2,k_3=1\\k_1<k_2<k_3}}^{n} P\left(A_{k_1}\cap A_{k_2}\cap A_{k_3}\right) \\
&\quad + \cdots + (-1)^{n-1}\cdot P(A_1\cap\cdots\cap A_n).
\end{aligned}
$$

(Siehe auch PowerPoint-Präsentation zu diesem Beweis.)

■

Literatur zu Kapitel 1

Folgende Bücher werden als begleitende Literatur zu diesem Kapitel des Skriptes empfohlen:

- M. Fisz:
  Wahrscheinlichkeitsrechnung und mathematische Statistik,
  VEB, Deutscher Verlag der Wissenschaften 1989.
  ISBN: 3326000790

- P. Gänssler und W. Stute:
  Wahrscheinlichkeitstheorie,
  Springer–Verlag, Berlin, 1977.
  ISBN: 3540084185

- H. O. Georgii:
  Stochastik,
  2. Auflage, de Gruyter, 2004.
  ISBN: 3110172356

- K. Hinderer:
  Grundbegriffe der Wahrscheinlichkeitstheorie,
  Springer–Verlag, 1980.
  ISBN: 3540073094

- G. Hübner:
  Stochastik. Eine Einführung für Mathematiker, Informatiker und Ingenieure.,
  4. Auflage, Vieweg Verlag, 2003.
  ISBN: 3528254432

- U. Krengel:
  Einführung in die Wahrscheinlichkeitstheorie und Statistik,
  6. Auflage, Vieweg, 2002.
  ISBN: 3528672595
- J. Lehn/H. Wegmann:
  Einführung in die Statistik,
  4. Auflage, Teubner, 2004.
  ISBN: 3519320711
- R. Mathar/D. Pfeiffer:
  Stochastik für Informatiker,
  Teubner, 1990.
  ISBN: 3519022400

Kapitel 2

Diskrete Wahrscheinlichkeitsverteilungen

In diesem Kapitel werden anhand einfacher diskreter Zufallsexperimente die Grundprinzipien stochastischer Modellbildung und die Konstruktion von Wahrscheinlichkeitsmaßen erläutert.

Schlüsselwörter: Konstruktion diskreter Wahrscheinlichkeitsmaße, Wahrscheinlichkeitsvektor, Laplace–Experiment, Urnenmodelle, Hypergeometrische Verteilung, Binomialverteilung, Poisson–Verteilung, diskrete reellwertige Zufallsvariable, Bildmaß, Verteilung einer Zufallsvariablen, Kenngrößen einer diskreten Zufallsvariablen, Erwartungswert, Varianz, k–tes Moment, k–tes zentrales Moment, erzeugende Funktion.

2.1 Konstruktion diskreter Wahrscheinlichkeitsmaße

In der Einführung wurde festgestellt, dass σ–Algebren den natürlichen Definitionsbereich von Wahrscheinlichkeitsmaßen darstellen. Im Fall eines diskreten Zufallsexperimentes bietet es sich an, als Ereignisalgebra die Potenzmenge $\mathfrak{P}(\Omega)$ über Ω zu verwenden. Da selbst bei einem endlichen Wahrscheinlichkeitsraum die Mächtigkeit der Potenzmenge in Abhängigkeit von der Mächtigkeit des Ergebnisraumes sehr rasch anwächst, ist es meist schwierig, $P(A)$ für alle $A \subseteq \Omega$ explizit anzugeben. Im Fall des Würfelspiels hatte sich allerdings herausgestellt, dass $P(A)$ bereits durch die Werte $p_\omega := P(\{\omega\})$ für alle $\omega \in \Omega$ festgelegt ist. Dieser Sachverhalt soll jetzt näher untersucht werden.

2.1 Definition (Wahrscheinlichkeitsvektor, diskreter Wahrscheinlichkeitsraum):
Es sei Ω eine diskrete Stichprobenmenge. Ein Vektor $p := (p_\omega)_{\omega\in\Omega}$ mit den Eigenschaften

(i) $p_\omega \geq 0$ *für alle* $\omega \in \Omega$

(ii) $\sum\limits_{\omega\in\Omega} p_\omega = 1$ *(d.h. die Reihe ist (abs.) konvergent gegen 1)*

heißt Wahrscheinlichkeitsvektor über Ω. Das Tupel (Ω, p) wird auch diskreter Wahrscheinlichkeitsraum genannt.

2.2 Satz:
Es sei Ω eine diskrete Stichprobenmenge und $p := (p_\omega)_{\omega\in\Omega}$ ein Wahrscheinlichkeitsvektor über Ω.

a) Ist P ein Wahrscheinlichkeitsmaß über $(\Omega, \mathfrak{P}(\Omega))$, so wird durch

$$p_\omega := P(\{\omega\}) \quad \forall\, \omega \in \Omega$$

ein Wahrscheinlichkeitsvektor p über Ω definiert.

b) Ist $p := (p_\omega)_{\omega\in\Omega}$ ein Wahrscheinlichkeitsvektor über Ω, so gibt es genau ein Wahrscheinlichkeitsmaß P über $(\Omega, \mathfrak{P}(\Omega))$ mit

$$p_\omega = P(\{\omega\}) \quad \forall\, \omega \in \Omega.$$

Damit ist die Zuordnung $P \mapsto (p_\omega)_{\omega\in\Omega}$ eine Bijektion zwischen Wahrscheinlichkeitsmaßen und Wahrscheinlichkeitsvektoren.

Beweis:

a) Da $P(A) \geq 0$ ist für alle $A \in \mathfrak{P}(\Omega)$, gilt natürlich auch $p_\omega := P(\{\omega\}) \geq 0$ für alle $\omega \in \Omega$. Aus $\Omega = \bigcup_{\omega\in\Omega}\{\omega\}$ folgt, dass

$$\sum_{\omega\in\Omega} p_\omega = \sum_{\omega\in\Omega} P(\{\omega\}) \overset{\sigma-add}{=} P\left(\bigcup_{\omega\in\Omega}\{\omega\}\right) = P(\Omega) = 1.$$

b) Wir definieren $P(A) := \sum_{\omega\in A} p_\omega$ für alle $A \subseteq \Omega$. Offensichtlich ist $P(A) \geq 0$ für alle $A \subseteq \Omega$ und $P(\Omega) = \sum_{\omega\in\Omega} p_\omega = 1$. Für paarweise disjunkte Mengen gilt außerdem

$$P\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{\omega\in\bigcup_{i\in\mathbb{N}} A_i} p_\omega \overset{abs.Konv.}{=} \sum_{i\in\mathbb{N}}\sum_{\omega\in A_i} p_\omega \overset{abs.Konv.}{=} \sum_{i\in\mathbb{N}} P(A_i),$$

d.h. P ist auch σ–additiv. Mit der σ-Additivität ergibt sich auch die Eindeutigkeit der Definition von P:

$$P(A) = P\left(\bigcup_{\omega \in A} \omega\right) \overset{\sigma-add.}{=} \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} p_\omega.$$

2.2 Laplacescher Wahrscheinlichkeitsraum

Bei der Betrachtung der Wahrscheinlichkeit für das Auftreten einer bestimmten Augenzahl beim Würfelspiel stellt man fest, dass alle Elementarereignisse mit der annähernd gleichen relativen Häufigkeit auftreten. Dieser Sachverhalt gibt Anlass zu folgender Definition:

2.3 Definition (Laplacescher Wahrscheinlichkeitsraum):
Ein endlicher Wahrscheinlichkeitsraum (Ω, P) heißt Laplacescher Wahrscheinlichkeitsraum, falls für alle $\omega \in \Omega$

$$p_\omega = \frac{1}{|\Omega|}$$

gilt. P heißt Gleichverteilung oder Laplacesche Wahrscheinlichkeitsverteilung über Ω. Die Wahrscheinlichkeit eines Ereignisses $A \subset \Omega$ ist dann:

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

(Siehe auch Lebensdaten von Laplace im Anhang D.)

2.4 Beispiel (Spiel mit zwei Würfeln):
Beim gleichzeitigen Spiel mit zwei Würfeln ergibt sich als Ergebnisraum

$$\Omega := \{(\omega_1, \omega_2) \mid \omega_i \in \{1, \ldots, 6\}; i = 1, 2\}.$$

Es gilt $|\Omega| = 36$. Das Ereignis A:=„Die Summe der Augenzahlen ist 7“ entspricht der Menge

$$A = \{(\omega_1, \omega_2) \mid \omega_1 + \omega_2 = 7\} = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}.$$

Es ist $|A| = 6$. Unter der Annahme eines Laplaceschen Wahrscheinlichkeitsraums folgt

$$P(A) = \frac{|A|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}.$$

2.3 Das Urnenmodell

In einer Urne befinden sich gut durchmischt N Kugeln. Darunter sind R schwarze und $N - R$ weiße Kugeln. Es werden der Urne zufällig n Kugeln entnommen. Es stellt sich die Frage:

> Wie groß ist die Wahrscheinlichkeit, dass sich unter den n gezogenen Kugeln genau k ($k \leq n$, $k \leq R$) schwarze befinden?

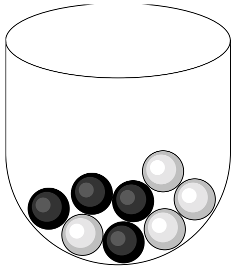


Abbildung 2.1: Urnenmodell

Ziehen ohne Zurücklegen

Offensichtlich lassen sich die Kugeln auf zwei verschiedene Arten ziehen und wieder zurücklegen (Ziehen mit Zurücklegen und Ziehen ohne Zurücklegen). Zuerst wird das Ziehen ohne Zurücklegen betrachtet, d.h. es wird davon ausgegangen, dass die Kugeln nacheinander gezogen und außerhalb der Urne belassen werden. Zur Lösung des Problems wird angenommen, dass die R schwarzen Kugeln von 1 bis R und die $N-R$ weißen Kugeln von $R+1$ bis N durchnumeriert seien:

$$\begin{array}{ccccc|ccc} \bullet & \bullet & \bullet & \dots & \bullet & \circ & \dots & \circ \\ 1 & 2 & 3 & \dots & R & R+1 & \dots & N \\ \multicolumn{5}{c|}{\text{schwarze Kugeln}} & \multicolumn{3}{c}{\text{weiße Kugeln}} \end{array}$$

Der Ergebnisraum ist dann

$$\Omega := \{(\omega_1, \dots, \omega_n) | \omega_j \in \{1, \dots, N\}, 1 \le j \le n, \omega_i \ne \omega_j \text{ für } i \ne j\}.$$

Um die <u>Mächtigkeit</u> von Ω, d.h. die Anzahl $|\Omega|$ der Elemente von Ω zu bestimmen, wird angenommen, dass die Kugeln in n Kästchen abgelegt seien.

$$\square\square\square\square\square \dots \square$$

Die jeweilige Anzahl der Möglichkeiten, die n Kästchen zu belegen, ergibt sich wie folgt:

1.	Kästchen:	N Möglichkeiten
2.	Kästchen:	$N-1$ Möglichkeiten
$\vdots$		
n.	Kästchen:	$N-(n-1)$ Möglichkeiten

Da sich die Möglichkeiten von Kästchen zu Kästchen multiplizieren, ergibt sich

$$|\Omega| = (N)_n := N \cdot (N-1) \cdot \dots \cdot (N-(n-1)) = \frac{N!}{(N-n)!}.$$

$(N)_n$ heißt <u>n–te untere Faktorielle</u> von N.

Das Ereignis A_k, dass sich unter den n gezogenen Kugeln genau k schwarze befinden, besteht aus allen n–Tupeln $(\omega_1, \dots, \omega_n)$, für die genau k Komponenten kleiner oder gleich R sind. Zunächst gibt es genau

$$\frac{n \cdot (n-1) \cdots (n-k+1)}{k \cdot (k-1) \cdots 2 \cdot 1} =: \binom{n}{k}$$

Möglichkeiten, die k schwarzen Kugeln auf die n Kästchen zu verteilen, denn: Für die erste Kugel gibt es n mögliche Kästchen, für die zweite $n-1$ usw. und schließlich für die k–te $n-k+1$ mögliche Kästchen. Da die schwarzen Kugeln untereinander nicht unterscheidbar sind, kommt es nicht auf die Reihenfolge an, in der sie auf die Kästchen aufgeteilt werden, d.h. schwarze Kugel 1 in Kästchen 1 und schwarze Kugel 2 in Kästchen 2 führt zu demselben Ergebnis wie schwarze Kugel 1 in Kästchen 2 und schwarze Kugel 2 in Kästchen 1. Die auf diese Weise doppelt gezählten Möglichkeiten müssen also wieder herausdividiert werden: Vertauscht man also die schwarzen Kugeln untereinander, so gibt es für die erste schwarze Kugel k mögliche Positionen, für die zweite $k-1$ usw. und schließlich für die k–te Kugel nur noch eine mögliche Position. Damit ergibt sich obiger Quotient.

Ein gewähltes Muster wird festgehalten und es wird wieder nach der Anzahl der Möglichkeiten, die einzelnen Kästchen zu belegen, gefragt.

1.	schwarzes Kästchen:	R	Möglichkeiten
2.	schwarzes Kästchen:	$R-1$	Möglichkeiten
$\vdots$			
k.	schwarzes Kästchen:	$R-k+1$	Möglichkeiten
1.	weißes Kästchen:	$N-R$	Möglichkeiten
2.	weißes Kästchen:	$N-R-1$	Möglichkeiten
$\vdots$			
$n-k$.	weißes Kästchen:	$N-R-(n-k)+1$	Möglichkeiten

Zusammenzählen der Möglichkeiten ergibt

$$|A_k| = \binom{n}{k}(R)_k(N-R)_{n-k}.$$

Damit wird

$$\begin{aligned} p_k = P(A_k) = \frac{|A_k|}{|\Omega|} &= \frac{\binom{n}{k}(R)_k(N-R)_{n-k}}{(N)_n} \\ &= \frac{\frac{n!}{k!(n-k)!}(R)_k(N-R)_{n-k}}{(N)_n} \\ &= \frac{\frac{(R)_k}{k!}\,\frac{(N-R)_{n-k}}{(n-k)!}}{\frac{(N)_n}{n!}} \\ &= \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}} \qquad (k=0,1,\ldots,n). \end{aligned}$$

Offensichtlich ist $p_k \geq 0$ für $k = 0,1,\ldots,n$. Um nachzuweisen, dass die Folge $(p_k)_{k=0}^n$ einen Wahrscheinlichkeitsvektor über $\Omega' = \{0,1,\ldots,n\}$ darstellt, muss die Identität

$$\sum_{k=0}^{n}\binom{R}{k}\binom{N-R}{n-k} = \binom{N}{n}$$

bewiesen werden. Der Beweis folgt sofort aus der Tatsache, dass Ω die disjunkte Vereinigung der A_k mit $k = 0,\ldots,n$ ist. Alternativ lässt sich die Gleichung auch durch Nachrechnen verifizieren. Zunächst wird folgende Gleichung betrachtet:

$$(1+x)^N = (1+x)^R(1+x)^{N-R}.$$

Mit dem Binomischen Lehrsatz folgt weiter:

$$\begin{aligned}
(1+x)^N = \sum_{n=0}^{N} \binom{N}{n} x^n &= (1+x)^R(1+x)^{N-R} \\
&= \sum_{k=0}^{R} \binom{R}{k} x^k \sum_{j=0}^{N-R} \binom{N-R}{j} x^j \\
&= \sum_{k=0}^{R} \sum_{j=0}^{N-R} \binom{R}{k} \binom{N-R}{j} x^{k+j} \\
&= \sum_{n=0}^{N} \left(\sum_{k=0}^{n} \binom{R}{k} \binom{N-R}{n-k} \right) x^n.
\end{aligned}$$

Die Behauptung ergibt sich nun durch Koeffizientenvergleich.

2.5 Definition (hypergeometrische Verteilung):
Es seien $N, R, n \in \mathbb{N}$ mit $N \geq R$ und $n \leq N$. Das durch den Wahrscheinlichkeitsvektor

$$p_k := \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}} \qquad (k = 0, 1, \ldots, n)$$

auf $(\Omega', \mathfrak{P}(\Omega'))$ definierte Wahrscheinlichkeitsmaß heißt hypergeometrische Verteilung mit den Parametern N, R und n; kurz $Hg(n, R, N)$.

2.6 Beispiel (Lotto „6 aus 49“):
Mit $N := 49$, $R := 6$, $n := 6$ und $k \in \{0, 1, \ldots, 6\}$ gilt:

$$\begin{aligned}
p_0 &= \frac{\binom{6}{0}\binom{43}{6}}{\binom{49}{6}} \approx 0.4359, \\
p_1 &= \frac{\binom{6}{1}\binom{43}{5}}{\binom{49}{6}} \approx 0.4130, \\
p_2 &= \frac{\binom{6}{2}\binom{43}{4}}{\binom{49}{6}} \approx 0.1323, \\
p_3 &= \frac{\binom{6}{3}\binom{43}{3}}{\binom{49}{6}} \approx 0.0176, \\
p_4 &= \frac{\binom{6}{4}\binom{43}{2}}{\binom{49}{6}} \approx 0.9686 \cdot 10^{-3}, \\
p_5 &= \frac{\binom{6}{5}\binom{43}{1}}{\binom{49}{6}} \approx 0.1845 \cdot 10^{-4}, \\
p_6 &= \frac{\binom{6}{6}\binom{43}{0}}{\binom{49}{6}} \approx 0.7150 \cdot 10^{-7}.
\end{aligned}$$

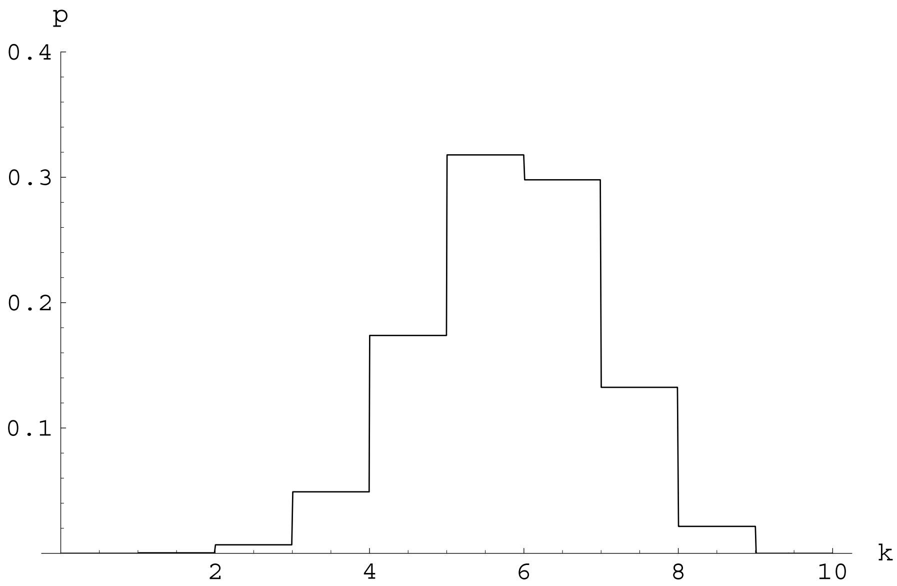


Abbildung 2.2: Hypergeometrische Verteilung mit $N := 30$, $R := 8$ und $n := 20$

Ziehen mit Zurücklegen

Beim Ziehen mit Zurücklegen wird jede gezogene Kugel sofort wieder in die Urne zurückgelegt. Nach erneutem Durchmischen wird die nächste Kugel gezogen. In diesem Fall kann jedes der n Kästchen mit jeder der N Kugeln belegt werden. Folglich ist

$$\Omega := \{1, \ldots, N\}^n = \{(\omega_1, \ldots, \omega_n) \mid \omega_j \in \{1, \ldots, N\}, j = 1, \ldots, n\}.$$

Damit ist $|\Omega| = N^n$.
A_k besteht aus allen $(\omega_1, \ldots, \omega_n) \in \Omega$, für die genau k Komponenten kleiner oder gleich R sind. Da die Kugeln zurückgelegt werden, kann jedes der k schwarzen Kästchen mit jeder der R schwarzen und jedes der $n-k$ weißen Kästchen mit jeder der $N-R$ weißen Kugeln belegt werden. Folglich ist

$$|A_k| = \binom{n}{k} R^k (N-R)^{n-k} \quad (k = 0, 1, \ldots, n)$$

und

$$P(A_k) = \binom{n}{k} \left(\frac{R}{N}\right)^k \left(\frac{N-R}{N}\right)^{n-k} \quad (k = 0, 1, \ldots, n).$$

Mit den Vereinbarungen $p := R/N$ und $q := 1-p = (N-R)/N$ ergibt sich

$$p_k = P(A_k) = \binom{n}{k} p^k q^{n-k} \quad (k = 0, 1, \ldots, n).$$

Damit definiert $(p_k)_{k=0}^n$ einen Wahrscheinlichkeitsvektor auf $\Omega' = \{0, \ldots, n\}$, denn es gilt $p_k \geq 0$ für $k = 0, 1, \ldots, n$ und

$$\sum_{k=0}^n p_k = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1.$$

2.7 Definition (Binomialverteilung):
Es seien $0 \le p \le 1$, $q := 1-p$ und $n \in \mathbb{N}$, dann heißt die durch den Wahrscheinlichkeitsvektor

$$p_k := \binom{n}{k} p^k q^{n-k} \qquad (k = 0, 1, \ldots, n)$$

auf der Menge $\Omega' := \{0, \ldots, n\}$ definierte Wahrscheinlichkeitsverteilung Binomialverteilung mit den Parametern n und p, kurz $\mathcal{B}(n, p)$.

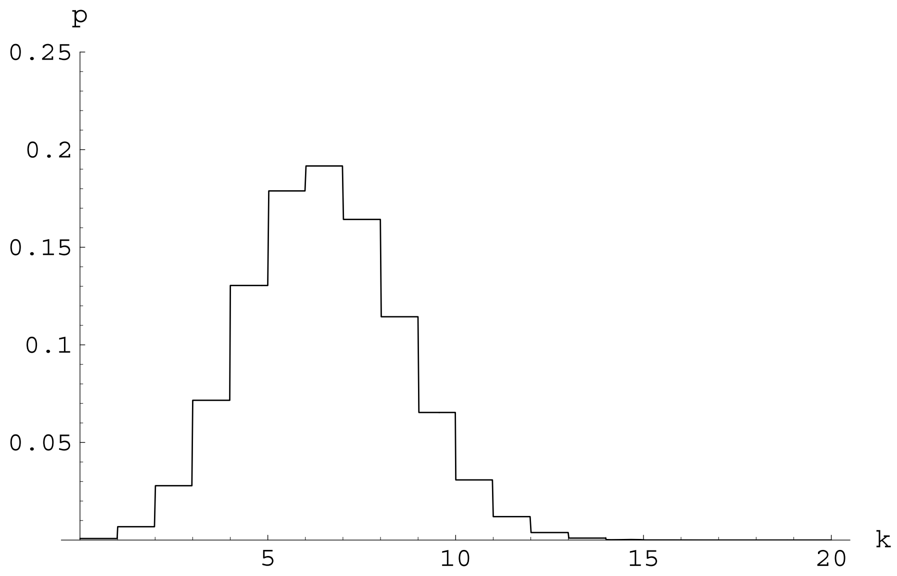


Abbildung 2.3: Binomialverteilung mit $p := 0.3$ und $n := 20$

2.8 Beispiel (Roulette–Spiel):
Beim Roulette–Spiel sei p_k die Wahrscheinlichkeit, in einer Spielserie vom Umfang n genau k–mal „Zero" zu haben. Die Berechnung von p_k kann auf die Binomialverteilung zurückgeführt werden, indem $N := 37$, $R := 1$, $p := R/N = 1/37$ und $q := 36/37$ gesetzt wird. Damit gilt

$$p_k = \binom{n}{k} \left(\frac{1}{37}\right)^k \left(\frac{36}{37}\right)^{n-k} \qquad (k = 0, 1, \ldots, n).$$

Allgemein lässt sich der Parameter p der Binomialverteilung als Wahrscheinlichkeit für das Eintreten eines bestimmten Ereignisses A interpretieren. p_k ist dann die Wahrscheinlichkeit für das k–malige Eintreten bei n gleichwertigen (unabhängigen) Wiederholungen des Zufallsexperimentes.

2.4 Beziehungen zwischen den Verteilungen

Wenn sich in der Urne wesentlich mehr Kugeln befinden, als gezogen werden, sollte es keine Rolle spielen, ob die gezogenen Kugeln wieder zurückgelegt werden oder nicht. Für praktische Rechnungen kann man sich daher auf das einfachere Modell der Binomialverteilung zurückziehen.

Formaler ausgedrückt besteht zwischen der hypergeometrischen und der Binomialverteilung folgende Beziehung:

2.9 Satz:
Es seien $0 < p \leq 1$ und $n \in \mathbb{N}$ fest gewählt. Falls $N, R \to \infty$, so dass $R/N \to p$, dann gilt:

$$Hg(n, R, N)(k) \longrightarrow \mathcal{B}(n, p)\,(k), \qquad (k = 0, 1, \ldots, n).$$

Bemerkung:
In der Stochastik II wird diese Art der Konvergenz auch als ,,schwache Konvergenz" oder ,,Konvergenz in Verteilung" bezeichnet (vgl. Kapitel 14 ,,Schwache Konvergenz und zentraler Grenzwertsatz").

Beweis:
Für $p < 1$ gelten die Implikationen

$$\frac{R}{N} \xrightarrow[N\to\infty]{} p \Longrightarrow \frac{N-R}{N} \xrightarrow[N\to\infty]{} 1-p \Longrightarrow N-R \xrightarrow[N\to\infty]{falls\ \ p<1} \infty,$$

wohingegen $\left(\frac{N-R}{N}\right) \to 0$ für $p = 1$ gilt. Somit folgt zunächst für $p < 1$:

$$\begin{aligned}
\mathrm{Hg}(n, R, N)(k) &= \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}} = \frac{(N-R)!}{(n-k)!(N-R-n+k)!}\frac{R!}{(R-k)!k!}\frac{(N-n)!n!}{N!} \\
&= \binom{n}{k}\frac{(R)_k(N-R)_{n-k}}{(N)_n} \\
&= \binom{n}{k}\left(\frac{R}{N}\right)^k\left(\frac{N-R}{N}\right)^{n-k}\frac{\frac{(R)_k}{R^k}\frac{(N-R)_{n-k}}{(N-R)^{n-k}}}{\frac{(N)_n}{N^n}} \\
&\xrightarrow[N\to\infty]{} \binom{n}{k}p^k(1-p)^{n-k} = B(n,p)(k).
\end{aligned}$$

Falls $p = 1$ ist, so gilt wegen $(N_j - R_j)_{n-k} \leq (N_j - R_j)^{n-k}$ und $\frac{(N_j)_n}{N_j^n}, \frac{(R_j)_k}{R_j^k} \xrightarrow[j\to\infty]{} 1$, dass

$$\mathrm{Hg}(n, R_r, N_j)(k) = \binom{n}{k}\underbrace{\left(\frac{R}{N}\right)^k}_{\to 1^k}\underbrace{\left(\frac{N-R}{N}\right)^{n-k}}_{\to 0}\underbrace{\frac{\frac{(R)_k}{R^k}\frac{(N-R)_{n-k}}{(N-R)^{n-k}}}{\frac{(N)_n}{N^n}}}_{beschr.} \xrightarrow[N\to\infty]{} 0 = \mathrm{B}(n,1)(k)$$

für $k < n$ und

$$\mathrm{Hg}(n, R_r, N_j)(n) = \frac{\binom{R}{n}\binom{N-R}{0}}{\binom{N}{n}} = \frac{\binom{R}{n}}{\binom{N}{n}} \xrightarrow[j\to\infty]{} 1 = \mathrm{B}(n,1)(n).$$

■

2.10 Beispiel:
In einer Stadt mit 2 Millionen Einwohnern stimmen 800.000 (40%) für eine bestimmte Partei. 100 Personen werden zufällig ausgewählt. Die Verteilung der Anzahl der Einwohner unter jenen 100, die für diese Partei stimmen, ist:

$$\mathrm{Hg}(100, 800\,000, 2\,000\,000) \approx \mathcal{B}(100, 0.4)\,.$$

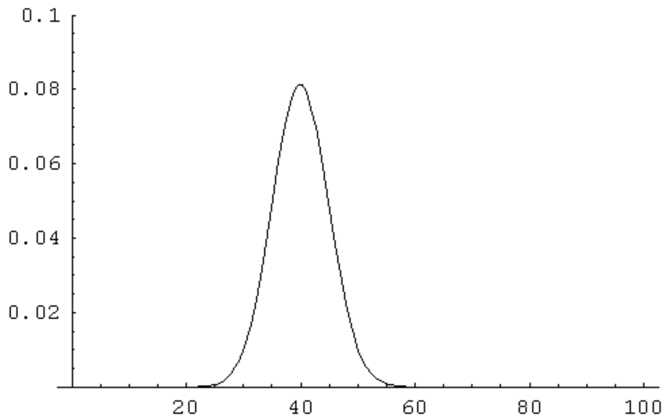


Abbildung 2.4: Binomialverteilung mit $n := 100$ und $p := 0.4$ sowie hypergeometrische Verteilung mit $N := 2\,000\,000$, $R := 800\,000$ und $n := 100$; es ist in der Grafik kein Unterschied zwischen den beiden Verteilungen erkennbar.

Es ist

$$\max_{k \in \{0,100\}} |\mathrm{Hg}(100, 800\,000, 2\,000\,000)(k) - \mathcal{B}(100, 0.4)\,(k)| < 3 \cdot 10^{-6}.$$

2.11 Satz:
Es seien $k \in \mathbb{N}_0$, $k < n$ und $\lambda > 0$ fest gewählt. Setze $p_n = \lambda/n$, so folgt:

$$\lim_{n\to\infty} \mathcal{B}(n, p_n)\,(k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Beweis:
Allgemein lässt sich folgendes Teleskop-Produkt schreiben:

$$\mathcal{B}(n,p_n)\,(k) = \frac{\mathcal{B}(n,p_n)\,(k)}{\mathcal{B}(n,p_n)\,(k-1)} \cdot \frac{\mathcal{B}(n,p_n)\,(k-1)}{\mathcal{B}(n,p_n)\,(k-2)} \cdot \ldots \cdot \frac{\mathcal{B}(n,p_n)\,(1)}{\mathcal{B}(n,p_n)\,(0)} \cdot \mathcal{B}(n,p_n)\,(0).$$

Für $k = 0$ gilt

$$\begin{aligned}
\lim_{n\to\infty} \mathcal{B}(n,p_n)\,(0) &= \lim_{n\to\infty} \binom{n}{0} p_n^0 (1-p_n)^n \;=\; \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n \\
&= \lim_{n\to\infty} \exp\left(n \,\ln\left(1 - \frac{\lambda}{n}\right)\right) \;=\; \exp\left(\lim_{n\to\infty} n\,\ln\left(1-\frac{\lambda}{n}\right)\right) \\
&= \exp\left(-\lambda \cdot \lim_{n\to\infty} -\frac{n}{\lambda}\,\ln\left(1-\frac{\lambda}{n}\right)\right) \;=\; \exp(-\lambda \cdot\, \ln'(1)) \;=\; e^{-\lambda}.
\end{aligned}$$

Ferner gilt:

$$\begin{aligned}
\frac{\mathcal{B}(n,p_n)\,(k)}{\mathcal{B}(n,p_n)\,(k-1)} &= \frac{\binom{n}{k} p_n^k (1-p_n)^{n-k}}{\binom{n}{k-1} p_n^{k-1} (1-p_n)^{n-k+1}} = \frac{\frac{n!}{k!\cdot(n-k)!} \cdot p_n}{\frac{n!}{(k-1)!\cdot(n-k+1)!} \cdot (1-p_n)} \\
&= \frac{n-k+1}{k} \cdot \frac{\frac{\lambda}{n}}{1-\frac{\lambda}{n}} = \frac{\lambda}{k} \cdot \frac{n-k+1}{n} \cdot \frac{1}{1-\frac{\lambda}{n}} \\
&\xrightarrow[n\to\infty]{} \frac{\lambda}{k}.
\end{aligned}$$

Folglich gilt

$$\mathcal{B}(n,p)\,(k) \xrightarrow[n\to\infty]{} \frac{\lambda^k}{k!}e^{-\lambda}.$$

■

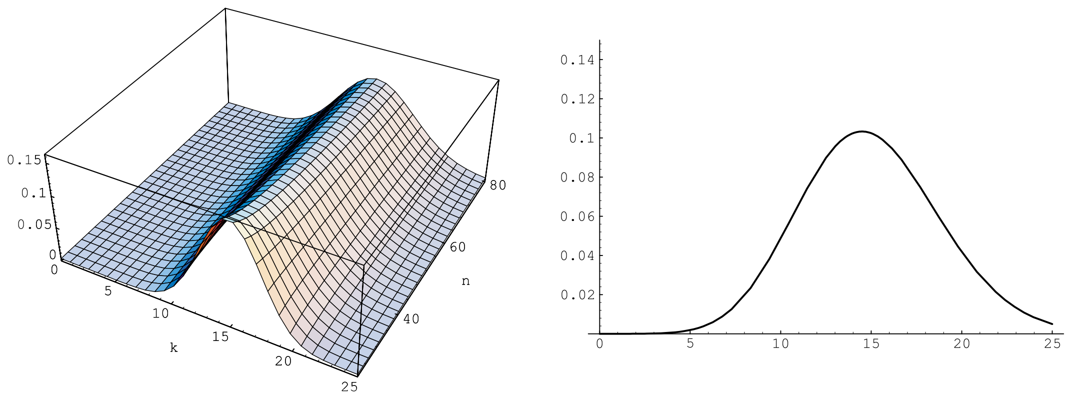


Abbildung 2.5: Die Grafik verdeutlicht die Annäherung der Binomialverteilung an die Poissonverteilung im Falle $\lambda := 15$. Die Grenzfunktion ist rechts dargestellt.

2.12 Bemerkung:
Die Größen

$$p_k := \frac{\lambda^k}{k!}e^{-\lambda} \quad (k = 0, 1, \ldots)$$

bilden einen Wahrscheinlichkeitsvektor über $\Omega := \mathbb{N}_0$, denn es ist $p_k \geq 0$ für alle k und es gilt

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

2.13 Definition (Poissonverteilung):
Es sei $\lambda \in \mathbb{R}^+$. Das durch den Wahrscheinlichkeitsvektor

$$p_k := \frac{\lambda^k}{k!}e^{-\lambda} \qquad (k = 0, 1, 2, \ldots)$$

definierte Wahrscheinlichkeitsmaß über $(\mathbb{N}_0, \mathfrak{P}(\mathbb{N}_0))$ heißt Poissonverteilung mit dem Parameter λ, kurz $P(\lambda)$.

2.14 Beispiel:
Es sei bekannt, dass pro Jahr 0.005% einer Bevölkerungsgruppe durch einen gewissen Unfall verletzt wird. Bei einer Versicherung sind 10.000 Personen gegen diesen Unfall versichert. Gesucht ist die Wahrscheinlichkeit, dass in einem Jahr mindestens drei Versicherungsnehmer verunglücken. Diesem Zufallsexperiment liegt offensichtlich die Binomialverteilung mit

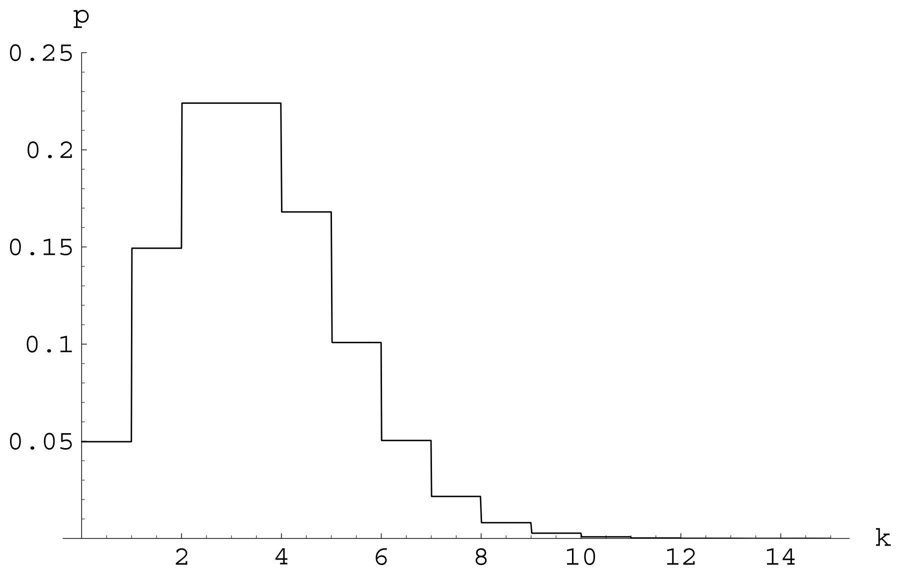


Abbildung 2.6: Poissonverteilung mit $\lambda := 3$

den Parametern $n := 10000$ und $p := 0.00005$ zugrunde. Für die Approximation durch die Poissonverteilung wird $\lambda := n \cdot p = 0.5$ gesetzt. Es ergibt sich:

$$\begin{aligned} P(\text{es verunglücken mindestens 3}) &= 1 - P(\text{es verunglücken maximal 2}) \\ &= 1 - \sum_{k=0}^{2} p_k = 1 - \sum_{k=0}^{2} \frac{\lambda^k}{k!} e^{-\lambda} = 1 - e^{-0.5}\left(1 + \frac{1}{2} + \frac{1}{8}\right) \\ &\approx 0.0144 = 1.44\%. \end{aligned}$$

2.5 Diskrete Zufallsvariable

Im Zusammenhang mit dem Würfelspiel wurde gefragt, wie groß die Wahrscheinlichkeit ist, eine der Zahlen 1 bis 6 bzw. eine gerade oder ungerade Augenzahl zu würfeln. In beiden Fällen wurde als Ereignisalgebra die Potenzmenge $\mathfrak{P}(\Omega)$ von $\Omega := \{1, \ldots, 6\}$ zugrundegelegt, die bereits $2^6 = 64$ verschiedene Elemente enthält. Hätte man sich von Anfang an auf die Ereignisse $A := \{1, 3, 5\}$ (ungerade Augenzahl) und $\overline{A} := \{2, 4, 6\}$ (gerade Augenzahl) beschränkt, so hätte als Ereignisalgebra ebensogut das Mengensystem $\mathfrak{F} := \{\varnothing, A, \overline{A}, \Omega\}$ verwendet werden können, in der die Ereignisse A und $\overline{A}$ sogar als Elementarereignisse auftreten. Es kann sich also durchaus als sinnvoll erweisen, einen gegebenen Wahrscheinlichkeitsraum den praktischen Erfordernissen entsprechend einzuschränken bzw. zu modifizieren. Der Übergang von einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{F}, P)$ zu einem anderen $(\Omega', \mathfrak{F}', P')$ wird im Allgemeinen mit Hilfe einer Abbildung $X\colon \Omega \to \Omega'$ vollzogen. In diesem Zusammenhang stellt sich die Frage, wie man die über $(\Omega, \mathfrak{F})$ gegebene Bewertung P zu einer Bewertung P' auf $(\Omega', \mathfrak{F}')$ ausdehnen kann. Da man das Ereignis $A' \in \mathfrak{F}'$ immer dann beobachtet, wenn im ursprünglichen Experiment ein ω mit $X(\omega) \in A'$ eintritt, definiert man

$$P'(A') := P_X(A') := P(X^{-1}(A')) = P(\{\omega \in \Omega \mid X(\omega) \in A'\}) \quad \forall\, A' \in \mathfrak{F}'.$$

Wenn man in dieser Weise vorgeht, muss sichergestellt werden, dass für alle $A' \in \mathfrak{F}'$ die Bedingung
$$X^{-1}(A') = \{\omega \in \Omega \mid X(\omega) \in A'\} \in \mathfrak{F}$$
erfüllt ist, denn nur für Elemente $A \in \mathfrak{F}$ ist $P(A)$ definiert. Eine Abbildung $X\colon \Omega \to \Omega'$ mit dieser Eigenschaft heißt Zufallsvariable. Ersetzt man Ω und Ω' durch diskrete Mengen und $\mathfrak{F}$ und $\mathfrak{F}'$ durch die zugehörigen Potenzmengen $\mathfrak{P}(\Omega)$ und $\mathfrak{P}(\Omega')$, dann ist diese Bedingung aufgrund der Beziehung

$$X^{-1}(A') = X^{-1}(A' \cap \Omega') = \bigcup_{\omega' \in A' \cap \Omega'} X^{-1}(\{\omega'\})$$

stets erfüllt (siehe Abbildung 2.7). Außerdem ist sofort ersichtlich, dass

$$P_X(\{\omega_i'\}) := P(X^{-1}(\{\omega_i'\}))$$

einen Wahrscheinlichkeitsvektor auf Ω' darstellt und P_X deshalb für alle $A' \in P(\Omega')$ eindeutig definiert ist (siehe Satz 2.2).

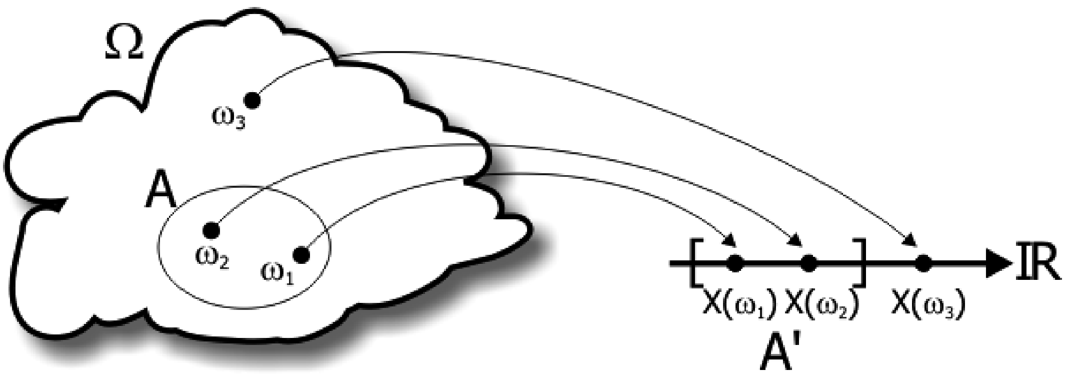


Abbildung 2.7: Zufallsvariable

2.15 Definition (diskrete Zufallsvariable, Bildmaß, Verteilung):
Es seien Ω und Ω' diskrete Stichprobenmengen und die Potenzmengen $\mathfrak{P}(\Omega)$ und $\mathfrak{P}(\Omega')$ seien die zugehörigen σ–Algebren. Jede Abbildung $X\colon \Omega \to \Omega'$ heißt diskrete Zufallsvariable. Ist $\Omega' \subseteq \mathbb{R}$ spricht man von einer reellwertigen diskreten Zufallsvariable. Das durch X induzierte Wahrscheinlichkeitsmaß P_X auf $\mathfrak{F}' := \mathfrak{P}(\Omega')$ heißt Bildmaß von P unter X bzw. die Verteilung von X unter P.

2.16 Beispiel (Spiel mit zwei Würfeln I):
Es soll die Abbildung $X\colon (\omega_1, \omega_2) \to \min(\omega_1, \omega_2)$ betrachtet werden.

$\min(\omega_1, \omega_2)$:

$\omega_1 \backslash \omega_2$	1	2	3	4	5	6
1	1	1	1	1	1	1
2	1	2	2	2	2	2
3	1	2	3	3	3	3
4	1	2	3	4	4	4
5	1	2	3	4	5	5
6	1	2	3	4	5	6

Somit ist

$$P_X(\{1\}) = P(X^{-1}(\{1\})) = P(\{(1,1),(1,2),\ldots,(1,6),(2,1),(3,1),\ldots,(6,1)\}) = \frac{11}{36}.$$

Analog gilt

$$\begin{aligned} P_X(\{1\}) &= \frac{11}{36}, & P_X(\{4\}) &= \frac{5}{36}, \\ P_X(\{2\}) &= \frac{9}{36}, & P_X(\{5\}) &= \frac{3}{36}, \\ P_X(\{3\}) &= \frac{7}{36}, & P_X(\{6\}) &= \frac{1}{36}. \end{aligned}$$

(Siehe auch PowerPoint-Präsentation zu diesem Beispiel.)

2.17 Beispiel (Spiel mit zwei Würfeln II):
Es soll die Abbildung $X\colon \Omega \times \Omega \to \Omega' = \{2,\ldots,12\}$ mit $X\colon (\omega_1,\omega_2) \to \omega_1 + \omega_2$ betrachtet werden.

$\omega_1 + \omega_2$:

$\omega_1 \backslash \omega_2$	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

d.h.

$$\begin{aligned} P_X(\{2\}) &= \frac{1}{36}, & P_X(\{6\}) &= \frac{5}{36}, & P_X(\{10\}) &= \frac{3}{36}, \\ P_X(\{3\}) &= \frac{2}{36}, & P_X(\{7\}) &= \frac{6}{36}, & P_X(\{11\}) &= \frac{2}{36}, \\ P_X(\{4\}) &= \frac{3}{36}, & P_X(\{8\}) &= \frac{5}{36}, & P_X(\{12\}) &= \frac{1}{36}. \\ P_X(\{5\}) &= \frac{4}{36}, & P_X(\{9\}) &= \frac{4}{36}. \end{aligned}$$

(Siehe auch PowerPoint-Präsentation zu diesem Beispiel.)

2.6 Kenngrößen einer diskreten Zufallsvariablen

Ein Hochschullehrer der Stochastik möchte herausfinden, ob sich das Leistungsniveau seiner Studentinnen und Studenten gegenüber früheren Jahren verändert hat. Aus diesem Grunde vergleicht er die Klausurergebnisse des laufenden mit denen eines früheren Jahrgangs. Als Klausurergebnis mögen die Punktzahlen x_i, $i = 1,\ldots,r$, mit den absoluten Häufigkeiten $H_n(x_i)$ auftreten. Dann wird man

$$\begin{aligned} \overline{x} &:= \frac{1}{n}(H_n(x_1)\cdot x_1 + \ldots + H_n(x_r)\cdot x_r) \\ &= h_n(x_1)\cdot x_1 + \ldots + h_n(x_r)\cdot x_r, \end{aligned}$$

als die mittlere oder erwartete Punktzahl pro Teilnehmer bezeichnen (wobei $h_n(x_i)$ die relative Häufigkeit des Elementarereignisses x_i bezeichnet) und denjenigen Jahrgang als den besseren ansehen, für den dieser Wert der größere ist. Wenn sich die relativen Häufigkeiten für $n \to \infty$ gegen die Grenzwerte $p_i := P(\{x_i\})$ stabilisieren kann die Fragestellung mit Mitteln der Stochastik untersucht werden. Allgemein kann definiert werden:

2.18 Definition (Erwartungswert):
Es sei X eine reellwertige Zufallsvariable über dem diskreten Wahrscheinlichkeitsraum (Ω, P). Ist die Reihe

$$\mathrm{E}[X] := \sum_{x \in X(\Omega)} x \cdot P(X = x) = \sum_{x \in X(\Omega)} x \cdot P_X(\{x\})$$

absolut konvergent, so heißt ihr Wert Erwartungswert von X.

2.19 Bemerkung:
Die absolute Konvergenz der Reihe wird gefordert, damit $\mathrm{E}[X]$ von der Reihenfolge der Summation unabhängig ist.
Bei der Definition des Erwartungswertes wurde das Bildmaß P_X zugrundegelegt. Aufgrund der Beziehung

$$\begin{aligned} E[X] &:= \sum_{x \in X(\Omega)} x \cdot P_X(X = x) = \sum_{x \in X(\Omega)} x \sum_{\omega \mid X(\omega) = x} p_\omega \\ &= \sum_{x \in X(\Omega)} \sum_{\omega \mid X(\omega) = x} X(\omega) p_\omega = \sum_{\omega \in \Omega} X(\omega) \cdot p_\omega \end{aligned}$$

kann jedoch ebensogut das originäre Maß verwendet werden.

2.20 Beispiel (Erwartungswert beim Spiel mit einem Würfel):
Es seien $x_i = i$ und $p_i = 1/6$ für $i = 1, \ldots, 6$. Damit gilt also:

$$\mathrm{E}[X] = \sum_{i=1}^{6} i \cdot p_i = \frac{1}{6} \sum_{i=1}^{6} i = \frac{21}{6} = 3.5\,.$$

2.21 Beispiel (Erwartungswert beim Spiel mit zwei Würfeln):
Es ergibt sich $\Omega := \{(k, \ell) \mid k, \ell = 1, \ldots, 6\}$, $\mathfrak{F} = \mathfrak{P}(\Omega)$ und $\Omega' :=$ 'Summe der Augenzahlen' $= \{2, \ldots, 12\}$. Somit wird die Zufallsgröße $X \colon \Omega \to \Omega'$ mit $X(k, \ell) := k + \ell$ betrachtet. Der Erwartungswert von X berechnet sich unter Zugrundelegung einer Gleichverteilung wie folgt:

$$\begin{aligned} \mathrm{E}[X] &= \sum_{j=2}^{12} j \cdot P(\{\omega = (k, \ell) \in \Omega \mid k + \ell = j\}) \\ &= 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} + \\ &\quad + 8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} \\ &= 7. \end{aligned}$$

2.22 Bemerkung:
Ist $X\colon \Omega \to \Omega'$ eine reellwertige diskrete Zufallsvariable und $g\colon \Omega' \to \mathbb{R}$ eine reellwertige Funktion auf Ω', so ist $Y := g \circ X$ eine reellwertige Zufallsvariable. Mit $\mathbf{E}_P\,[Y]$ bezeichnen wir den Erwartungswert von Y bzgl. P und mit

$$\mathbf{E}_{P_X}\,[g] := \sum_{\omega' \in \Omega'} g(\omega') \cdot P_X(\{\omega'\})$$

den Erwartungswert von g bzgl. P_X.

2.23 Satz:
Der Erwartungswert $\mathbf{E}_P\,[Y]$ *existiert genau dann, wenn* $\mathbf{E}_{P_X}\,[g]$ *existiert und es ist* $\mathbf{E}_P\,[Y] = \mathbf{E}_{P_X}\,[g]$.

Beweis:
Existiert einer der beiden Erwartungswerte, so lässt sich mit Hilfe des Umordnungssatzes der jeweils andere berechnen:

$$\begin{aligned} E_P[Y] &= \sum_{\omega \in \Omega} g(X(\omega)) \cdot p_\omega = \sum_{\omega' \in \Omega'} g(\omega^{'}) \cdot \sum_{\omega \in X^{-1}(\{\omega\}')} p_\omega \\ &= \sum_{\omega' \in \Omega'} g(\omega^{'}) \cdot P(X^{-1}(\{\omega^{'}\}) = \sum_{\omega' \in \Omega'} g(\omega^{'}) \cdot P_X(\{\omega^{'}\}) = E_{P_X}[g]. \end{aligned}$$

2.24 Definition (k–te Moment, k–te zentrale Moment, Varianz):
Es sei X eine reellwertige diskrete Zufallsvariable mit den Werten x_1, x_2, … und den Wahrscheinlichkeiten $p_1 := P(X = x_1)$, $p_2 := P(X = x_2)$, … Gilt $\sum_{i=1}^{\infty} \left|x_i^k\right| p_i < \infty$, dann nennt man

$$\mathrm{E}\Big[X^k\Big] := \sum_{i=1}^{\infty} x_i^k \cdot p_i$$

das <u>k–te Moment</u> von X. Gilt $\sum_{i=1}^{\infty} \left|(x_i - \mathrm{E}[X])^k\right| p_i < \infty$, dann nennt man

$$\mathrm{E}\Big[(X - \mathrm{E}[X])^k\Big] := \sum_{i=1}^{\infty} (x_i - \mathrm{E}[X])^k p_i$$

das <u>k–te zentrale Moment</u> von X. Das zweite zentrale Moment wird auch die <u>Varianz</u> von X genannt, kurz $\mathrm{Var}[X]$.

2.25 Satz:
Es seien X und Y reellwertige diskrete Zufallsvariablen über demselben Wahrscheinlichkeitsraum (Ω, p), für die $\mathrm{E}[X]$ und $\mathrm{E}[Y]$ existieren. Dann gilt für $a, b \in \mathbb{R}$:

$$\mathrm{E}[aX + b] = a \cdot \mathrm{E}[X] + b$$

und

$$\mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y]\,.$$

Für $\mathrm{E}\big[X^2\big] < \infty$ gilt außerdem

$$\mathrm{Var}[X] = \mathrm{E}\big[X^2\big] - (\mathrm{E}[X])^2 \qquad \textit{(Verschiebungssatz)}$$

und

$$\mathrm{Var}[aX + b] = a^2 \cdot \mathrm{Var}[X]\,.$$

Beweis:
Mit $p_i := P(X = x_i)$ gilt

$$\begin{aligned} \mathrm{E}[aX+b] &= \sum_{i=1}^{\infty}(ax_i+b)p_i = a\sum_{i=1}^{\infty}x_ip_i + b\sum_{i=1}^{\infty}p_i \\ &= a\sum_{i=1}^{\infty}x_ip_i + b = a\cdot\mathrm{E}[X]+b \end{aligned}$$

bzw.

$$\mathrm{E}[aX+b] = \sum_{\omega\in\Omega}[aX(\omega)+b]p_\omega = a\sum_{\omega\in\Omega}X(\omega)p_\omega + b\sum_{\omega\in\Omega}p_\omega = a\cdot\mathrm{E}[X]+b.$$

Entsprechend gilt

$$\mathrm{E}[X+Y] = \sum_{\omega\in\Omega}[X(\omega)+Y(\omega)]p_\omega = \sum_{\omega\in\Omega}X(\omega)p_\omega + \sum_{\omega\in\Omega}Y(\omega)p_\omega = \mathrm{E}[X]+\mathrm{E}[Y]\,.$$

Weiter ist

$$\begin{aligned} \mathrm{Var}[X] &= \mathrm{E}\big[(X-\mathrm{E}[X])^2\big] = \mathrm{E}\big[X^2 - 2X\mathrm{E}[X] + (\mathrm{E}[X])^2\big] \\ &= \mathrm{E}\big[X^2\big] - 2\mathrm{E}[X]\,\mathrm{E}[X] + (\mathrm{E}[X])^2 = \mathrm{E}\big[X^2\big] - (\mathrm{E}[X])^2 \end{aligned}$$

und

$$\mathrm{Var}[aX+b] = \mathrm{E}\big[(aX+b-a\mathrm{E}[X]-b)^2\big] = \mathrm{E}\big[a^2(X-\mathrm{E}[X])^2\big] = a^2\,\mathrm{Var}[X]\,.$$

■

2.26 Beispiel:
Die Zufallsvariable X sei binomialverteilt mit den Parametern n und p. Dann gilt:

$$\begin{aligned} \mathrm{E}[X] &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\ &= \sum_{k=0}^{n}\frac{kn!}{k!(n-k)!}p^k(1-p)^{n-k} \\ &= n\cdot p\cdot\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}(1-p)^{n-k} \\ &= n\cdot p\cdot\sum_{k=0}^{n-1}\frac{(n-1)!}{k!(n-k-1)!}p^k(1-p)^{n-k-1} \\ &= n\cdot p\cdot\sum_{k=0}^{n-1}\binom{n-1}{k}p^k(1-p)^{n-1-k} \\ &= n\cdot p. \\ \mathrm{E}\big[X^2\big] &= \mathrm{E}[X(X-1)] + \mathrm{E}[X] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} + np \\
&= \sum_{k=2}^{n} k(k-1)\frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} + np \\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2}(1-p)^{n-k} + np \\
&= n(n-1)p^2 \sum_{k=0}^{n-2} \binom{n-2}{k} p^k (1-p)^{n-2-k} + np \\
&= n^2p^2 - np^2 + np.
\end{aligned}
$$

Insgesamt folgt damit:

$$\mathrm{Var}[X] = \mathrm{E}\big[X^2\big] - (\mathrm{E}[X])^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p).$$

2.27 Beispiel:
Die Zufallsgröße X sei Poisson–verteilt mit dem Parameter λ. Dann gilt:

$$
\begin{aligned}
\mathrm{E}[X] &= \sum_{k=0}^{\infty} k \cdot p_k = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\
&= \lambda \cdot e^{\lambda} \cdot e^{-\lambda} = \lambda. \\
\mathrm{E}\big[X^2\big] &= \mathrm{E}[X(X-1)+X] = \mathrm{E}[X(X-1)] + \mathrm{E}[X] \\
&= \sum_{k=2}^{\infty} k \cdot (k-1) \cdot \frac{\lambda^k}{k!} e^{-\lambda} + \lambda = \lambda^2 \cdot \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} + \lambda \\
&= \lambda^2 \cdot e^{\lambda} \cdot e^{-\lambda} + \lambda = \lambda^2 + \lambda.
\end{aligned}
$$

Folglich ist:

$$\mathrm{Var}[X] = \mathrm{E}\big[X^2\big] - (\mathrm{E}[X])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

(Siehe auch PowerPoint-Präsentation zu diesem Beispiel.)

2.7 Erzeugende Funktion

Oftmals ist die direkte Berechnung von Erwartungswert und Varianz aufwendig. Die in diesem Abschnitt betrachtete erzeugende Funktion stellt einen alternativen Weg zur Berechnung der Kenngrößen dar.

2.28 Definition (erzeugende Funktion):
Für eine auf $\mathbb{N}_0$ verteilte Zufallsvariable X mit der Verteilung $p_k := P_X(\{k\}) = P(X = k)$, $k = 0, 1, 2, \ldots$, $z \in \mathbb{C}$ heißt die Potenzreihe

$$G(z) := \mathrm{E}\big[z^X\big] = \sum_{k=0}^{\infty} p_k z^k, \qquad |z| \leq 1,$$

erzeugende Funktion von X bzw. $(p_k)_{k=0}^{\infty}$.

2.29 Bemerkung:

- Wegen $\sum_{k=0}^{\infty} p_k = 1$ konvergiert $G(z)$ mindestens für alle $|z| \leq 1$, $z \in \mathbb{C}$.
- Der Wahrscheinlichkeitsvektor $p := (p_k)_{k=0}^{\infty}$ ist durch $G(z)$ eindeutig bestimmt (und umgekehrt). Insbesondere gilt:

$$p_k = P(X = k) = \frac{G^{(k)}(0)}{k!} \qquad (k = 0, 1, 2, \ldots),$$

  wobei $G^{(k)}$ die k–te Ableitung von G bedeutet.

2.30 Satz:
Es sei $G(z)$ die erzeugende Funktion der diskreten Zufallsvariable X mit Werten in $\mathbb{N}_0$. Dann gilt:

$$\mathrm{E}[X \cdot (X-1) \cdot \ldots \cdot (X-k+1)] = \lim_{z \to 1-} G^{(k)}(z) = \lim_{z \to 1-} \frac{d^k G(z)}{dz^k}, \quad k \in \mathbb{N},$$

genau dann, wenn $\mathrm{E}\left[X^k\right] < \infty$ *ist.*

Beweis:
Mit $\mathrm{E}\left[X^k\right]$ sind auch alle $\mathrm{E}[X^m]$ für $0 \leq m \leq k$ endlich, da $|X(\omega)|^m \leq |X(\omega)|^k$ gilt, falls $|X(\omega)| \geq 1$ ist. Für $|z| < 1$ gilt:

$$G^{(k)}(z) = \sum_{\ell=0}^{\infty} \underbrace{\ell \cdot (\ell-1) \cdot \ldots \cdot (\ell-k+1)}_{\leq \ell^k} \cdot P(X = \ell) \cdot z^{\ell-k}.$$

und $|\ell^k p_\ell z^{\ell-k}| \leq \ell^k p_\ell$. Wenn $\mathrm{E}\left[X^k\right] < \infty$ ist, so folgt mit $\mathrm{E}\left[X^k\right] = \sum_{\ell=0}^{\infty} \ell^k p_\ell$ die folgende Identität (und damit insbesondere die absolute Konvergenz der Reihe)

$$\mathrm{E}[X \cdot (X-1) \cdot \ldots \cdot (X-k+1)] = \lim_{z \to 1-} G^{(k)}(z) = \lim_{z \to 1-} \frac{d^k G(z)}{dz^k}.$$

Die Umkehrung kann durch Induktion über k gezeigt werden, wobei die Identität

$$X^{m+1} = X(X-1) \cdots (X-m) + \sum_{k=1}^{m} c_k X^k$$

für passende $c_k \in \mathbb{Z}$ benutzt wird. ■

2.31 Beispiel:
Die Zufallsvariable X sei binomialverteilt mit den Parametern n und p, d.h.

$$p_k := P(X = k) := \binom{n}{k} p^k (1-p)^{n-k} \qquad (k = 0, 1, \ldots, n).$$

Dann ist

$$\begin{aligned} G(z) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} z^k = \sum_{k=0}^{n} \binom{n}{k} (pz)^k (1-p)^{n-k} \\ &= (pz+q)^n, \ z \in \mathbb{C}, \\ \mathrm{E}[X] &= \lim_{z \to 1-} G'(z) = p\left(n \cdot (pz+q)^{n-1}\right)\Big|_{z=1} = n \cdot p. \end{aligned}$$

2.32 Beispiel:
Die Zufallsvariable X sei Poisson–verteilt mit dem Parameter $\lambda > 0$, d.h.

$$p_k := P(X = k) := \frac{\lambda^k}{k!} e^{-\lambda} \qquad (k = 0, 1, 2, \ldots).$$

Dann gilt

$$\begin{aligned} G(z) &= \sum_{k=0}^{\infty} p_k z^k = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} z^k = \sum_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} e^{\lambda z} = e^{-\lambda(1-z)},\ z \in \mathbb{C}, \\ \mathrm{E}[X] &= \lim_{z \to 1-} G'(z) = \left(e^{-\lambda(1-z)} \lambda \right) \Big|_{z=1} = \lambda. \end{aligned}$$

2.33 Beispiel:
Die Zufallsvariable X sei hypergeometrisch verteilt mit den Parametern n, R und N, d.h.

$$p_k := P(X = k) := \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}} \quad (k = 0, \ldots, R).$$

Dann gilt

$$\begin{aligned} G(z) &= \sum_{k=0}^{\infty} p_k z^k = \sum_{k=0}^{\infty} \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}} z^k, \\ \mathrm{E}[X] &= \sum_{k=0}^{R} k p_k = \sum_{k=0}^{R} \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}} k = \sum_{k=1}^{R} \frac{R!}{(k-1)!(R-k)!} \cdot \frac{\binom{N-R}{n-k}}{\binom{N}{n}} \\ &= R \sum_{k=1}^{R} \frac{(R-1)!}{(k-1)!((R-1)-(k-1))!} \cdot \frac{\binom{(N-1)-(R-1)}{(n-1)-(k-1)}}{\frac{N(N-1)!}{n(n-1)!((N-1)-(n-1))!}} \\ &= n\frac{R}{N} \sum_{k=0}^{R-1} \frac{(R-1)!}{k!((R-1)-k)!} \cdot \frac{\binom{(N-1)-(R-1)}{(n-1)-k}}{\binom{N-1}{n-1}} \\ &= n\frac{R}{N} \sum_{k=0}^{R-1} Hg(n-1, R-1, N-1)(k) = n\frac{R}{N}. \end{aligned}$$

2.34 Bemerkung:
Die Varianz einer Zufallsgröße berechnet sich mit Hilfe der erzeugenden Funktion wie folgt:

$$\begin{aligned} \mathrm{Var}[X] &= \mathrm{E}\big[X^2\big] - (\mathrm{E}[X])^2 = \mathrm{E}[X(X-1) + X] - (\mathrm{E}[X])^2 \\ &= G''(1) + G'(1) - (G'(1))^2, \end{aligned}$$

falls G in $z = 1$ zweimal stetig differenzierbar ist.

2.35 Beispiel:
Für die Binomialverteilung gilt

$$\begin{aligned} G(z) &= (pz+q)^n, & z \in \mathbb{C}. \\ G'(z) &= n \cdot (pz+q)^{n-1} \cdot p, & z \in \mathbb{C}. \\ G''(z) &= n \cdot (n-1) \cdot (pz+q)^{n-2} \cdot p^2, & z \in \mathbb{C}. \end{aligned}$$

Also

$$\begin{aligned}\mathrm{Var}[X] &= G''(1) + G'(1) - (G'(1))^2 = n \cdot (n-1) \cdot p^2 + n \cdot p - (n \cdot p)^2 \\ &= n^2 \cdot p^2 - n \cdot p^2 + n \cdot p - n^2 \cdot p^2 = n \cdot p \cdot (1-p).\end{aligned}$$

2.36 Beispiel:
Für die Poissonverteilung gilt

$$\begin{aligned} G(z) &= e^{-\lambda(1-z)}, & z \in \mathbb{C}. \\ G'(z) &= e^{-\lambda(1-z)} \cdot \lambda, & z \in \mathbb{C}. \\ G''(z) &= e^{-\lambda(1-z)} \cdot \lambda^2, & z \in \mathbb{C}. \end{aligned}$$

Also

$$\mathrm{Var}[X] = G''(1) + G'(1) - (G'(1))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Literatur zu Kapitel 2

Folgende Bücher werden als begleitende Literatur zu diesem Kapitel des Skriptes empfohlen:

- M. Fisz:
  Wahrscheinlichkeitsrechnung und mathematische Statistik,
  VEB, Deutscher Verlag der Wissenschaften 1989.
  ISBN: 3326000790

- P. Gänssler und W. Stute:
  Wahrscheinlichkeitstheorie,
  Springer–Verlag, Berlin, 1977.
  ISBN: 3540084185

- H. O. Georgii:
  Stochastik,
  2. Auflage, de Gruyter, 2004.
  ISBN: 3110172356

- K. Hinderer:
  Grundbegriffe der Wahrscheinlichkeitstheorie,
  Springer–Verlag, 1980.
  ISBN: 3540073094

- G. Hübner:
  Stochastik. Eine Einführung für Mathematiker, Informatiker und Ingenieure.,
  4. Auflage, Vieweg Verlag, 2003.
  ISBN: 3528254432

- U. Krengel:
  Einführung in die Wahrscheinlichkeitstheorie und Statistik,
  6. Auflage, Vieweg, 2002.
  ISBN: 3528672595

- J. Lehn/H. Wegmann:
  Einführung in die Statistik,
  4. Auflage, Teubner, 2004.
  ISBN: 3519320711

- R. Mathar/D. Pfeiffer:
  Stochastik für Informatiker,
  Teubner, 1990.
  ISBN: 3519022400

Kapitel 3

Bedingte Wahrscheinlichkeiten

Manchmal erhält man bei der Durchführung eines Zufallsexperimentes eine Teilinformation darüber, wie das Experiment verlaufen wird. Die Kenntnis solcher Zusatzinformationen macht eine Revision der ursprünglichen Wahrscheinlichkeitsbewertung erforderlich und führt auf den Begriff der bedingten Wahrscheinlichkeit. In der Praxis werden bedingte Wahrscheinlichkeiten verwendet, um die Wahrscheinlichkeiten komplizierter, zusammengesetzter Ereignisse zu berechnen. In diesem Zusammenhang spielt auch der Begriff der stochastischen Unabhängigkeit eine zentrale Rolle.

Schlüsselwörter: Bedingte relative Häufigkeit, bedingte Wahrscheinlichkeit, Spur-σ-Algebra, Satz von der totalen Wahrscheinlichkeit, a–priori–Verteilung, a–posteriori–Verteilung, Multiplikationssatz, stochastische Unabhängigkeit, paarweise und vollständige Unabhängigkeit.

3.1 Begriffe und Zusammenhänge

In der Praxis treten oftmals Problemstellungen auf, bei denen Teilinformationen über den Ausgang eines Zugfallsexperiments bekannt werden, bevor das endgültige Ergebnis eintritt. Diese Informationen können die Wahrscheinlichkeit für das Eintreten eines bestimmten Ereignisses beeinflussen, wie die folgenden Beispiele verdeutlichen:

3.1 Beispiel:

- Ein Pokerspiel enthält genau 52 Karten, bestehend aus den vier verschiedenen Farben ($\clubsuit$, $\spadesuit$, $\heartsuit$, $\diamondsuit$) mit jeweils 13 Karten. Eine Pokerhand erhält immer 5 Karten.

Abbildung 3.1: „Kartenspieler" von Paul Cezanne, siehe www.Heinrich-Boell.de

Gefragt wird nach der Wahrscheinlichkeit des Ereignisses F, einen „Flush" zu erhalten, d.h. dass alle Karten in einer Hand von derselben Farbe sind.

$$|\Omega| = \binom{52}{5}, \quad |F| = 4 \cdot \binom{13}{5} \Longrightarrow P(F) = \frac{4 \cdot \binom{13}{5}}{\binom{52}{5}} \approx 0.002.$$

Erhält man nun die Vorabinformation, dass alle Karten, die man bekommen hat, rot sind ($\heartsuit$, $\diamondsuit$), so ändert sich die Wahrscheinlichkeit für das Eintreten von F.

$$|\Omega'| = \binom{26}{5}, \quad |F| = 2 \cdot \binom{13}{5} \Longrightarrow P(F) = \frac{2 \cdot \binom{13}{5}}{\binom{26}{5}} \approx 0.039.$$

- Wir betrachten die Wartezeiten in der Mensa.

  Es werden zwei Essen ausgegeben. W_1 sei das Ereignis, dass die Wartezeit eines Studenten kleiner gleich 10 Minuten ist und W_2 die Wahrscheinlichkeit, dass die Wartezeit größer ist als 10 Minuten. Eine Umfrage ergab folgende Statistik:

	Anzahl	Essen I	Essen II
W_1	7000	3500	3500
W_2	3000	2500	500

  Hieraus folgt

$$h(W_1) = \frac{7000}{7000 + 3000} = 0.7 \quad \text{und} \quad h(W_2) = \frac{3000}{7000 + 3000} = 0.3.$$

Abbildung 3.2: Mensa der TU Clausthal

Kennt man die Präferenz eines Studenten, ändern sich die Wahrscheinlichkeiten gemäß

$$\begin{aligned} h(W_i \mid I) &= \frac{H(W_i \cap I)}{H(I)} = \frac{H(W_i \cap I)}{n} \cdot \frac{n}{H(I)} \\ &= \frac{h(W_i \cap I)}{h(I)} = \begin{cases} \frac{3500}{6000} \approx 0.583 & \text{für } i = 1 \\ \frac{2500}{6000} \approx 0.417 & \text{für } i = 2 \end{cases}, \end{aligned}$$

dabei sei $h(W_i|I)$ die relative Häufigkeit für das Ereignis W_i unter der Bedingung, dass der Student Essen I gewählt hat.

Für $h(B) > 0$ heißt $h(A \mid B) = \frac{h(A \cap B)}{h(B)}$ bedingte relative Häufigkeit von A unter B.

Beobachtung: Bedingte relative Häufigkeiten stabilisieren sich für $n \to \infty$ ebenso wie die relativen Häufigkeiten.

3.2 Definition (bedingte Wahrscheinlichkeit):
Es seien $(\Omega, \mathfrak{F}, P)$ *ein Wahrscheinlichkeitsraum und* $B \in \mathfrak{F}$ *mit* $P(B) > 0$. *Dann heißt für jedes* $A \in \mathfrak{F}$

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

die bedingte Wahrscheinlichkeit von A *unter* B *bzgl.* P.

Folgerung:
$A \cap B = \varnothing \Longrightarrow P(A \mid B) = 0$ und $B \subset A \Longrightarrow P(A \mid B) = 1$.

Wahrscheinlichkeitstheoretisch bedeutet der Übergang von $P(A)$ zu $P(A|B)$ die Einschränkung der ursprünglichen σ–Algebra $\mathfrak{F}$ auf das Mengensystem $\mathfrak{F}'$, das aus $\mathfrak{F}$ hervorgeht, indem man alle ihre Elemente mit B schneidet. Es stellt sich heraus, dass $\mathfrak{F}'$ eine σ–Algebra über $\Omega' = B$ ist. Entsprechend definieren $P(A|B)$ ein Wahrscheinlichkeitsmaß auf $\mathfrak{F}$ und die Einschränkung von $P(A|B)$ auf $\mathfrak{F}'$ ein Wahrscheinlichkeitsmaß auf $\mathfrak{F}'$. Diesem Nachweis dienen die beiden nachfolgenden Sätze:

3.3 Satz:
Es seien $\mathfrak{F}$ *eine* σ*–Algebra über* Ω *und* Ω' *eine nichtleere Menge mit* $\Omega' \subseteq \Omega$. *Dann ist* $\mathfrak{F}' := \{\Omega' \cap A \mid A \in \mathfrak{F}\}$ *eine* σ*–Algebra über* Ω', *die sogenannte Spur–*σ*–Algebra.*

Beweis:

1. $\Omega' \in \mathfrak{F}'$:
   Wegen $\Omega' \subseteq \Omega$ ist $\Omega' \cap \Omega = \Omega'$. Da $\Omega \in \mathfrak{F}$ ist, ist deshalb auch $\Omega' \in \mathfrak{F}'$.

2. $A' \in \mathfrak{F}' \;\Rightarrow\; \overline{A}' \in \mathfrak{F}'$:
   Es sei $A' \in \mathfrak{F}'$. Aufgrund der Definition von $\mathfrak{F}'$ gibt es ein $A \in \mathfrak{F}$ mit $A' = \Omega' \cap A$. Da $\mathfrak{F}$ eine σ–Algebra ist, ist $\overline{A} \in \mathfrak{F}$. Dann ist $\widetilde{A} = \Omega' \cap \overline{A} \in \mathfrak{F}'$. $\widetilde{A}$ ist aber das Komplement von A' in $\mathfrak{F}'$, denn es gilt:
   $$A' \cap \widetilde{A} = (\Omega' \cap A) \cap (\Omega' \cap \overline{A}) = \Omega' \cap (A \cap \overline{A}) = \Omega' \cap \varnothing = \varnothing\,,$$
   $$A' \cup \widetilde{A} = (\Omega' \cap A) \cup (\Omega' \cap \overline{A}) = \Omega' \cap (A \cup \overline{A}) = \Omega' \cap \Omega = \Omega'\,.$$

3. Für jede Folge $(A'_n)_{n\in\mathbb{N}}$ von Elementen aus $\mathfrak{F}'$ ist $\bigcup\limits_{n\in\mathbb{N}} A'_n \in \mathfrak{F}'$:
   Es sei $(A'_n)_{n\in\mathbb{N}}$ eine Folge von Elementen aus $\mathfrak{F}'$. Dann gibt es aufgrund der Definition von $\mathfrak{F}'$ eine Folge $(A_n)_{n\in\mathbb{N}}$ von Elementen aus $\mathfrak{F}$, so dass $A'_n = \Omega' \cap A_n$ für alle $n \in \mathbb{N}$ gilt. Da $\mathfrak{F}$ eine σ–Algebra ist, ist $\bigcup_{n\in\mathbb{N}} A_n \in \mathfrak{F}$. Mit der Definition von $\mathfrak{F}'$ folgt weiter:
   $$\Omega' \cap \bigcup_{n\in\mathbb{N}} A_n \in \mathfrak{F}'.$$
   Es gilt aber
   $$\Omega' \cap \bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} (\Omega' \cap A_n) = \bigcup_{n\in\mathbb{N}} A'_n, \quad \text{also} \quad \bigcup_{n\in\mathbb{N}} A'_n \in \mathfrak{F}.$$

■

3.4 Satz:
Es seien $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $B \in \mathfrak{F}$ mit $P(B) > 0$. $P(\cdot \mid B)\colon \mathfrak{F} \to \mathbb{R}$ definiert durch
$$P(A \mid B) := \frac{P(A \cap B)}{P(B)} \qquad (A \in \mathfrak{F})$$
ein Wahrscheinlichkeitsmaß über Ω.

Beweis:

1) Aus der Definition der bedingten Wahrscheinlichkeit folgt für alle $A \in \mathfrak{F}$ sofort $P(A \mid B) \geq 0$.

2) Es gilt $P(\Omega \mid B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$, d.h. $P(\cdot \mid B)$ ist normiert.

3) Es sei $(A_i)_{i\in\mathbb{N}}$ eine Folge von paarweise fremden Elementen von $\mathfrak{F}$. Dann gilt:
$$\begin{aligned} P\left(\bigcup_{i\in\mathbb{N}} A_i \mid B\right) &= \frac{P\left(\left(\bigcup_{i\in\mathbb{N}} A_i\right) \cap B\right)}{P(B)} = \frac{P\left(\bigcup_{i\in\mathbb{N}} (A_i \cap B)\right)}{P(B)} \\ &= \frac{\sum_{i\in\mathbb{N}} P(A_i \cap B)}{P(B)} \quad (\sigma\text{–Additivität von } P) \\ &= \sum_{i\in\mathbb{N}} P(A_i \mid B) \end{aligned}$$
und es folgt die σ–Additivität von $P(\cdot \mid B)$. ■

Dieser Satz rechtfertigt die Bezeichnung bedingte **Wahrscheinlichkeit**.

3.5 Beispiel (Regenwahrscheinlichkeit):
An einer Wetterstation (nicht in Clausthal) wurde die Regenwahrscheinlichkeit in Abhängigkeit des Monats aufgezeichnet:

Monat	Jan	Feb	Mär	Apr	Mai	Jun
Wahrscheinlichkeit für Regentag	15%	15%	25%	25%	20%	25%

Monat	Jul	Aug	Sep	Okt	Nov	Dez
Wahrscheinlichkeit für Regentag	30%	20%	25%	30%	25%	25%

Bekannt sind also die bedingten Wahrscheinlichkeiten

$$P(\text{„Regentag"}|\text{Monat} = i) \quad \text{für } i = 1, \dots, 12.$$

Gesucht ist die durchschnittliche Regenwahrscheinlichkeit für das ganze Jahr.
Mit A_i :=„ein zufällig gewählter Tag liegt in Monat i" ist $P(A_1) = \frac{31}{365}$, $P(A_2) = \frac{28}{365}$, $P(A_3) = \frac{30}{365}$, $P(A_4) = \frac{31}{365}$ usw. und damit lässt sich allgemein schreiben:

$$P(\text{„Regentag"}) = \sum_{i=1}^{12} P(\text{„Regentag"}|\text{Monat} = i) \cdot P(A_i).$$

Also:

$$\begin{aligned} P(\text{„Regentag"}) &= \frac{31}{365} \cdot P(\text{„Regentag"}|\text{Monat} = 1) + \\ &\quad \frac{28}{365} \cdot P(\text{„Regentag"}|\text{Monat} = 2) + \\ &\quad \frac{31}{365} \cdot P(\text{„Regentag"}|\text{Monat} = 3) + \\ &\quad \frac{30}{365} \cdot P(\text{„Regentag"}|\text{Monat} = 4) + \dots \\ &\approx 0.234. \end{aligned}$$

In dem Beispiel wurde also rückwärts von einer bedingten Wahrscheinlichkeit auf eine unbedingte Wahrscheinlichkeit geschlossen. Dieser Sachverhalt soll nun im folgenden Satz allgemein bewiesen werden:

3.6 Satz (Satz von der totalen Wahrscheinlichkeit):
Es seien $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $(B_n)_{n\in\mathbb{N}}$ eine Folge paarweise fremder Mengen aus $\mathfrak{F}$ mit $\Omega = \bigcup_{n\in\mathbb{N}} B_n$ und $P(B_n) > 0 \;\forall\, n \in \mathbb{N}$. Dann gilt für jedes Ereignis $A \in \mathfrak{F}$:

$$P(A) = \sum_{n\in\mathbb{N}} P(A \mid B_n) \cdot P(B_n).$$

Beweis:
Es sei $A = A \cap \Omega = A \cap \bigcup\limits_{n\in\mathbb{N}} B_n = \bigcup\limits_{n\in\mathbb{N}} (A \cap B_n)$ und $(A \cap B_n) \cap (A \cap B_m) = \varnothing$ für $m \neq n$. Es folgt

$$P(A) = P\left(\bigcup_{n\in\mathbb{N}} (A \cap B_n)\right) = \sum_{n\in\mathbb{N}} P(A \cap B_n) \overset{\text{Def. 3.2}}{=} \sum_{n\in\mathbb{N}} P(A \mid B_n) \cdot P(B_n).$$

■

Mit Hilfe des Satzes von der totalen Wahrscheinlichkeit kann von bedingten Wahrscheinlichkeiten $P(A|B_i)$ auf die unbedingte Wahrscheinlichkeit $P(A)$ geschlossen werden (sofern die $P(B_i)$ bekannt sind). Im Folgenden soll noch ein Schritt weiter gegangen werden: Es soll jetzt $P(B_i|A)$ bestimmt werden. Gerade diese Fragestellung hat sehr viele praktische Anwendungen, wie sich in Beispiel 3.9 zeigen wird.

3.7 Satz (Formel von Bayes):
Es seien $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $(B_n)_{n\in\mathbb{N}}$ eine Folge paarweise fremder Mengen aus $\mathfrak{F}$ mit $\Omega = \bigcup_{n\in\mathbb{N}} B_n$ und $P(B_n) > 0 \;\forall\, n \in \mathbb{N}$. Dann gilt für jedes Ereignis $B \in \mathfrak{F}$ mit $P(B) > 0$ und alle $k \in \mathbb{N}$:

$$P(B_k \mid B) = \frac{P(B \mid B_k) \cdot P(B_k)}{\sum\limits_{n\in\mathbb{N}} P(B \mid B_n) \cdot P(B_n)}.$$

(Siehe auch Lebensdaten von Bayes im Anhang D.)

Beweis:
Aus Satz 3.6 folgt

$$P(B) = \sum_{n\in\mathbb{N}} P(B \mid B_n) \cdot P(B_n)\,.$$

Außerdem ist

$$P(B_k \cap B) = P(B_k \mid B) \cdot P(B) = P(B \mid B_k) \cdot P(B_k)\,.$$

Es ergibt sich

$$P(B_k \mid B) = \frac{P(B_k \cap B)}{P(B)} = \frac{P(B \mid B_k) \cdot P(B_k)}{\sum\limits_{n\in\mathbb{N}} P(B \mid B_n) \cdot P(B_n)}\,.$$

■

3.8 Definition (a–priori–Verteilung, a–posteriori–Verteilung):

- $(P(B_k))_{k\in N}$ *nennt man a–priori–Verteilung (vor dem Eintreten von B).*
- $(P(B_k \mid B))_{k\in N}$ *nennt man a–posteriori–Verteilung (nach dem Eintreten von B).*

3.9 Beispiel:
„Let's make a deal" (3-Tore-Problem):
In einer Spielshow wird der Kandidat vor drei verschlossene Türen gestellt. Hinter einer der Türen wartet ein Gewinn, in allen anderen Fällen geht der Kandidat leer aus. Der Kandidat wählt eine Tür aus, die aber nicht geöffnet wird. Stattdessen öffnet der Quizmaster eine der beiden anderen Türen, hinter der sich jedoch nichts befindet. Nun wird der Kandidat vor die Wahl gestellt, bei der gewählten Tür zu bleiben oder sich für die andere, noch nicht geöffnete Tür zu entscheiden. Es stellt sich die Frage, ob der Kandidat seine Gewinnchance erhöhen kann, wenn er sich umentscheidet.
Zur Klärung des Sachverhalts werden folgende Bezeichnungen eingeführt:
U :=„eine Umentscheidung führt zu Gewinn" und r :=„richtige Tür war von Anfang an gewählt". Wurde bereits von Anfang an die richtige Tür gewählt, so führt eine Umentscheidung zum Verlust ($P(U|r) = 0$). Wurde allerdings von Anfang an die falsche Tür gewählt, so führt eine Umentscheidung zum Gewinn ($P(U|\overline{r}) = 1$). Die Wahrscheilichkeit, von Anfang an

die richtige Tür gewählt zu haben, beträgt $\frac{1}{3}$, also $P(r) = \frac{1}{3}$. Damit ergibt sich mit dem Satz von der totalen Wahrscheinlichkeit:

$$P(U) = P(U|r) \cdot P(r) + P(U|\overline{r}) \cdot P(\overline{r}) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}.$$

Das Ergebnis zeigt, dass sich im Allgemeinen die Gewinnwahrscheinlichkeit erhöht, wenn sich der Kandidat umentscheidet.

Fertigungs–Ausschussanteil:
Die Produktion in einer Fabrik erfolgt durch drei Maschinen. Diese sind zu verschiedenen Anteilen an der Gesamtproduktion beteiligt und weisen folgende Ausschussraten auf:

	Maschine A	Maschine B	Maschine C
Anteil an der Gesamtproduktion	20%	50%	30%
Ausschussrate	1%	3%	4%

Abbildung 3.3: Eine der drei Maschinen

Demnach lauten die Parameter des Systems:

$$\begin{array}{llllll} P(A) & = 20\%, & P(B) & = 50\%, & P(C) & = 30\% \\ P(\text{,,Ausschuss''}|A) & = \ \ 1\%, & P(\text{,,Ausschuss''}|B) & = \ \ 3\%, & P(\text{,,Ausschuss''}|C) & = \ \ 4\%. \end{array}$$

Wir berechnen zunächst den Ausschussanteil bezogen auf die gesamte Produktion. Dieser lässt sich mit Hilfe des Satzes von der totalen Wahrscheinlichkeit berechnen:

$$\begin{aligned} P(\text{,,Ausschuss''}) &= \sum_{M \in \{A,B,C\}} P(\text{,,Ausschuss''}|M) \cdot P(M) \\ &= 1\% \cdot 20\% + 3\% \cdot 50\% + 4\% \cdot 30\% = 2.9\%. \end{aligned}$$

Mit Hilfe des Satzes von Bayes lässt sich jetzt auch noch bestimmen, wie gross die Wahrscheinlichkeit dafür ist, dass ein Ausschussteil von Maschine A stammt:

$$P(A|\text{,,Ausschuss''}) = \frac{P(\text{,,Ausschuss''}|A) \cdot P(A)}{\sum_{M \in \{A,B,C\}} P(\text{,,Ausschuss''}|M) \cdot P(M)} = \frac{1\% \cdot 20\%}{2{,}9\%} \approx 6{,}90\%.$$

Die entsprechenden Wahrscheinlichkeiten für die anderen beiden Maschinen sind:

$$\begin{aligned} P(B|\text{,,Ausschuss''}) &= \frac{3\% \cdot 50\%}{2{,}9\%} \approx 51{,}72\% \quad \text{und} \\ P(C|\text{,,Ausschuss''}) &= \frac{4\% \cdot 30\%}{2{,}9\%} \approx 41{,}38\%. \end{aligned}$$

Multiple–Choice–Test:
Eine Prüfungsfrage in einem Multiple–Choice–Test hat n mögliche Antworten, von denen k $(0 \leq k \leq n)$ richtig sind. Hat ein Student sich vorbereitet, so sollte er die richtigen Antworten auswählen können. Hat ein Student sich nicht vorbereitet, so muss er raten und wählt j der n möglichen Antworten zufällig aus.

Es sei $P(\text{,,Student hat sich vorbereitet''}) =: p$.

Gesucht ist die Wahrscheinlichkeit (in Abhängigkeit von p und n), dass der Student sich vorbereitet hat, wenn man weiß, dass er die richtigen Antworten gewählt hat.

Lösungsmenge: $\Omega := \{(0,0); (0,1); (1,0); (1,1)\}$

(0,0) := „nicht vorbereitet, falsche Antwort“

(0,1) := „nicht vorbereitet, richtige Antwort“

(1,0) := „vorbereitet, falsche Antwort“

(1,1) := „vorbereitet, richtige Antwort“

$A := \{(0,1); (1,1)\} =$ „richtige Antwort“

$B := \{(1,0); (1,1)\} =$ „der Student hat sich vorbereitet“

Es gilt $P(B) = p \in (0,1]$, außerdem wird angenommen, dass es Studenten gibt, die die richtigen Antworten gegeben haben, d.h. dass $P(A) > 0$ gilt. Ein Student, der sich vorbereitet hat, gebe immer die richtigen Antworten, d.h. $P(A \mid B) = 1$ und ein nicht vorbereiteter Student wähle zufällig j der n Antworten. In diesem Fall ist $P(A \mid \overline{B}) = \frac{1}{2^n}$, da er bei jeder der n möglichen Antworten raten muss, ob sie richtig ist oder nicht und dabei jeweils eine Chance von $\frac{1}{2}$ hat. Gesucht ist $P(B \mid A)$. Zur Lösung des Problems wird die Formel von Bayes verwendet:

$$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A \mid B) \cdot P(B) + P(A \mid \overline{B}) \cdot P(\overline{B})} = \frac{1 \cdot p}{1 \cdot p + \frac{1}{2^n} \cdot (1-p)} = \frac{p}{p + \frac{1-p}{2^n}} \xrightarrow[n \to \infty]{} 1.$$

D.h., je größer die Anzahl der Antworten ist, desto wahrscheinlicher ist es, dass der Student sich vorbereitet hat, wenn man weiß, dass er die richtige Antwort gegeben hat.

Beachte: Es wurde in dem Beispiel vorausgesetzt, dass die n möglichen Antworten der Frage unabhängig voneinander sind, d.h. es gibt keine Antworten der Form ,,Es gilt A.” und ,,Es gilt nicht A.” Wenn es Antworten gibt, die sich gegenseitig ausschließen, so gilt nicht mehr $P(A \mid \overline{B}) = \frac{1}{2^n}$, sondern nur noch $\frac{1}{2} \geq P(A \mid \overline{B}) > \frac{1}{2^n}$.

Da aber in jedem Fall $P(\text{Student hat eine Frage richtig beantwortet} \mid \overline{B}) < 1$ ist, lässt sich trotzdem $P(B \mid A) \to 1$ durch die Erhöhung der Anzahl der Fragen erreichen.

Spam–Filter:
In der eMail–Kommunikation stellen Spam-Mails (unerwünschte Werbemails) ein ständig

wachsendes Ärgernis dar. Der naheliegenste Ansatz, Spam automatisiert auszufiltern, besteht darin, eine Liste von Wörtern anzulegen, die fast nur in Spam-Mails vorkommen. Tritt eines der Wörter aus dieser Liste in der Mail auf, so wird diese als Spam aussortiert. Mit einem solchen Ansatz werden aber auch erwünschte Mails aussortiert, sofern diese eines der Schlüsselwörter enthalten. Außerdem muss die Liste ständig aktualisiert und erweitert werden, um möglichst viel Spam zu erfassen.

Einen anderen Ansatz, der diese beiden Nachteile weitgehend ausgleicht, stellen die lernfähigen Bayes–Filter dar. (Die Spam-Filter von Mozilla und Thunderbird funktionieren nach diesem Prinzip.) Zunächst muss der Filter einige Zeit trainiert werden, indem man Mails manuell als Spam markiert. Der Spam–Filter liest alle Spam– und alle Nicht–Spam–Mails und zählt dabei, wie häufig welches Wort auftritt. (Dabei werde jedes Wort nur jeweils ein Mal pro Mail gezählt).

Nach dem Scannen von 2000 Mails (je 1000 Spam– und Nicht–Spam–Mails) könnte sich z.B. folgende Tabelle mit Wort–Häufigkeiten ergeben haben:

	absolute Häufigkeit		
Wort	**in Spam–Mails**	**in Nicht–Spam–Mails**	**gesamt**
Viagra	500	0	500
enlargement	400	0	400
money	350	3	353
buy	300	4	304
credit	250	10	260
⋮			
eigener Name	250	800	1050
Software	150	200	350
⋮			
Dir	0	500	500
und	10	700	710
der	0	950	950
die	0	950	950

Mit Hilfe dieser Daten lassen sich nun u.a. folgende bedingte Wahrscheinlichkeiten aufstellen:

$$
\begin{aligned}
P(\text{,,Viagra'' tritt auf} \mid \text{Spam}) &= 50\% \\
P(\text{,,enlargement'' tritt auf} \mid \text{Spam}) &= 40\% \\
P(\text{,,money'' tritt auf} \mid \text{Spam}) &= 35\% \\
P(\text{,,buy'' tritt auf} \mid \text{Spam}) &= 30\% \\
P(\text{,,credit'' tritt auf} \mid \text{Spam}) &= 25\% \\
P(\text{,,}\textit{eigener Name}\text{'' tritt auf} \mid \text{Spam}) &= 25\% \\
&\vdots \\
P(\text{,,der'' tritt auf} \mid \text{kein Spam}) &= 95\% \\
P(\text{,,die'' tritt auf} \mid \text{kein Spam}) &= 95\% \\
P(\text{,,}\textit{eigener Name}\text{'' tritt auf} \mid \text{kein Spam}) &= 80\% \\
P(\text{,,und'' tritt auf} \mid \text{kein Spam}) &= 70\%
\end{aligned}
$$

$$P(\text{„Dir" tritt auf} \mid \text{kein Spam}) = 50\%.$$
$$\vdots$$

Das Verhältnis von Spam zu normalen Mails sei bekannt. In diesem Beispiel sind es 50%. Für die absoluten Eintrittswahrscheinlichkeiten gilt dann:

$$\begin{aligned} P(\text{„money" tritt auf}) &= 17{,}65\% \\ P(\text{„credit" tritt auf}) &= 13\% \\ P(\text{„}\textit{eigener Name}\text{" tritt auf}) &= 52{,}5\% \\ &\vdots \end{aligned}$$

Erhält man nun eine neue Mail, so wird diese vom Spam–Filter nach den bekannten Schlüsselwörtern durchsucht. Wenn die Mail jetzt die Wörter „money", „credit" oder den Namen des Empfängers enthält (und sonst keine bekannten Schlüsselwörter), so lässt sich mit Hilfe der Formel von Bayes Folgendes berechnen:

$$\begin{aligned} P(\text{Spam}|\text{„money" tritt auf}) &= \frac{P(\text{„money" tritt auf} \mid \text{Spam}) \cdot P(\text{Spam})}{P(\text{„money" tritt auf})} \\ &= \frac{35\% \cdot 50\%}{17{,}65\%} \approx 99,15\%, \\ P(\text{Spam}|\text{„credit" tritt auf}) &= \frac{P(\text{„credit" tritt auf} \mid \text{Spam}) \cdot P(\text{Spam})}{P(\text{„credit" tritt auf})} \\ &= \frac{25\% \cdot 50\%}{13\%} \approx 96,15\%, \\ P(\text{Spam}|\text{„}\textit{eigener Name}\text{" tritt auf}) &= \frac{P(\text{„}\textit{eigener Name}\text{" tritt auf} \mid \text{Spam}) \cdot P(\text{Spam})}{P(\text{„}\textit{eigener Name}\text{" tritt auf})} \\ &= \frac{25\% \cdot 50\%}{52{,}5\%} \approx 23{,}81\%. \end{aligned}$$

Abschließend wird der Mittelwert über alle drei Wahscheinlichkeiten gebildet:

$$\frac{1}{3} \cdot (P(\text{Spam}|\text{„money"}) + P(\text{Spam}|\text{„credit"}) + P(\text{Spam}|\text{„}\textit{eigener Name}\text{"})) \approx 73{,}04\%.$$

Überschreitet dieser Wert eine vorgegebene Grenze, so wird die Mail als Spam aussortiert. Zur technischen Ausführung und zu weiteren Details siehe:

- Paul Graham: „A Plan for Spam"
- Gary Robinson: „Spam Detection"

3.10 Satz (Multiplikationssatz):
Sei $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum. Für $A_0, \ldots, A_n \in \mathfrak{F}$ gelte $P(A_0 \cap A_1 \cap \cdots \cap A_{n-1}) > 0$. Dann gilt:

$$P\left(\bigcap_{j=0}^{n} A_j\right) = P(A_0) \cdot P(A_1 \mid A_0) \cdot P(A_2 \mid A_0 \cap A_1) \cdot \cdots \cdot P(A_n \mid A_0 \cap A_1 \cap \cdots \cap A_{n-1}).$$

Beweis:
Mit $A_0 \cap A_1 \cap \cdots \cap A_{n-1} \subset A_0 \cap A_1 \cap \cdots \cap A_{n-2} \subset \cdots \subset A_0 \cap A_1 \subset A_0$ folgt

$$0 < P(A_0 \cap \cdots \cap A_{n-1}) \leq P(A_0 \cap \cdots \cap A_{n-2}) \leq \cdots \leq P(A_0).$$

Der Rest wird durch Induktion bewiesen. Den Induktionsanfang stellt die Definition der bedingten Wahrscheinlichkeit dar:

$$P(A_0 \cap A_1) = P(A_0) \cdot P(A_1 \mid A_0).$$

Der Induktionsschluss von n auf $n+1$ erfolgt durch:

$$\begin{aligned} P(A_0 \cap \cdots \cap A_n \cap A_{n+1}) &= P(A_0 \cap \cdots \cap A_n) \cdot P(A_{n+1} \mid A_0 \cap \cdots \cap A_n) \quad \text{(Def.)} \\ &= P(A_0) \cdot \cdots \cdot P(A_n \mid A_0 \cap \ldots \cap A_{n-1}) \cdot P(A_{n+1} \mid A_0 \cap \cdots \cap A_n). \end{aligned}$$

■

3.11 Beispiel (Urnenmodell von Pólya):
Gegeben sei eine Urne, die r rote Kugeln und s schwarze Kugeln enthält. Nach der zufälligen Entnahme einer Kugel wird diese mit c weiteren desselben Typs zurück in die Urne gelegt.

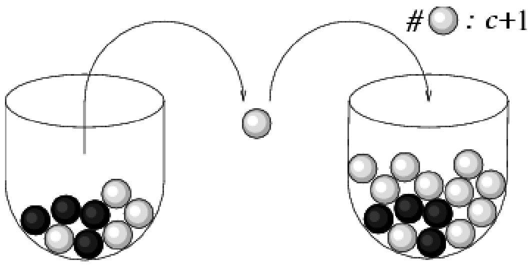


Abbildung 3.4: Urnenmodell von Pólya

Gesucht ist die Wahrscheinlichkeit dafür, bei n Ziehungen jeweils eine rote Kugel zu ziehen. Hierzu wird das Ereignis $A_i :=$ „im i–ten Versuch eine rote Kugel ziehen“ definiert. Nun kann die gesuchte Wahrscheinlichkeit wie folgt beschrieben werden: $P(A_1 \cap A_2 \cap \cdots \cap A_n)$. Mit Hilfe des Multiplikationssatzes lässt sich die gesuchte Wahrscheinlichkeit berechnen:

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) &= P(A_1) \cdot P(A_2 \mid A_1) \cdot \cdots \cdot P(A_n \mid A_1 \cap \cdots \cap A_{n-1}) \\ &= \frac{r}{r+s} \cdot \frac{r+c}{r+s+c} \cdot \frac{r+2c}{r+s+2c} \cdot \ldots \cdot \frac{r+(n-1)c}{r+s+(n-1)c}. \end{aligned}$$

(Siehe auch PowerPoint-Präsentation zu diesem Beispiel.)

3.2 Stochastische Unabhängigkeit

Diese Abschnitt behandelt den Begriff der stochastischen Unabhängigkeit. Als Motivation diene das folgende Beispiel:
Es wird eine Population Ω medizinischer Probanden betrachtet, in der sich solche befinden, die rauchen (R), die nicht rauchen ($\overline{R}$), die an Lungenkrebs erkrankt sind (L) oder nicht erkrankt sind ($\overline{L}$). Unter der Annahme, dass das Rauchen keinen Einfluss auf die Entstehung von

Lungenkrebs habe, erwartet man, dass in der Gruppe der Raucher genauso viele Probanden an Lungenkrebs erkrankt sind wie in der Gruppe der Nichtraucher, d.h. im Fall einer hinreichend großen Population sollte gelten:

$$\begin{aligned}
& \frac{|L \cap R|}{|R|} = \frac{|L \cap \overline{R}|}{|\overline{R}|} \\
\iff \quad & \frac{|L \cap R| \,/\, |\Omega|}{|R| \,/\, |\Omega|} = \frac{|L \cap \overline{R}| \,/\, |\Omega|}{|\overline{R}| \,/\, |\Omega|} \\
\text{bzw.} \quad & \frac{P(L \cap R)}{P(R)} = \frac{P(L \cap \overline{R})}{P(\overline{R})} \\
\iff \quad & P(L \mid R) = P(L \mid \overline{R})\,.
\end{aligned}$$

Diese Feststellung gibt Anlass zu folgender Definition:

3.12 Definition (stochastisch unabhängig):
Es sei $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $A, B \in \mathfrak{F}$ mit $P(B) > 0$ und $P(\overline{B}) > 0$. A heißt stochastisch unabhängig von B bzgl. P, falls

$$P(A \mid B) = P(A \mid \overline{B})$$

gilt.

3.13 Satz:
Es sei $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $A, B \in \mathfrak{F}$ mit $P(B) > 0$. Dann sind folgende Aussagen äquivalent:

a) $P(A \mid B) = P(A \mid \overline{B})$, *falls* $P(\overline{B}) > 0$.

b) $P(A \mid B) = P(A)$.

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

Beweis:
a) $\Longrightarrow$ b):

$$\begin{aligned}
P(A) &= P(A \mid B) \cdot P(B) + P(A \mid \overline{B}) \cdot P(\overline{B}) \\
&= P(A \mid B) \cdot P(B) + P(A \mid B) \cdot P(\overline{B}) \\
&= P(A \mid B) \cdot \big(P(B) + P(\overline{B})\big) = P(A \mid B) \cdot P(\Omega) = P(A \mid B)\,.
\end{aligned}$$

b) $\Longrightarrow$ a):

$$\begin{aligned}
P(A) &= P(A \mid B) \cdot P(B) + P(A \mid \overline{B}) \cdot P(\overline{B}) \\
&= P(A) \cdot P(B) + P(A \mid \overline{B}) \cdot P(\overline{B}) \\
\iff \quad P(A) \cdot (1 - P(B)) &= P(A \mid \overline{B}) \cdot P(\overline{B}) \\
\iff \quad P(A) \cdot P(\overline{B}) &= P(A \mid \overline{B}) \cdot P(\overline{B}) \\
\iff \quad P(A) &= P(A \mid \overline{B})\,.
\end{aligned}$$

b) $\Longrightarrow$ c):

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \Longrightarrow \quad P(A \cap B) = P(A \mid B) \cdot P(B) = P(A) \cdot P(B)\,.$$

c) $\Longrightarrow$ b):

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

■

3.14 Beispiel:
Es wird das Werfen eines Würfels betrachtet. Es seien $A := \{1,3,5\}$ und $B := \{3,4,5,6\}$ und somit $P(A) = 1/2$ und $P(B) = 2/3$. A und B sind unabhängig, da

$$P(A \cap B) = P(\{3,5\}) = \frac{1}{3} = \frac{1}{2} \cdot \frac{2}{3} = P(A) \cdot P(B)\,.$$

3.15 Bemerkung:
Die stochastische Unabhängigkeit zweier Ereignisse geht beim Übergang zu einem anderen Wahrscheinlichkeitsmaß im allgemeinen verloren.

3.16 Beispiel:
Es seien $\Omega := \{1,2,3\}$, $A := \{1\}$ und $B := \{1,2\}$. Man wähle zwei Wahrscheinlichkeitsmaße P und P' wie folgt: $P(\{1\}) := 1$, $P(\{2\}) := P(\{3\}) := 0$ und $P'(\{i\}) := \frac{1}{3}$ für $i = 1,2,3$.
Unter dem Wahrscheinlichkeitsmaß P sind A und B stochastisch unabhängig, denn es gilt:

$$P(A \cap B) = P(\{1\}) = 1 = 1 \cdot 1 = P(A) \cdot P(B).$$

Unter P' sind A und B jedoch nicht unabhängig:

$$P'(A \cap B) = P'(\{1\}) = \frac{1}{3} \neq \frac{2}{9} = \frac{1}{3} \cdot \frac{2}{3} = P'(A) \cdot P'(B).$$

3.17 Definition (paarweise Unabhängigkeit):
Endlich viele Ereignisse $A_i \in \mathfrak{F}$ $(i = 1, \ldots, n)$ heißen <u>*paarweise stochastisch unabhängig*</u> *bzgl. P, wenn gilt:*

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j) \quad \textit{für } i \neq j.$$

Eine Frage in diesem Zusammenhang ist, ob aus $P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3)$ auch die paarweise Unabhängigkeit der Ereignisse A_1, A_2 und A_3 folgt. Dieses Problem führt zum Begriff der vollständigen Unabhängigkeit (siehe Definition 3.19).

3.18 Beispiel:
Es seien $\Omega := \{1,2,3\}$, $P(\{1\}) := \frac{1}{2}$, $P(\{2\}) := \frac{1}{2}$, $P(\{3\}) := 0$, $A_1 = \{1\}$, $A_2 := \{2\}$ und $A_3 = \{3\}$. Dann gilt:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3),$$

aber

$$P(A_1 \cap A_2) = 0 \neq \frac{1}{4} = P(A_1) \cdot P(A_2).$$

3.19 Definition (vollständig stochastisch unabhängig):
Endlich oder abzählbar unendlich viele Ereignisse $A_i \in \mathfrak{F}$ heißen vollständig stochastisch unabhängig bzgl. P, wenn gilt:

$$P(A_{i_1} \cap \cdots \cap A_{i_m}) = P(A_{i_1}) \cdot \cdots \cdot P(A_{i_m})$$

für je endlich viele paarweise verschiedene Indizes $i_1, \ldots, i_m \in \mathbb{N}$, $m \in \mathbb{N}$.

3.20 Bemerkung:
Die vollständige Unabhängigkeit ist eine schärfere Forderung als die paarweise Unabhängigkeit. Aus der paarweisen Unabhängigkeit folgt nicht notwendig auch die vollständige Unabhängigkeit.

3.21 Beispiel:
In einer Urne befinden sich vier Lose mit den Zahlen 6, 7, 10 und 15. Der Urne wird zufällig ein Los entnommen. T_k sei das Ereignis, dass die gezogene Zahl durch k teilbar ist. Es gilt:

$$\begin{aligned}
P(T_2) &= P(\{6, 10\}) = 0.5 \\
P(T_3) &= P(\{6, 15\}) = 0.5 \\
P(T_5) &= P(\{10, 15\}) = 0.5 \\
P(T_2 \cap T_3) &= P(\{6\}) = 0.25 = 0.5 \cdot 0.5 = P(T_2) \cdot P(T_3) \\
P(T_2 \cap T_5) &= P(\{10\}) = 0.25 = 0.5 \cdot 0.5 = P(T_2) \cdot P(T_5) \\
P(T_3 \cap T_5) &= P(\{15\}) = 0.25 = 0.5 \cdot 0.5 = P(T_3) \cdot P(T_5) \\
P(T_2 \cap T_3 \cap T_5) &= P(\varnothing) = 0 \neq 0.5 \cdot 0.5 \cdot 0.5 = P(T_2) \cdot P(T_3) \cdot P(T_5)\,.
\end{aligned}$$

(Siehe auch PowerPoint-Präsentation zu diesem Beispiel.)

Literatur zu Kapitel 3

Folgende Bücher werden als begleitende Literatur zu diesem Kapitel des Skriptes empfohlen:

- M. Fisz:
  Wahrscheinlichkeitsrechnung und mathematische Statistik,
  VEB, Deutscher Verlag der Wissenschaften 1989.
  ISBN: 3326000790

- P. Gänssler und W. Stute:
  Wahrscheinlichkeitstheorie,
  Springer–Verlag, Berlin, 1977.
  ISBN: 3540084185

- H. O. Georgii:
  Stochastik,
  2. Auflage, de Gruyter, 2004.
  ISBN: 3110172356

- K. Hinderer:
  Grundbegriffe der Wahrscheinlichkeitstheorie,
  Springer–Verlag, 1980.
  ISBN: 3540073094

- G. Hübner:
  Stochastik. Eine Einführung für Mathematiker, Informatiker und Ingenieure.,
  4. Auflage, Vieweg Verlag, 2003.
  ISBN: 3528254432

- U. Krengel:
  Einführung in die Wahrscheinlichkeitstheorie und Statistik,
  6. Auflage, Vieweg, 2002.
  ISBN: 3528672595

- J. Lehn/H. Wegmann:
  Einführung in die Statistik,
  4. Auflage, Teubner, 2004.
  ISBN: 3519320711

- R. Mathar/D. Pfeiffer:
  Stochastik für Informatiker,
  Teubner, 1990.
  ISBN: 3519022400

Kapitel 4

Statistische Methoden der Qualitätssicherung

Sowohl während seiner Entwicklung als auch während seiner Herstellung durchläuft ein Produkt verschiedene Qualifizierungsstufen. Die Sicherstellung seiner technischen Spezifikationen durch prozessintegrierte Meßsysteme und computergestützte Auswerteverfahren stellt für alle Unternehmen eine besondere Herausforderung dar. Besonders Großserien–Hersteller wie PC–, Automobil-, Elektronik- oder Lebensmittelhersteller müssen sich darauf verlassen können, dass die von ihnen verwendeten Komponenten und Zutaten einwandfrei sind. Aber auch der Kunde möchte sicher sein, ein fehlerfreies Produkt zu erwerben. Eventuell erforderlich werdende Garantieleistungen und Rückrufaktionen sind nicht nur mit hohen Kosten sondern auch mit einem erheblichen Vertrauensschwund bei den Kunden verbunden. Aus diesem Grund betreiben fast alle großen Firmen ein umfassendes Qualitätsmanagement, das die mit der Qualitätsüberwachung einhergehenden Geschäftsprozesse koordiniert und die für die Erfassung und Auswertung der Qualitätsdaten erforderlichen Methoden und Verfahren zur Verfügung stellt. Aufgrund der Datenmengen werden dabei auch große Anforderungen an die Informationstechnik gestellt.
Die Organisation des Qualitätsmanagement und die Installation von Qualitätssystemen wird in verschiedenen Richtlinien geregelt. Entsprechende Empfehlungen findet man u.a. in der DIN ISO 900x. Unter dem Druck des Wettbewerbs sehen sich viele Unternehmen gezwungen, ihre Qualitätssysteme von unabhängigen Institutionen zertifizieren zu lassen.

Schlüsselwörter: Gut–Schlecht–Prüfung, Stichprobenplan, Stichprobe, Annahmezahl, Test, Null–Hypothese, Alternativ–Hypothese, Lieferantenrisiko, Konsumentenrisiko, Fehler 1. Art, Fehler 2. Art, Signifikanzniveau, Gütefunktion, Operationscharakteristik, Gutgrenze, Schlechtgrenze, Steilheit, Indifferenzpunkt, Philips–Stichprobenplan, Maximaler mittlerer Durchschlupf, Mittlerer Prüfaufwand.

4.1 Hypothesentest

Die statistische Qualitätskontrolle ist ein Teilgebiet der Stochastik. Sie beschäftigt sich sowohl mit Methoden für die laufende Prozesskontrolle als auch mit Verfahren für die Wareneingangs- und Endkontrolle. Im ersten Fall wird anhand von Messungen überprüft, ob die beobachteten Toleranzen noch akzeptabel sind oder ein Eingreifen in den Prozess erforderlich machen (sogenannte messende oder Attribut–Prüfung). Im zweiten Fall geht es um die Frage, ob die angelieferten Komponenten bzw. die ausgelieferten Produkte den vereinbarten Qualitätsanforderungen genügen oder nicht (sogenannte zählende oder Gut–Schlecht–Prüfung). Dieser Abschnitt beschäftigt sich zunächst mit der Gut–Schlecht–Prüfung.

Der mit einer Vollkontrolle einhergehende technische und personelle Aufwand steht oftmals in keinem Verhältnis zum Erlös, der mit dem Produkt erzielt werden kann, oder zum Risiko, das mit dem Versagen des Produkts verbunden ist. Hinzu kommt, dass viele Prüfverfahren zerstörenden Charakter haben, weshalb eine Vollkontrolle nicht in Frage kommt. Umfangreiche Tests verlängern außerdem die Durchlaufzeiten durch die Fertigung, was im Rahmen einer schlanken Produktion nicht erwünscht ist. Die mathematischen Verfahren der Gut–Schlecht–Prüfung zielen darauf ab, die mit der Qualitätsüberprüfung verbundenen Kosten aufgrund einer Stichprobenprüfung zu reduzieren und die Möglichkeit einer Fehlentscheidung unter Kontrolle zu halten.

Abbildung 4.1: Prinzip der Stichprobenprüfung

Im Fall einer Stichprobenprüfung stellt sich die Frage, wie man anhand der Anzahl fehlerhafter Stücke in der Stichprobe auf den Ausschussanteil in der gesamten Lieferung schließen kann. Diese Frage soll mit Hilfe von statistischen Tests beantwortet werden. Unter einem statistischen Test versteht man ein Verfahren zur Überprüfung einer Hypothese über den Parameter einer Wahrscheinlichkeitsverteilung. Im einfachsten Fall entnimmt man dem Los vom Umfang N eine zufällige Stichprobe vom Umfang n und verwendet als sogenannte Testgröße X die Anzahl fehlerhafter Stücke in der Stichprobe. Liegt X unterhalb einer kritischen Grenze c (sogenannte Annahmezahl) wird die Lieferung angenommen, andernfalls zurückgewiesen. Diese Vorgehensweise wird als Einfach– oder $(n-c)$–Stichprobenplan bezeichnet, wobei n und c noch genauer zu bestimmen sind.

Wichtige Begriffe der Testtheorie

Wenn man im Urnenmodell in der Fassung ohne Zurücklegen (siehe 2.3) die Anzahl der Kugeln N mit der Anzahl der angelieferten Stücke und die Anzahl der schwarzen Kugeln R mit der Anzahl der fehlerhaften Stücke identifiziert, stellt man fest, dass X einer hypergeometrischen Verteilung mit den Parametern N, R und n genügt. Dabei bezeichnet $p = R/N$ den wahren, aber unbekannten Ausschussanteil. Die Hypothesen über die zugrundeliegende Verteilung lauten damit:

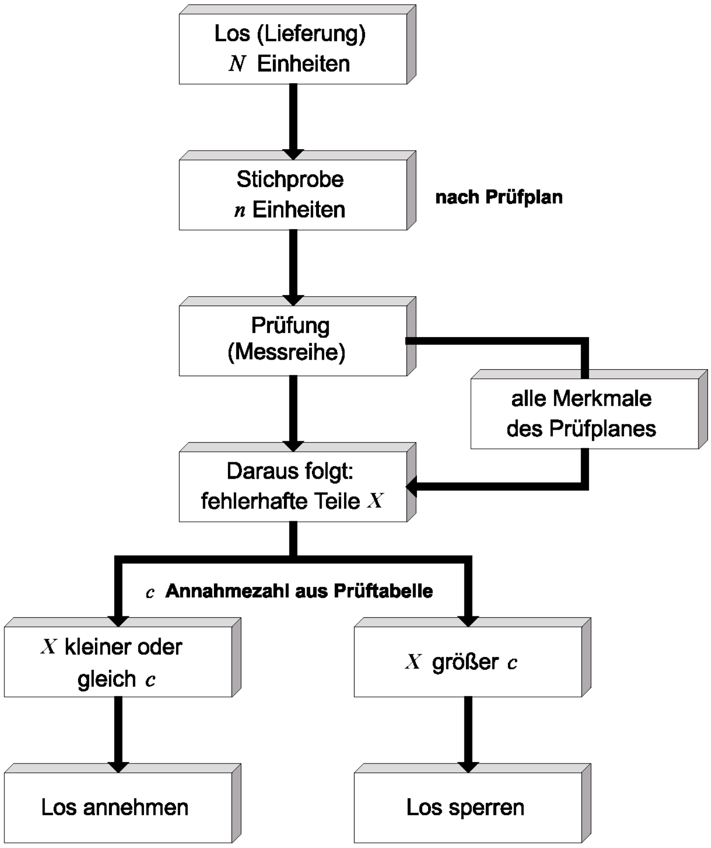


Abbildung 4.2: Schema des einfachen Stichprobenplans

$H_0 : p \leq p_0$ Der Ausschussanteil liegt unterhalb einer kritischen Grenze p_0, d.h. die Lieferung erfüllt die geforderten Qualitätsmerkmale (sogenannte Null–Hypothese).

$H_1 : p > p_0$ Der Ausschussanteil liegt oberhalb einer kritischen Grenze p_0, d.h. die Lieferung erfüllt die geforderten Qualitätsmerkmale nicht (sogenannte Alternativ–Hypothese).

Je kleiner der Wert der Testgröße, umso deutlicher spricht er für H_0 und gegen H_1. Diese Vorgehensweise birgt offensichtlich zwei Risiken:
Der Lieferant hat das Risiko α, dass er aufgrund eines zufällig schlechten Stichprobenergebnisses die Lieferung zurückerhält, obgleich sie eigentlich den vereinbarten Qualitätsanforderungen genügt.
Der Abnehmer hat das Risiko β, dass er aufgrund eines zufällig guten Stichprobenergebnisses die Lieferung akzeptiert, obgleich sie den vereinbarten Qualitätsanforderungen nicht genügt.
Im Hinblick auf die wirtschaftlichen Auswirkungen von Fehlentscheidungen bei der Qualitätsprüfung sollten das Lieferanten- und Abnehmerrisiko möglichst klein sein.
Zur Beurteilung eines Tests stehen verschiedene Kriterien zur Verfügung. Die Funktion

$$G(p) := P(X \in \{c+1, \ldots, n\}) = P(X > c)$$

wird als Gütefunktion bezeichnet. Diese stellt einen Zusammenhang zwischen der Ablehnwahrscheinlichkeit für H_0 und dem unbekannten Parameter p her. Der Begrenzung des Lieferantenrisikos dient die Forderung

$$G(p) = P(X > c) < \alpha \quad \text{für alle } p \leq p_0. \tag{4.1}$$

Ein Test, der diese Bedingung erfüllt, nennt man Test zum Signifikanzniveau α. Üblicherweise wählt man $\alpha = 0.1$, 0.05 oder 0.01.
Die Beurteilung des Konsumentenrisikos geschieht mit Hilfe der Funktion

$$L(p) := P(X \in \{0, 1, \dots, c\}) = P(X \leq c),$$

die als Operationscharakteristik bezeichnet wird. Der Zusammenhang zwischen $G(p)$ und $L(p)$ ist

$$G(p) = 1 - L(p).$$

Damit lässt sich die Bedingung (4.1) auch in der Form

$$1 - L(p) < \alpha \quad \text{für alle } p \leq p_0$$

bzw.

$$L(p) > 1 - \alpha \quad \text{für alle } p \leq p_0$$

schreiben.

4.2 Konstruktion von (n-c)-Stichprobenplänen

Wie bereits erwähnt, tragen Lieferant und Abnehmer bei einer Stichprobenprüfung unterschiedliche Risiken. Sie verfolgen deshalb auch unterschiedliche Ziele:

Zielsetzung des Lieferanten
Wenn der (unbekannte) Ausschussanteil p der Lieferung nicht größer als eine Gutgrenze (AQL; Acceptance Quality Limit) $p_{1-\alpha}$ ist, soll die Lieferung mit einer möglichst großen Wahrscheinlichkeit, nämlich $1-\alpha$, angenommen werden. α beschreibt das Risiko des Lieferanten, dass die Lieferung zurückgewiesen wird, obgleich sie den vereinbarten Qualitätsanforderungen genügt. Lehnt man die Hypothese H_0 („Lieferung genügt den vereinbarten Qualitätsanforderungen") zu Unrecht ab, spricht man von einem Fehler 1. Art. Üblicherweise verwendet man $\alpha = 0.1$, $\alpha = 0.05$ oder $\alpha = 0.01$.

Zielsetzung des Abnehmers
Wenn der Ausschussanteil p größer oder gleich einer Schlechtgrenze (LQ; Limiting Quality) p_β, $p_\beta > p_{1-\alpha}$, ist, soll die Lieferung mit einer möglichst großen Wahrscheinlichkeit $1 - \beta$ abgelehnt werden. β ist folglich das Risiko des Abnehmers, dass er eine unzureichende Lieferung akzeptiert. Nimmt man die Hypothese H_0 („Lieferung genügt den vereinbarten Qualitätsanforderungen") zu Unrecht an, spricht man von einem Fehler 2. Art. Üblicherweise verwendet man $\beta = 0.1$, $\beta = 0.05$ oder $\beta = 0.01$.

Gemeinsame Zielsetzung
Da jede Kontrolle mit Zeit und Geld verbunden ist, haben beide Vertragsparteien ein Interesse daran, dass der Prüfaufwand n möglichst gering ist.
Die Wahrscheinlichkeit, dass ein angeliefertes Los akzeptiert wird, ist bei hypergeometrisch verteiltem X

$$L = L_{N,n,c}(p) = P(X \leq c) = \sum_{k=0}^{c} \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}}, \tag{4.2}$$

wobei $p = R/N$ den wahren (aber unbekannten) Ausschussanteil in der Lieferung bezeichnet. Die Annahmewahrscheinlichkeit $L_{N,n,c}(p)$ als Funktion von p wird Operationscharakteristik des $(n-c)$–Stichprobenplans genannt.

Es bleibt die Frage, wie die Werte n und c eines Stichprobenplanes bestimmt werden. Gibt man die Werte α, β, $p_{1-\alpha}$ und p_β vor, so besteht die Idee darin, die Zahlen n und c so zu wählen, dass die Operationscharakteristik durch die beiden Punkte $(p_{1-\alpha}, 1-\alpha)$ und (p_β, β) verläuft. Dadurch werden sowohl die Zielsetzung des Lieferanten als auch die des Abnehmers und, bei der Wahl eines möglichst kleinen n, auch die gemeinsame Zielsetzung erfüllt. Da es aber in Abhängigkeit von n und c nur endlich viele Operationscharakteristiken gibt, ist dieses Problem in seiner strengen Form im allgemeinen nicht lösbar, weshalb man die folgende Abschwächung wählt:
Bestimme n und c, so dass gilt:

$$L_{N,n,c}(p_{1-\alpha}) \geq 1-\alpha, \quad L_{N,n,c}(p_\beta) \leq \beta, \quad n \to \text{Min.} \tag{4.3}$$

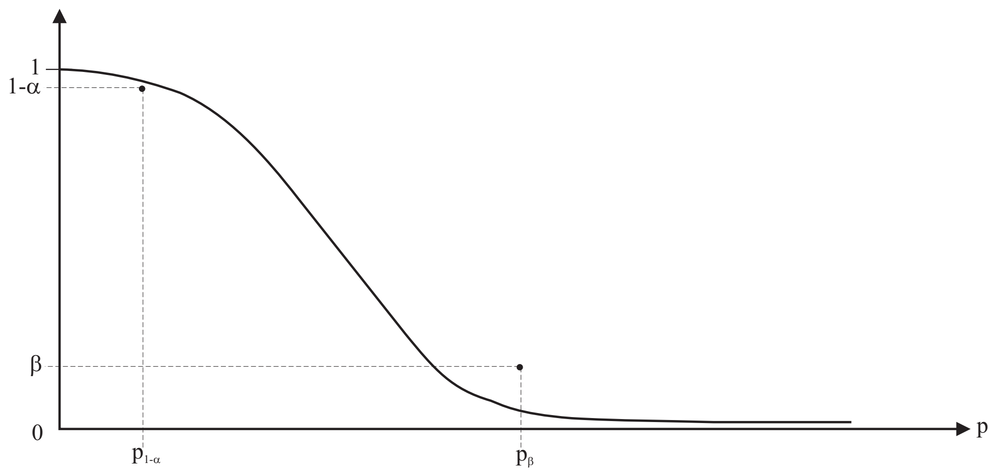


Abbildung 4.3: Beispiel einer Operationscharakteristik

Im Folgenden werden einige Verfahren zur Bestimmung von n und c vorgestellt:

Das Verfahren von Günther

Das **Verfahren von Günther** beruht auf der Feststellung, dass sich die Operationscharakteristik für größer werdende n immer stärker an die x–Achse anschmiegt und für größer werdende c angehoben wird, d.h.

$$\begin{cases} L_{N,n+1,c}(p) \leq L_{N,n,c}(p), \ c \text{ fest}, \ n = c, c+1, \dots, N-1 \\ L_{N,N,c}(p) = \begin{cases} 1 & \text{für } c \geq Np \\ 0 & \text{für } c < Np \end{cases} \\ L_{N,n,c+1}(p) \geq L_{N,n,c}(p), \ n \text{ fest}, \ c = 0, 1, \dots, n-1 \\ L_{N,c,c}(p) = 1. \end{cases}$$

Beweis:

(i) Wir unterscheiden zwei Fälle:

Bei der Prüfung der ersten n Teile der Lieferung mögen sich bereits c als defekt erwiesen haben. Ist das $(n+1)$-te Element in Ordnung, so wird die Lieferung angenommen. Ist es defekt, so wird die Lieferung abgelehnt.

Waren unter den ersten n Teilen weniger als c defekt, so wird die Lieferung auf jeden Fall angenommen, unabhängig davon ob das $(n+1)$-te defekt ist oder nicht. (Da in diesem Fall die Gesamtzahl der defekten Teile höchsten c betragen kann.)

Durch die Erhöhung des Losumfangs kann folglich die Annahmewahrscheinlichkeit nur gleich bleiben oder abnehmen.

(ii) Für $n = N$ gilt:

$$L_{N,N,c}(p) = \sum_{k=0}^{c} \frac{\binom{R}{k}\binom{N-R}{N-k}}{\binom{N}{N}} = \sum_{k=0}^{c} \binom{R}{k}\binom{N-R}{N-k}.$$

Für $k > R$ ist $\binom{R}{k} = 0$ und für $k < R$ ist $N-k > N-R$ bzw. es ist $\binom{N-R}{N-k} = 0$. Damit kann nur der Summand mit $k = R$ ungleich 0 sein:

$$L_{N,N,c}(p) = \sum_{k=R}^{\min\{R,c\}} \binom{R}{k}\binom{N-R}{N-k}.$$

Ist $c < R := Np$, so ist die Summe leer und somit $= 0$, andernfalls ist

$$L_{N,N,c}(p) = \sum_{k=R}^{\min\{R,c\}} \underbrace{\binom{R}{k}}_{=1}\underbrace{\binom{N-R}{N-k}}_{=1} = 1.$$

(iii) Da in der Reihe

$$L_{N,n,c}(p) := \sum_{k=0}^{c} \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}}$$

alle Summanden ≥ 0 sind, kann der Wert der Summe nicht kleiner werden kann, wenn ein Summand hinzugefügt wird. Somit muss also $L_{N,n,c+1}(p) \geq L_{N,n,c}(p)$ gelten.

■

Praktische Vorgehensweise zur Bestimmung von n und c:
Indem man mit $c := 0$ anfängt, versucht man zunächst durch Vergrößern von n die Bedingung $L_{N,n,0}(p_\beta) < \beta$ für alle $p \geq p_\beta$ zu befriedigen. Hat man erstmals ein n gefunden mit $L_{N,n,0}(p_\beta) < \beta$, überprüft man die Bedingung $L_{N,n,0}(p_{1-\alpha}) \geq 1 - \alpha$. Der Prozess bricht ab, wenn diese Bedingung erfüllt ist. Andernfalls setzt man $c := 1$ und wiederholt die Prozedur, bis man ein n findet mit $L_{N,n,1}(p_\beta) \leq \beta$. Ist zugleich $L_{N,n,1}(p_{1-\alpha}) \geq 1 - \alpha$, endet das Verfahren. Andernfalls setzt man den Algorithmus mit $c := 2$ in entsprechender Form fort.

4.1 Beispiel:
Es seien $N := 100$, $AQL{:=}\ 0.01$, $LQ{:=}\ 0.15$ und $\alpha := \beta := 0.1$. Dann ergibt sich nach dem Algorithmus von Günther:

$$\begin{array}{llll}
c := 0 & n := 1 & L(p_\beta) = 0.85 & \\
 & n := 2 & L(p_\beta) = 0.721 & \\
 & \vdots & & \\
 & n := 13 & L(p_\beta) = 0.1039 & \\
 & n := 14 & L(p_\beta) = 0.0860 < \beta = 0.1 & L(p_{1-\alpha}) = 0.86 < 1 - \alpha = 0.9 \\
c := 1 & n := 1 & L(p_\beta) = 1 & \\
 & \vdots & & \\
 & n := 22 & L(p_\beta) = 0.1061 & \\
 & n := 23 & L(p_\beta) = 0.0902 < \beta = 0.1 & L(p_{1-\alpha}) = 1 > 1 - \alpha = 0.9
\end{array}$$

Der gesuchte Stichprobenplan ist von der Form $(23 - 1)$. (Siehe auch Mathematica-Notebook und PowerPoint-Präsentation zu diesem Beispiel.)

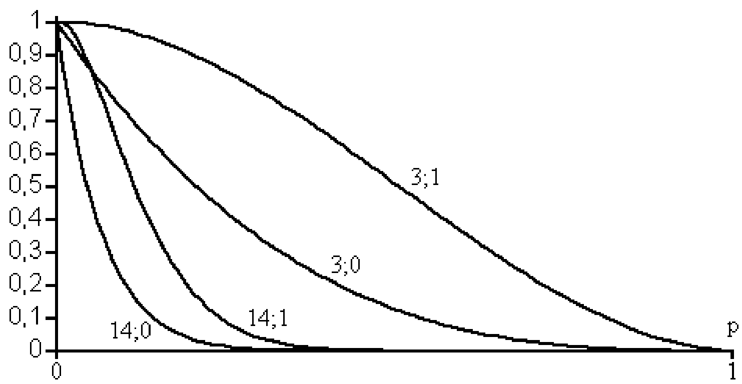


Abbildung 4.4: Beispiel einer Operationscharakteristik $L_{N,n;c}(p)$ für $N = 100$ und verschiedene Werte von $(n; c)$.

Zur Vereinfachung der Berechnung von $L_{N,n,c}(p)$ können auch die Approximationen durch die Binomial– und Poisson–Verteilung verwendet werden. Bezeichnet

$$L_{n,c}(p) := \sum_{m=0}^{c} \binom{n}{m} p^m (1-p)^{n-m}, \quad 0 \leq p \leq 1,$$

die Operationscharakteristik bzgl. der Binomialverteilung mit den Parametern n und p, so gilt

$$\lim_{N\to\infty} L_{N,n,c}(p) = L_{n,c}(p).$$

Diese Näherung ist für $\frac{n}{N} \leq 0.1$ hinreichend genau.
Die Operationscharakteristik $L_{n,c}(p)$ kann ihrerseits durch die Operationscharakteristik bzgl. der Poissonverteilung

$$L^*_{n,c}(p) := \sum_{m=0}^{c} \frac{(np)^m}{m!} e^{-np}$$

angenähert werden, falls $n \geq 100$ und $p < 0.05$.

Das Verfahren von Günther kann nicht nur mit der hypergeometrischen Verteilung sondern in derselben Weise auch mit der Binomial– und der Poissonverteilung als Modellverteilung durchgeführt werden.

Die χ^2-Methode

Will man n und c numerisch bestimmen, bietet sich das Verfahren von P. Peach und S.B. Littauer an, das auf der Poisson–Verteilung als Modellverteilung basiert und den folgenden Zusammenhang ausnutzt:

4.2 Satz:
Für $c = 0, 1, 2, \ldots$ und alle $\lambda \geq 0$ gilt:

$$\sum_{k=0}^{c} p_\lambda(k) \;=\; \sum_{k=0}^{c} \frac{\lambda^k}{k!} e^{-\lambda} \;=\; 1 - G\big(2\lambda; 2(c+1)\big) \;=\; 1 - \int_0^{2\lambda} \frac{1}{2^{c+1} c!} y^c e^{-y/2}\, dy, \qquad (4.4)$$

wobei G die sogenannte χ^2–Verteilung mit $2(c+1)$ Freiheitsgraden bedeutet.

Beweis:
Es seien

$$f(\lambda) \;:=\; \sum_{k=0}^{c} \frac{\lambda^k}{k!} e^{-\lambda} \qquad \text{und} \qquad g(\lambda) \;:=\; 1 - \int_0^{2\lambda} \frac{1}{2^{c+1} \cdot c!} y^c e^{-y/2} dy, \qquad \lambda \geq 0.$$

Im Folgenden wird gezeigt, dass $f(0) = g(0)$ und $f'(\lambda) = g'(\lambda)$ für alle $\lambda \geq 0$ gilt.

(i) Offensichtlich ist $f(0) = 1 = g(0)$.

(ii) Es gilt

$$\begin{aligned} f'(\lambda) \;&=\; \left(e^{-\lambda} \sum_{k=0}^{c} \frac{\lambda^k}{k!}\right)' \;=\; e^{-\lambda} \left[\sum_{k=1}^{c} \frac{k\lambda^{k-1}}{k!} - \sum_{k=0}^{c} \frac{\lambda^k}{k!}\right] \\ &=\; e^{-\lambda} \left[\sum_{k=1}^{c} \frac{\lambda^{k-1}}{(k-1)!} - \sum_{k=1}^{c+1} \frac{\lambda^{k-1}}{(k-1)!}\right] \;=\; -\, e^{-\lambda} \frac{\lambda^c}{c!} \end{aligned}$$

und

$$\begin{aligned} g'(\lambda) &= -\frac{1}{2^{c+1}\cdot c!}\cdot\frac{d}{d\lambda}\int_0^{2\lambda} y^c e^{-y/2}\,dy \;=\; -\frac{1}{2^{c+1}\cdot c!}\cdot\frac{d}{d\lambda}\int_0^{\lambda} 2^c x^c e^{-x}\cdot 2\,dx \\ &= -\frac{1}{c!}\cdot\frac{d}{d\lambda}\int_0^{\lambda} x^c e^{-x}\,dx \end{aligned}$$

Es sei $F(x) := \int x^c e^{-x}\,dx$. Damit folgt weiter:

$$g'(\lambda) \;=\; -\frac{1}{c!}\cdot\frac{d}{d\lambda}(F(\lambda)-F(0)) \;=\; -\frac{1}{c!}\lambda^c e^{-\lambda},$$

was zu zeigen war. ■

Für die Annahmewahrscheinlichkeit eines $(n-c)$–Stichprobenplans unter Zugrundelegung der Poisson–Verteilung gilt:

$$\begin{aligned} L^*_{n,c}(p) &:= \sum_{k=0}^{c}\frac{(np)^k}{k!}e^{-np} \;=\; 1-\int_0^{2np}\frac{1}{2^{c+1}c!}y^c\cdot e^{-y/2}\,dy \\ &= 1-G(2np;2(c+1)), \qquad np\geq 0. \end{aligned} \tag{4.5}$$

Setzt man diese Beziehung in die beiden Abschätzungen der Operationscharakteristik (4.3) ein, erhält man

$$\begin{aligned} L^*_{n,c}(p_{1-\alpha})\geq 1-\alpha \quad &\Longleftrightarrow\quad 1-G(2np_{1-\alpha};2(c+1))\geq 1-\alpha \\ &\Longleftrightarrow\quad G(2np_{1-\alpha};2(c+1))\leq \alpha \\ &\Longleftrightarrow\quad G^{-1}(\alpha;2(c+1))\geq 2np_{1-\alpha} \\ &\Longleftrightarrow\quad \frac{G^{-1}(\alpha;2(c+1))}{2p_{1-\alpha}}\geq n, \end{aligned}$$

bzw.

$$\begin{aligned} L^*_{n,c}(p_{\beta})\leq \beta \quad &\Longleftrightarrow\quad 1-G(2np_{\beta};2(c+1))\leq \beta \\ &\Longleftrightarrow\quad G(2np_{\beta};2(c+1))\geq 1-\beta \\ &\Longleftrightarrow\quad G^{-1}(1-\beta;2(c+1))\leq 2np_{\beta} \\ &\Longleftrightarrow\quad \frac{G^{-1}(1-\beta;2(c+1))}{2p_{\beta}}\leq n. \end{aligned}$$

Fasst man diese Ungleichungen zusammen, erhält man das folgende abschließende Resultat.

4.3 Satz:
Ein $(n-c)$ Stichprobenplan erfüllt die Bedingungen

$$L^*_{n,c}(p_{1-\alpha})\geq 1-\alpha \qquad \textit{und} \qquad L^*_{n,c}(p_{\beta})\leq \beta$$

genau dann, wenn gilt:

$$\frac{G^{-1}(1-\beta;2(c+1))}{2p_{\beta}}\leq n\leq \frac{G^{-1}(\alpha;2(c+1))}{2p_{1-\alpha}}. \tag{4.6}$$

Indem man für c nacheinander die natürlichen Zahlen $c = 0, 1, 2, \ldots$ einsetzt, ermittelt man das kleinste c, zu dem es eine natürliche Zahl $n > c$ gibt, die der Bedingung (4.6) genügt.

4.4 Beispiel:
Es seien $N := 6000$, $p_{1-\alpha} := 0.015$, $\alpha := 0.05$, $1 - \alpha := 0.95$, $p_\beta := 0.12$ und $\beta := 0.025$, $1 - \beta := 0.975$. In diesem Fall tritt an die Stelle von (4.6) folgende Ungleichung:

$$\frac{G^{-1}(0.975; 2(c+1))}{0.24} \leq n \leq \frac{G^{-1}(0.05; 2(c+1))}{0.03}.$$

Mit Hilfe der Tabelle für die χ^2–Verteilung (siehe Anhang A) findet man

c	$\frac{G^{-1}(0.975;2(c+1))}{0.24}$		$\frac{G^{-1}(0.05;2(c+1))}{0.03}$
0	$7.38/0.24 = 30.75$	$>$	$3.33 = 0.1/0.03$
1	$11.14/0.24 = 46.42$	$>$	$23.67 = 0.71/0.03$
2	$14.45/0.24 = 60.21$	$>$	$54.67 = 1.64/0.03$
3	$17.53/0.24 = 73.04$	$<$	$91 = 2.73/0.03$

Hieraus folgt $c = 3$ und $n = 74$. (Siehe auch PowerPoint-Präsentation zu diesem Beispiel.)

Vorgabe des Indifferenzpunktes und der Steilheit

Aus der Operationscharakteristik wird klar, dass ein $(n - c)$–Stichprobenplan umso besser zwischen einer guten und einer schlechten Lieferung trennt, je steiler die Annahmekennlinie verläuft. Als Maß für die Trennschärfe verwendet man den Quotienten

$$\text{Trennschärfe} \;:=\; \frac{\text{Gutgrenze}}{\text{Schlechtgrenze}} \;=\; \frac{\text{AQL}}{\text{LQ}}, \tag{4.7}$$

der im Idealfall in der Nähe von 1 liegen würde, was aber nur durch einen sehr hohen Prüfaufwand zu erreichen wäre.
Das Stichprobensystem der Firma Philips orientiert sich an der Steilheit der Operationscharakteristik. Die Steilheit h_0 wird dabei im sogenannten Indifferenzpunkt $p_{0.5}$ gemessen, der durch die Gleichung

$$L(p_{0.5}) \;=\; 0.5 \tag{4.8}$$

bestimmt ist. Als Steilheit wird

$$h_0 \;:=\; -\frac{p}{L(p)} \cdot \frac{dL(p)}{dp}\bigg|_{p=p_{0.5}} \;=\; -\frac{p_{0.5}}{0.5} \cdot \frac{dL(p)}{dp}\bigg|_{p=p_{0.5}}$$

definiert. Wählt man als Modellverteilung die Poissonverteilung, so ergibt sich

$$h_0 \;=\; -2p_{0.5}\frac{dL^*_{n,c}(p)}{dp}\bigg|_{p=p_{0.5}} \;=\; \frac{2(np_{0.5})^{c+1}}{c!}e^{-np_{0.5}}.$$

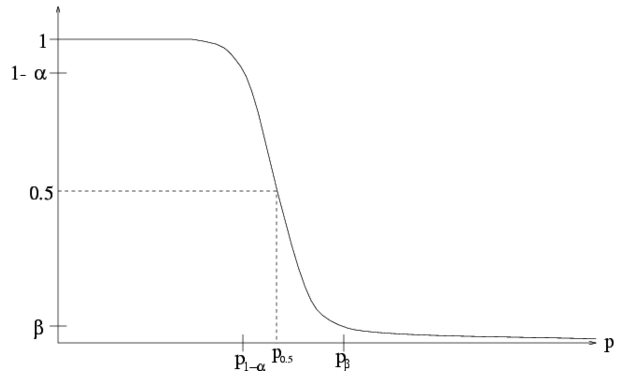


Abbildung 4.5: Indifferenzpunkt und Steilheit

Andererseits gilt aufgrund früherer Überlegungen:

$$\begin{aligned} & L^*_{n,c}(p_{0.5}) = 1 - G(2np_{0.5}; 2(c+1)) = 0.5 \\ \iff & G(2np_{0.5}; 2(c+1)) = 0.5 \\ \iff & G^{-1}(0.5; 2(c+1)) = 2np_{0.5} \\ \iff & \frac{1}{2}G^{-1}(0.5; 2(c+1)) = np_{0.5}. \end{aligned}$$

Beim Philips Stichprobensystem wird nun folgendermaßen vorgegangen: Gegeben sind $\overline{h_0}$ und $p_{0.5}$. Es sind n und c so zu bestimmen, dass

$$h_0 = \frac{2(np_{0.5})^{c+1}}{c!} e^{-np_{0.5}} \geq \overline{h_0} \qquad \text{und} \qquad \frac{1}{2}G^{-1}(0.5; 2(c+1)) = np_{0.5}$$

gilt.
Um diese beiden Bedingungen zu erfüllen, wird genau wie beim Algorithmus von Günther vorgegangen. Es wird zunächst $c := 0$, $c := 1$, ... gesetzt, bis man zum ersten Mal

$$h_0 = \frac{2(n \cdot p_{0.5})^{c+1}}{c!} e^{-n \cdot p_{0.5}} \geq \overline{h_0}$$

erzielt hat. Der zugehörige Stichprobenumfang n wird mit Hilfe der Gleichung

$$n := \left\lceil \frac{G^{-1}(0.5; 2(c+1))}{2 \cdot p_{0.5}} \right\rceil$$

ermittelt, wobei $\lceil x \rceil$ die kleinste ganze Zahl größer oder gleich x bedeutet.

4.5 Beispiel:
Es seien $N := 1500$, $\overline{h_0} := 1.5$ und $p_{0.5} := 0.03$.

c	Tabelle der χ^2–Verteilung $G^{-1}(0.5;2(c+1))$	$x=\frac{G^{-1}(0.5;2(c+1))}{2}$	$h_0=\frac{2x^{c+1}e^{-x}}{c!}$	$\geq\overline{h_0}=1.5$	$n=\left\lceil\frac{G^{-1}(0.5;2(c+1))}{2\cdot p_{0.5}}\right\rceil$
0	1.39	0.695	0.694	nein	
1	3.36	1.68	1.052	nein	
2	5.35	2.675	1.319	nein	
3	7.34	3.67	1.541	ja	$\lceil 122.33\rceil=123$

Der gesuchte Stichprobenplan ist folglich von der Form $(n - c) = (123 - 3)$. (Siehe auch PowerPoint-Präsentation zu diesem Beispiel.)

4.6 Bemerkung:
Die Werte n und c eines Stichprobenplanes können bei gegebenem Indifferenzenpunkt $p_{0.5}$ und gegebener Steilheit h_0 auch durch Näherungsformeln bestimmt werden. Aus der Tabelle der χ^2-Verteilung lässt sich die Approximation $G^{-1}(0.5; 2(c+1)) \approx 2(c+1) - 0.66$ ableiten. Daraus ergeben sich folgende Abschätzungen: $c = \pi \cdot 2{h_0}^2 - 1$ und $n = c + 0.67 \cdot p_{0.5}$.

4.3 Maximaler mittlerer Durchschlupf

Trotz sorgfältiger Kontrolle kann bei der Anwendung eines $(n - c)$–Stichprobenplans nicht verhindert werden, dass defekte Einheiten die Prüfung passieren. Die Tests sind lediglich darauf ausgerichtet, dass eine Lieferung mit einer Qualitätslage $p \leq p_{1-\alpha}$ mit großer Wahrscheinlichkeit angenommen und eine Lieferung mit einer Qualitätslage $p \geq p_\beta$ mit großer Wahrscheinlichkeit abgelehnt wird. Der Anteil durchschlüpfender defekter Einheiten Y ist deshalb ein weiteres Beurteilungskriterium für einen $(n - c)$–Stichprobenplan.
Es wird zunächst angenommen, dass der Lieferant Lose mit konstanter Qualitätslage p anliefert. Die Darstellung der Zufallsvariablen Y hängt davon ab, wie mit den Einheiten der Stichprobe und, bei Ablehnung, auch mit den im Los verbleibenden Einheiten verfahren wird. Unter der Annahme, dass die Teile in der Stichprobe grundsätzlich nicht weiterverwendet werden und bei Ablehnung eine Vollkontrolle erfolgt, bei der sämtliche defekten Teile durch gute ersetzt werden, erhält man:

$$Y = \begin{cases} p, & \text{falls } X \leq c \\ 0, & \text{falls } X > c \end{cases},$$

wobei X wieder die Anzahl defekter Einheiten in der Stichprobe bedeutet. Der Erwartungswert von Y ist

$$\mathrm{E}[Y] = p \cdot P(X \leq c) + 0 \cdot P(X > c) = p \cdot L(p)$$

und wird <u>mittlerer Durchschlupf</u> (englisch: Average Outgoing Quality (AOQ)) genannt.
Für den Konsumenten ist der <u>maximale mittlere Durchschlupf</u> (Average Outgoing Quality Limit (AOQL)) von besonderem Interesse:

$$\text{AOQL} := \max_{0 \leq p \leq 1} \; p \cdot L(p).$$

Um den AOQL zu berechnen, wird wieder die Approximation durch die Poissonverteilung benutzt:

$$\text{AOQL} = \max_{p \geq 0} \; p \cdot L^*_{n,c}(p) = \max_{p \geq 0} \; p \cdot \sum_{k=0}^{c} \frac{(n \cdot p)^k}{k!} e^{-n \cdot p}.$$

4.7 Satz:
Es gibt genau ein $\overline{p} \in \left(0, \frac{c+2}{n}\right)$ *mit*

$$AOQL = \max_{p \geq 0} \; p \cdot L^*_{n,c}(p) = \overline{p} \cdot L^*_{n,c}(\overline{p}).$$

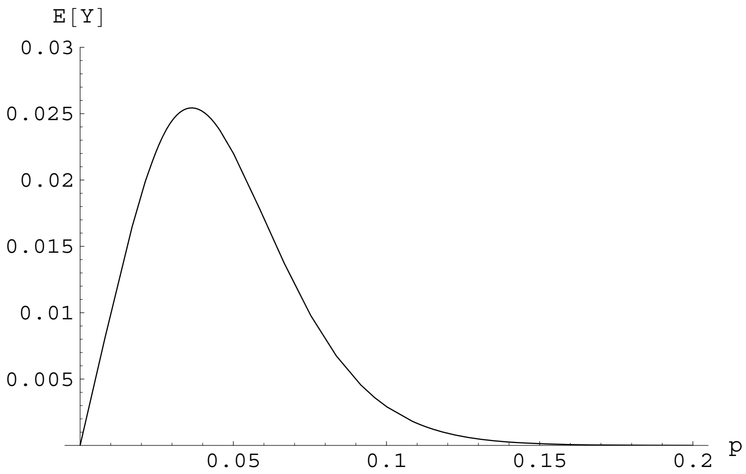


Abbildung 4.6: Mittlerer Durchschlupf E$[Y]$ bei $n := 100$ und $c := 4$

Dabei ist $\overline{p}$ ist eindeutig bestimmt durch die Gleichung

$$\sum_{k=0}^{c} \frac{(n \cdot \overline{p})^k}{k!} = \frac{(n \cdot \overline{p})^{c+1}}{c!}$$

und es gilt

$$AOQL = \frac{(n \cdot \overline{p})^{c+2}}{c! \cdot n} e^{-n\overline{p}}.$$

Beweis:
Zu zeigen sind:

- Existenz der Maximalstelle $\overline{p}$
- Bestimmung des Maximums durch
  (i) $\frac{d}{dp} p \cdot L^*_{n,c}(p) = 0$
  (ii) $\frac{d^2}{dp^2} p \cdot L^*_{n,c}(p) < 0$

Die Existenz einer Maximalstelle $\overline{p}$ von $p \cdot L^*_{n,c}(p)$ ist aufgrund der Beziehungen

$$p \cdot L^*_{n,c}(p) \begin{cases} = 0 \ , & \text{für } p = 0 \\ > 0 \ , & \text{für } p > 0 \\ \to 0 \ , & \text{für } p \to \infty \end{cases}$$

gesichert.

Zu (i): Es gilt

$$\frac{d}{dp} p \cdot L^*_{n,c}(p) = L^*_{n,c} + p \cdot \frac{d}{dp} L^*_{n,c}(p).$$

Mit $L^*_{n,c}(p) = f(np)$ wie im Beweis von Satz 4.2 und der dort berechneten Ableitung von f

$$\frac{d}{dp} L^*_{n,c}(p) = \frac{d}{dp} f(np) = n \cdot f'(np) = -n \cdot e^{-np} \frac{(np)^c}{c!}$$

folgt

$$\frac{d}{dp}p \cdot L^*_{n,c}(p) = \left(\sum_{k=0}^{c} \frac{(n \cdot p)^k}{k!} - \frac{(n \cdot p)^{c+1}}{c!}\right) e^{-np}$$

und somit insgesamt

$$\begin{aligned} & \frac{d}{dp}p \cdot L^*_{n,c}(p) = 0 \\ \iff \quad & \sum_{k=0}^{c} \frac{(np)^k}{k!} e^{-np} = e^{-np} \frac{(np)^{c+1}}{c!} \\ \iff \quad & \sum_{k=0}^{c} \frac{(n \cdot p)^k}{k!} = \frac{(n \cdot p)^{c+1}}{c!}. \end{aligned}$$

Zu (ii) Es gilt:

$$\begin{aligned} \frac{d^2}{dp^2} p L^*_{n,c}(p) &= \frac{d}{dp}\left(L^*_{n,c}(p) + p\frac{d}{dp}L^*_{n,c}(p)\right) \\ &= \frac{d}{dp}L^*_{n,c}(p) + \frac{d}{dp}L^*_{n,c}(p) + p \cdot \frac{d^2}{dp^2}L^*_{n,c}(p) \\ &= 2 \cdot \frac{d}{dp}L^*_{n,c}(p) + p \cdot \frac{d^2}{dp^2}L^*_{n,c}(p) \end{aligned}$$

und damit folgt durch Einsetzen

$$\begin{aligned} \frac{d^2}{dp^2} p L^*_{n,c}(p) &= -2e^{-np}\frac{n^{c+1}p^c}{c!} - \frac{pn^{c+1}}{c!} \cdot \frac{d}{dp}e^{-np}p^c \\ &= -2\frac{n^{c+1}p^c}{c!}e^{-np} - \frac{pn^{c+1}}{c!}e^{-np}(-np^c + cp^{c-1}) \\ &= \frac{n^{c+1}p^c}{c!}e^{-np}(np - (c+2)) \begin{cases} < 0 \text{ für } & 0 < p < \frac{c+2}{n} \\ = 0 \text{ für } & p = \frac{c+2}{n} \\ > 0 \text{ für } & p > \frac{c+2}{n}. \end{cases} \end{aligned}$$

Deswegen muss die Maximalstelle $\overline{p}$ im Intervall $\left(0, \frac{c+2}{n}\right)$ liegen.

Unter Verwendung der letzten Identität aus Punkt (i) erhält man nun mit

$$\begin{aligned} \text{AOQL} &= \max_{p \geq 0} \; p \cdot L^*_{n,c}(p) = \overline{p} \cdot L^*_{n,c}(\overline{p}) = \overline{p} \cdot \sum_{k=0}^{c} \frac{(n \cdot \overline{p})^k}{k!} \cdot e^{-n\overline{p}} \\ &= \overline{p} \cdot \frac{(n \cdot \overline{p})^{c+1}}{c!} \cdot e^{-n\overline{p}} = \frac{(n \cdot \overline{p})^{c+2}}{c! \cdot n} \cdot e^{-n\overline{p}} \end{aligned}$$

die behauptete Identität für den maximalen mittleren Durchschlupf AOQL. ■

An den letzten beiden Ausdrücken erkennt man, dass die Größen $x = n \cdot \overline{p}$ und $x' = n \cdot \text{AOQL}$ nur noch von c abhängen und deshalb leicht tabelliert werden können.

c	$n \cdot \overline{p}$	$n \cdot \text{AOQL}$
0	1.000	0.3679
1	1.618	0.840
2	2.270	1.371
3	2.945	1.942
4	3.640	2.544
5	4.349	3.168
6	5.071	3.812
7	5.804	4.472
8	6.546	5.146
9	7.297	5.831
10	8.055	6.528

Mit Hilfe dieser Tabelle lassen sich der maximale mittlere Durchschlupf und die dazugehörige Qualitätslage $\overline{p}$ für beliebige $(n-c)$–Stichprobenpläne leicht berechnen.

4.8 Beispiel:
Bei einem $(n-c)$–Stichprobenplan mit $n := 100$ und $c := 3$ ist der maximale mittlere Durchschlupf

$$\text{AOQL} = \frac{1.942}{100} = 0.01942.$$

Die zugehörige Qualitätslage ist

$$\overline{p} = \frac{2.945}{100} = 0.02945.$$

4.4 Mittlerer Prüfaufwand

Mit jeder Stichprobenprüfung ist ein gewisser technischer Aufwand verbunden, der Kosten verursacht. Bei der Berechnung des mittleren Prüfaufwands wird wie bei der Berechnung des maximalen mittleren Durchschlupfs vorgegangen und angenommen, dass bei einer Ablehnung eine Vollkontrolle stattfindet, bei der alle defekten Teile durch gute ersetzt werden. In diesem Fall ergibt sich für die Anzahl M der zu prüfenden Stücke

$$M = \begin{cases} n, & \text{falls } X \leq c \\ N, & \text{falls } X > c. \end{cases}$$

Die durchschnittliche Anzahl zu prüfender Einheiten ist deshalb

$$\text{E}[M] = n \cdot P(X \leq c) + N \cdot P(X > c) = n \cdot L(p) + N(1 - L(p)).$$

4.9 Beispiel:
Es seien $n := 150$ und $c := 3$. Die Lose vom Umfang $N := 2000$ werden mit einer mittleren Qualitätslage von 1% angeliefert.
Es ist

$$L^*_{150,3}(0.01) = 0.9344$$

und deshalb

$$\text{E}[M] = 150 \cdot 0.9344 + 2000 \cdot 0.0656 = 271.36.$$

Literatur zu Kapitel 4

Folgende Bücher werden als begleitende Literatur zu diesem Kapitel des Skriptes empfohlen:

- J. Banks:
  Principles of Quality Control,
  John Wiley and Sons, New York, 1989.
  ISBN: 0471635510
- D.C. Montgomery:
  Introduction to Statistical Quality Control,
  2nd edition, John Wiley and Sons, New York, 1991.
  ISBN: 0471656313
- H. Rinne und H.–J. Mittag:
  Statistische Methoden der Qualitätssicherung,
  3. Auflage, Carl Hanser Verlag, München, 1995.
  ISBN: 3446180060
- W. Uhlmann:
  Statistische Qualitätskontrolle,
  2. Auflage, Teubner–Verlag, Stuttgart, 1982.
  ISBN: 3519123061

Kapitel 5

Mengensysteme

In der Einführung wurde bereits festgestellt, dass σ–Algebren den natürlichen Definitionsbereich von Wahrscheinlichkeitsmaßen darstellen. Im Vorgriff auf die beiden Maß–Fortsetzungssätze, die im nächsten Kapitel behandelt werden, erweist es sich als zweckmäßig, als Vorstufen von σ–Algebren auch einfachere Mengensysteme zu untersuchen.

Schlüsselwörter: Semiring, Ring, Algebra, Dynkin–System, σ–Algebra, Erzeugendensystem, Darstellungssatz für Ringe, σ–Algebra der Borelschen Mengen.

5.1 Begriffe und Zusammenhänge

In diesem Abschnitt werden verschiedene Typen von Mengensysteme eingeführt und ihre Eigenschaften untersucht.

5.1 Definition (vereinigungsstabil, durchschnittsstabil):
Es seien Ω eine nichtleere Menge und $\mathfrak{M} \subseteq \mathfrak{P}(\Omega)$.

- $\mathfrak{M}$ *heißt* <u>*vereinigungsstabil*</u> *— $\cup$–stabil —, wenn mit $A, B \in \mathfrak{M}$ auch $A \cup B \in \mathfrak{M}$ ist.*
- $\mathfrak{M}$ *heißt* <u>*durchschnittsstabil*</u> *oder einfach* <u>*schnittstabil*</u> *— $\cap$–stabil —, wenn mit $A, B \in \mathfrak{M}$ auch $A \cap B \in \mathfrak{M}$ ist.*

5.2 Definition (Semiring):
Ein System S von Teilmengen von Ω heißt <u>*Semiring*</u> *über Ω, wenn es folgende Eigenschaften besitzt:*

(i) $\varnothing \in S$,

(ii) S *ist* $\cap$*–stabil,*

(iii) $A, B \in S \Longrightarrow$ *es existieren* $n \in \mathbb{N}$ *und* $C_1, \ldots, C_n \in S$ *mit* $C_i \cap C_j = \varnothing$ *für* $i \neq j$, *so dass* $A \backslash B = \bigcup_{i=1}^{n} C_i$ *ist.*

Es sei jetzt speziell $\Omega := \mathbb{R}^n$, $n \in \mathbb{N}$, gewählt. Bezeichnen $a := (a_1, \ldots, a_n)$ und $b := (b_1, \ldots, b_n)$ zwei Punkte des Euklidischen Raumes $\mathbb{R}^n$ mit $a \leq b$, d.h. $a_i \leq b_i$ für $i = 1, \ldots, n$, dann versteht man unter einem linksseitig offenen und rechtsseitig abgeschlossenen Intervall die folgende Punktmenge:

$$(a, b]_{(n)} := \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_i < x_i \leq b_i;\ i = 1, \ldots, n\}.$$

Im Folgenden soll gezeigt werden, dass die Mengen

$$\mathbb{I}^n := \{(a, b]_{(n)} \mid a, b \in \mathbb{R}^n,\ a \leq b\}$$

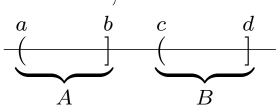
einen Semiring über $\Omega := \mathbb{R}^n$ bilden. Dies wird zunächst im folgenden Beispiel für $n = 1, 2$ anschaulich dargestellt und dann in Satz 5.4 allgemein bewiesen.

5.3 Beispiel:
Das Mengensystem $\mathbb{I}^n$ ist für $n = 1, 2$ ein Semiring über $\Omega := \mathbb{R}^n$. Denn:

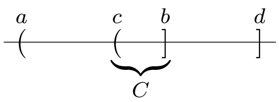
1. Der Fall $n := 1$:

   a) Die leere Menge $\varnothing$ wird durch $(a, a]$ hinzugezogen.

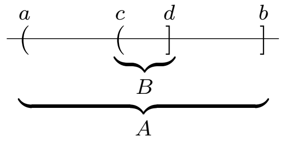
   b) Der Durchschnitt von zwei linksseitig offenen und rechtsseitig abgeschlossenen Intervallen $A := (a, b]$ und $B := (c, d]$ ist entweder leer oder wieder ein solches Intervall, wie die nachfolgenden Abbildungen zeigen:

   $A \cap B = \varnothing \in \mathbb{I}^1$,

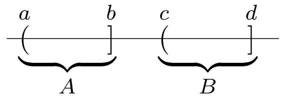
   

   $A \cap B = C = (c, b] \in \mathbb{I}^1$,

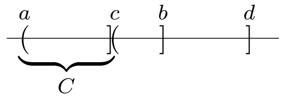


$B \subset A \Longrightarrow A \cap B = B \in \mathbb{I}^1$.

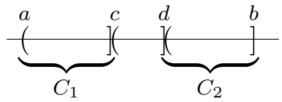
c) Schließlich ist $A \backslash B = A \cap \overline{B}$ gleich A (falls A und B disjunkt sind) oder wieder ein Intervall bzw. Vereinigung von höchstens zwei Intervallen desselben Typs:



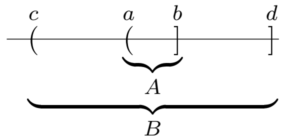
$A \cap B = \varnothing \Longrightarrow A \setminus B = A \cap \overline{B} = A \in \mathbb{I}^1$,



$A \setminus B = A \cap \overline{B} = C = (a, c] \in \mathbb{I}^1$,



$B \subset A \Longrightarrow A \setminus B = A \cap \overline{B} = C_1 \cup C_2$, wobei $C_1, C_2 \in \mathbb{I}^1$ und $C_1 \cap C_2 = \varnothing$,

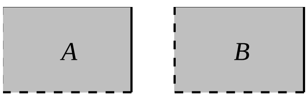


$A \subset B \Longrightarrow A \backslash B = A \cap \overline{B} = \varnothing \in \mathbb{I}^1$. (Siehe auch PowerPoint-Präsentation zu diesem Beispiel.)

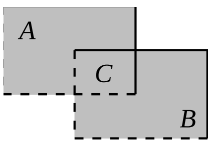
2. Der Fall $n := 2$:

   a) Die leere Menge $\varnothing$ wird durch $(a, a] \times (a, a]$ hinzugezogen.

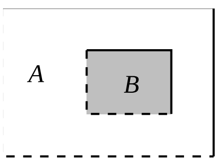
   b) Auch $\mathbb{I}^2$ ist $\cap$–stabil, wie die nachfolgenden Abbildungen zeigen:



$A \cap B = \varnothing \in \mathbb{I}^2$,

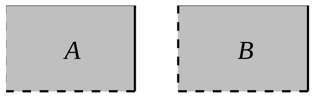


$A \cap B \neq \varnothing \Longrightarrow A \cap B = C \in \mathbb{I}^2$,

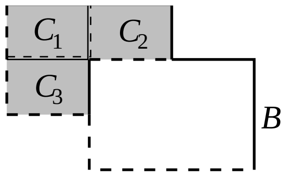


$B \subset A \Longrightarrow A \cap B = B \in \mathbb{I}^2$.

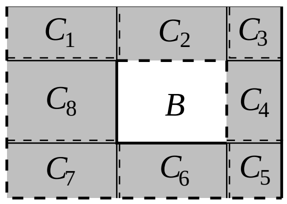
   c) Für alle möglichen Fälle ist $A \setminus B \in \mathbb{I}^2$:



$$A \backslash B = A \cap \overline{B} = A \in \mathbb{I}^2$$



$$A \backslash B = C_1 \cup C_2 \cup C_3 \in \mathbb{I}^2$$



$$A \backslash B = \bigcup_{i=1}^{8} C_i \in \mathbb{I}^2$$

(Siehe auch PowerPoint-Präsentation zu diesem Beispiel.)

5.4 Satz:
Das Mengensystem

$$\mathbb{I}^n := \{(a,b]_{(n)} \mid a,b \in \mathbb{R}^n,\ a \leq b\}$$

ist ein Semiring über $\Omega := \mathbb{R}^n$.

Beweis:
Folgende Eigenschaften sind nachzuweisen:

1. Es gilt $\varnothing \in \mathbb{I}^n$, denn es ist $\varnothing = (a,a]_{(n)} \in \mathbb{I}^n$.

2. Der Schnitt zweier Intervalle aus $\mathbb{I}^n$ ist wieder ein Intervall aus $\mathbb{I}^n$:
   Es seien $(a,b]_{(n)}$ und $(c,d]_{(n)}$ zwei Intervalle aus $\mathbb{I}^n$. Mit $e_i := \max\{a_i, c_i\}$ und $f_i := \min\{b_i, d_i\}$ gilt:

   $$(a,b]_{(n)} \ \cap \ (c,d]_{(n)} = (e,f]_{(n)} \in \mathbb{I}^n.$$

   Dabei ist zu beachten, dass $(e,f]_{(n)}$ leer ist, wenn für ein $1 \leq i \leq n$ gilt $e_i \geq f_i$.

3. Für zwei Elemente $A, B \in \mathbb{I}^n$ mit $A \subseteq B$ lässt sich $A \setminus B$ als endliche Vereinigung von Elementen aus $\mathbb{I}^n$ darstellen:
   Es seien $A := (a,b]_{(n)}$ und $B := (c,d]_{(n)}$ zwei Intervalle aus $\mathbb{I}^n$. Mit $e_i := \max\{a_i, c_i\}$ und $f_i := \min\{b_i, d_i\}$ gilt zunächst wie im 2. Punkt $A \cap B = C := (e,f]_{(n)}$ und damit $A \setminus B = A \setminus C$. Außerdem gilt

   $$a_i \leq e_i \leq f_i \leq b_i \ \forall \ i = 1, \ldots, n,$$

   da nach Wahl der c_i und der d_i gilt: $(c,d]_{(n)} \subseteq (a,b]_{(n)}$. Damit lässt sich allgemein schreiben:

$$
\begin{aligned}
(a,b]_{(n)} \setminus (c,d]_{(n)} &= \left(\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}\right]_{(n)} \setminus \left(\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}, \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix}\right]_{(n)} \\
&= \left(\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \begin{pmatrix} c_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}\right]_{(n)} \cup \left(\begin{pmatrix} d_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}\right]_{(n)} \\
&\quad \cup \left(\begin{pmatrix} c_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix}, \begin{pmatrix} d_1 \\ c_2 \\ b_3 \\ \vdots \\ b_n \end{pmatrix}\right]_{(n)} \cup \left(\begin{pmatrix} c_1 \\ d_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix}, \begin{pmatrix} d_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{pmatrix}\right]_{(n)} \\
&\quad \vdots \\
&\quad \cup \left(\begin{pmatrix} c_1 \\ \vdots \\ c_{n-1} \\ a_n \end{pmatrix}, \begin{pmatrix} d_1 \\ \vdots \\ d_{n-1} \\ c_n \end{pmatrix}\right]_{(n)} \cup \left(\begin{pmatrix} c_1 \\ \vdots \\ c_{n-1} \\ d_n \end{pmatrix}, \begin{pmatrix} d_1 \\ \vdots \\ d_{n-1} \\ b_n \end{pmatrix}\right]_{(n)}
\end{aligned}
$$

Dabei sind die Vereinigungen disjunkt und einige der 2^n Stücke eventuell leer.

■

5.5 Definition (Ring):
Ein System $\mathfrak{R}$ von Teilmengen einer nichtleeren Menge Ω heißt ein Ring über Ω, wenn es die folgenden Eigenschaften besitzt:

(i) $\varnothing \in \mathfrak{R}$,

(ii) $\mathfrak{R}$ *ist* $\cup$*–stabil,*

(iii) $A, B \in \mathfrak{R} \Longrightarrow A \backslash B \in \mathfrak{R}$.

5.6 Definition (Algebra):
Ein System $\mathfrak{A}$ von Teilmengen einer nichtleeren Menge Ω heißt eine Algebra über Ω, wenn gilt:

(i) $\Omega \in \mathfrak{A}$,

(ii) $\mathfrak{A}$ *ist* $\cup$*–stabil,*

(iii) $A \in \mathfrak{A} \Longrightarrow \overline{A} \in \mathfrak{A}$.

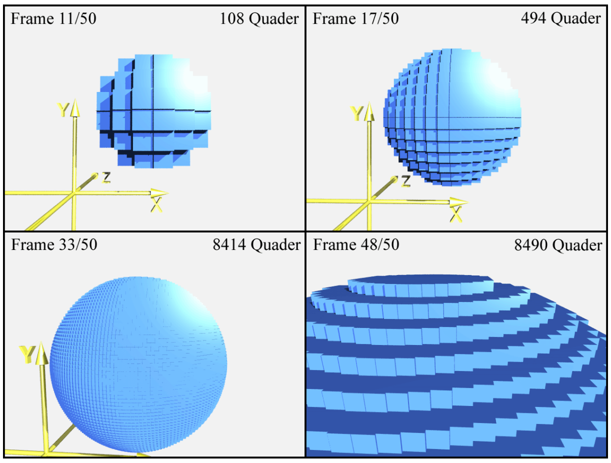
5.7 Definition (σ–Algebra):
Ein System $\mathfrak{F}$ von Teilmengen einer nichtleeren Menge Ω heißt σ–Algebra über Ω, wenn es die folgenden Eigenschaften besitzt:

(i) $\Omega \in \mathfrak{F}$,

(ii) $A \in \mathfrak{F} \Longrightarrow \overline{A} \in \mathfrak{F}$.

(iii) *Für jede Folge* $(A_n)_{n\in\mathbb{N}}$ *von Elementen aus* $\mathfrak{F}$ *ist* $\bigcup\limits_{n\in\mathbb{N}} A_n \in \mathfrak{F}$.

5.8 Beispiel (Visualisierung einer σ–Algebra):



5.9 Bemerkung:

1. $\mathfrak{R} := \{\varnothing\}$ ist der kleinste Ring über Ω für jedes Ω.
2. Es gilt $\{\sigma\text{–Algebra}\} \supseteq \{\text{Algebra}\} \supseteq \{\text{Ring}\} \supseteq \{\text{Semiring}\}$.
3. $\mathfrak{F} := \{\varnothing, \Omega\}$ ist die kleinste σ–Algebra über Ω für jedes Ω.
4. $\mathfrak{F} := \mathfrak{P}(\Omega)$ ist die größte σ–Algebra über Ω für jedes Ω.
5. Jede σ–Algebra über Ω ist eine Algebra über Ω.
6. Ist $\mathfrak{R}$ ein Ring über Ω, dann ist $\mathfrak{R}$ auch $\cap$–stabil, denn:

$$A, B \in \mathfrak{R} \overset{\text{(iii)}}{\Longrightarrow} A \backslash B \in \mathfrak{R},$$
$$A, A\backslash B \in \mathfrak{R} \overset{\text{(iii)}}{\Longrightarrow} A\backslash(A\backslash B) \in \mathfrak{R}.$$

Es gilt aber

$$\begin{aligned}
A\backslash(A\backslash B) &= A \cap (\overline{A\backslash B}) \;=\; A \cap (\overline{A \cap \overline{B}}) \\
&= A \cap (\overline{A} \cup B) && \text{(Reziprozitätsgesetz)} \\
&= (A \cap \overline{A}) \cup (A \cap B) && \text{(Distributivgesetz)} \\
&= A \cap B.
\end{aligned}$$

7. Ein Ring $\mathfrak{R}$ über Ω ist genau dann eine Algebra über Ω, wenn $\Omega \in \mathfrak{R}$ gilt. Um dies zu zeigen wird zunächst angenommen, dass $\mathfrak{R}$ ein Ring über Ω ist und dass $\Omega \in \mathfrak{R}$ ist. In diesem Fall ist noch zu zeigen, dass mit $A \in \mathfrak{R}$ auch $\overline{A} \in \mathfrak{R}$ gilt. Aus $A, \Omega \in \mathfrak{R}$ folgt aber sofort: $\Omega \backslash A = \overline{A} \in \mathfrak{R}$. Es wird jetzt angenommen, dass $\mathfrak{R}$ eine Algebra über Ω ist. In diesem Fall muss noch gezeigt werden, dass mit $A, B \in \mathfrak{A}$ auch $A \backslash B \in \mathfrak{A}$ ist. Aus $A, B \in \mathfrak{A}$ folgt jedoch $A \backslash B = A \cap \overline{B} = \overline{\overline{A} \cup B} \in \mathfrak{A}$. Wegen $\varnothing = \overline{\Omega}$ gehört auch die leere Menge zu $\mathfrak{A}$.

8. Es sei $\mathfrak{F}$ eine σ–Algebra über Ω. Dann gilt: Für jede Folge $(A_n)_{n\in\mathbb{N}}$ von Elementen aus $\mathfrak{F}$ ist $\bigcap_{n\in\mathbb{N}} A_n \in \mathfrak{F}$, denn:
$$A_n \in \mathfrak{F} \Longrightarrow \bar{A}_n \in \mathfrak{F} \Longrightarrow \bigcup_{n\in\mathbb{N}} \bar{A}_n \in \mathfrak{F} \Longrightarrow \overline{\bigcup_{n\in\mathbb{N}} \bar{A}_n} = \bigcap_{n\in\mathbb{N}} \bar{\bar{A}}_n = \bigcap_{n\in\mathbb{N}} A_n \in \mathfrak{F}.$$

Es ist mitunter schwierig, bei einem vorgegebenen Mengensystem direkt festzustellen, ob es sich um eine σ-Algebra handelt. Diese Schwierigkeit lässt sich jedoch mit Hilfe der nachfolgend eingeführten Dynkin-Systemen umgehen.

5.10 Definition (Dynkin–System):
Ein System $\mathfrak{D}$ von Teilmengen von Ω heißt Dynkin–System über Ω, wenn es die folgenden Eigenschaften besitzt:

(i) $\Omega \in \mathfrak{D}$,

(ii) *Für $D, E \in \mathfrak{D}$ mit $D \subseteq E$ gilt: $E \setminus D \in \mathfrak{D}$.*

(iii) *Für jede Folge $(A_n)_{n\in\mathbb{N}}$ paarweise disjunkter Mengen aus $\mathfrak{D}$ ist auch $\bigcup\limits_{n\in\mathbb{N}} A_n \in \mathfrak{D}$.*

5.11 Bemerkung:
- Jede σ-Algebra ist ein Dynkin-System, denn für $A, B \in \mathfrak{F}$ gilt $\overline{B} \in \mathfrak{F}$ und nach Bemerkung 5.9.7 gilt $A \setminus B = A \cap \overline{B} \in \mathfrak{F}$.
- Mit der Wahl $E := \Omega$ in (ii) folgt auch sofort: $A \in \mathfrak{D} \;\Rightarrow\; \overline{A} \in \mathfrak{D}$.

Der nachfolgende Satz 5.12 charakterisiert den Zusammenhang zwischen Dynkin–Systemen und σ–Algebren.

5.12 Satz:
Es sei $\mathfrak{D}$ ein Dynkin–System über Ω. $\mathfrak{D}$ ist genau dann eine σ–Algebra, wenn $\mathfrak{D}$ $\cap$–stabil ist.

Beweis:
Da die Richtung „$\mathfrak{D}$ ist σ-Algebra $\Rightarrow$ $\mathfrak{D}$ ist Dynkin-System" nach Bemerkung 5.11 gilt, ist nur die Umkehrung „$\mathfrak{D}$ ist $\cap$–stabiles Dynkin-System $\Rightarrow \mathfrak{D}$ ist σ-Algebra" zu zeigen.
Da sich die ersten beiden Eigenschaften einer σ-Algebra direkt aus der Definition der Dynkin-Systeme ergeben, bleibt nur noch der Nachweis zu führen, dass für jede Folge $A_1, A_2, \ldots \in \mathfrak{D}$ (A_n nicht notwendig paarweise disjunkt) auch $\bigcup_{n\in\mathbb{N}} A_n \in \mathfrak{D}$ ist.
$\mathfrak{D}$ ist nach Voraussetzung $\cap$–stabil, somit folgt $\bigcap_{m=1}^{n} \bar{A}_m \in \mathfrak{D}$ für alle n. Ferner bildet $(A_n \cap \bigcap_{m=1}^{n-1} \bar{A}_m)_{n\in\mathbb{N}}$ eine Folge paarweise fremder Mengen aus $\mathfrak{D}$. Es gilt
$$\bigcup_{n\in\mathbb{N}} A_n \supseteq \bigcup_{n\in\mathbb{N}} \left(A_n \cap \bigcap_{m=1}^{n-1} \bar{A}_m \right)$$

aber auch

$$\bigcup_{n\in\mathbb{N}} A_n \subseteq \bigcup_{n\in\mathbb{N}} \left(A_n \cap \bigcap_{m=1}^{n-1} \bar{A}_m \right),$$

wie man sich folgendermaßen klarmachen kann: Es sei $x \in \bigcup_{n\in\mathbb{N}} A_n$, dann existiert ein n_0, so dass $x \in A_{n_0}$ und $x \notin A_k$, $k = 1, \ldots, n_0 - 1$ (falls $n_0 > 1$, sonst trivial). Es folgt $x \in \bar{A}_k$, $k = 1, \ldots, n_0 - 1$, sowie $x \in \bigcap_{k=1}^{n_0-1} \bar{A}_k$ und damit $x \in A_{n_0} \cap \bigcap_{k=1}^{n_0-1} \bar{A}_k$.
Insgesamt folgt also

$$\underbrace{\bigcup_{n\in\mathbb{N}} \left(A_n \cap \bigcap_{m=1}^{n-1} \bar{A}_m \right)}_{\in\mathfrak{D}} = \bigcup_{n\in\mathbb{N}} A_n \in \mathfrak{D}\,.$$

5.2 Erzeugendensysteme

Im nächsten Kapitel wird gezeigt, wie Maße von kleineren Mengensystemen (Semiringen) auf größere Mengensysteme (σ-Algebren) fortgesetzt werden können. Dafür muss zunächst beschrieben werden, wie man aus kleinen Mengensystemen größere bilden kann. In diesem Abschnitt werden dafür die sogenannten Erzeugendensysteme eingeführt, die genau dies leisten.

5.13 Satz:
Es seien I eine beliebige Indexmenge und $\mathfrak{X}_i$ für jedes $i \in I$ ein Ring, eine Algebra, ein Dynkin–System oder eine σ–Algebra über Ω. Dann ist

$$\mathfrak{X} := \bigcap_{i\in I} \mathfrak{X}_i = \{A \subseteq \Omega | A \in \mathfrak{X}_i \ \forall \ i \in I\}$$

ein Mengensystem desselben Typs wie die $\mathfrak{X}_i$.

Beweis:
Der Satz wird im Folgenden exemplarisch für Ringe bewiesen. Für Algebren, Dynkin-Systeme und σ-Algebren verläuft der Beweis analog.

1. $\varnothing \in \mathfrak{X}_i \ \forall \ i \implies \varnothing \in \bigcap_{i\in I} \mathfrak{X}_i,$

2. $A \in \bigcap_{i\in I} \mathfrak{X}_i \wedge B \in \bigcap_{i\in I} \mathfrak{X}_i \implies A, B \in \mathfrak{X}_i \ \forall \ i \in I$, $\mathfrak{X}_i$ sind Ringe
   $\implies A \cup B \in \mathfrak{X}_i \ \forall \ i \in I \implies A \cup B \in \bigcap_{i\in I} \mathfrak{X}_i,$

3. $A \in \bigcap_{i\in I} \mathfrak{X}_i \wedge B \in \bigcap_{i\in I} \mathfrak{X}_i \implies A, B \in \mathfrak{X}_i \ \forall \ i \in I$, $\mathfrak{X}_i$ sind Ringe
   $\implies A \backslash B \in \mathfrak{X}_i \ \forall \ i \in I \implies A \backslash B \in \bigcap_{i\in I} \mathfrak{X}_i.$ ■

5.14 Bemerkung:
Der Durchschnitt von Semiringen ist im Allgemeinen **kein** Semiring mehr, wie das folgende Gegenbeispiel zeigt. Es seien

$$\begin{aligned} S_1 &:= \{\varnothing, \{1\}, \{2\}, \{3\}, \{1,2,3\}\} \quad \text{und} \\ S_2 &:= \{\varnothing, \{1\}, \{2,3\}, \{1,2,3\}\} \end{aligned}$$

zwei Semiringe über $\Omega := \{1,2,3\}$. Der Schnitt

$$S := S_1 \cap S_2 = \{\varnothing, \{1\}, \{1,2,3\}\}$$

ist zwar gegenüber der Durchschnittsbildung abgeschlossen und enthält die leere Menge, doch es gilt:

$$\begin{aligned} S_1 :&\ \{1,2,3\} \backslash \{1\} = \{2,3\} = \{2\} \cup \{3\} \text{ mit } \{2\}, \{3\} \in S_1, \\ S_2 :&\ \{1,2,3\} \backslash \{1\} = \{2,3\} \in S_2, \\ S :&\ \{1,2,3\} \backslash \{1\} = \{2,3\} \text{ und } \nexists \bigcup_{i=1}^{n} C_i = \{2,3\} \text{ mit } C_i \in S. \end{aligned}$$

5.15 Satz:
Es sei $\Omega \neq \varnothing$ und $\mathfrak{B}$ ein beliebiges System von Teilmengen von Ω. Dann gibt es unter den Ringen, Algebren, Dynkin–Systemen bzw. σ–Algebren, die $\mathfrak{B}$ enthalten, jeweils ein kleinstes solches System (symbolisch $\mathfrak{M}(\mathfrak{B}) = \mathfrak{R}(\mathfrak{B}), \mathfrak{A}(\mathfrak{B}), \mathfrak{D}(\mathfrak{B})$ bzw. $\sigma(\mathfrak{B})$) nämlich

$$\mathfrak{M}(\mathfrak{B}) := \bigcap \{\mathfrak{M}' \mid \mathfrak{M}' \supseteq \mathfrak{B},\ \mathfrak{M}' \textit{ ist Ring, Algebra, Dynkin–System bzw. } \sigma\textit{–Algebra}\}.$$

$\mathfrak{M}(\mathfrak{B})$ heißt das von $\mathfrak{B}$ erzeugte System und $\mathfrak{B}$ der Erzeuger des Systems.

Beweis:
Die Existenz eines solchen Systems folgt aus der Tatsache, dass die Potenzmenge $\mathfrak{P}(\Omega)$ die Menge $\mathfrak{B}$ umfaßt und alle Eigenschaften eines Ringes, einer Algebra, eines Dynkin–Systems bzw. einer σ–Algebra besitzt. Die Behauptung ergibt sich nun unmittelbar aus Satz 5.13, wonach der Durchschnitt von Ringen, Algebren, Dynkin–Systemen bzw. σ–Algebren wieder ein Ring, eine Algebra, ein Dynkin–System bzw. eine σ–Algebra ist. ■

5.16 Satz:
Es seien Ω eine nichtleere Menge und $\mathcal{E} \subseteq \mathfrak{P}(\Omega)$. Ist $\mathcal{E}$ $\cap$–stabil, so stimmen das von $\mathcal{E}$ erzeugte Dynkin–System $\mathfrak{D}(\mathcal{E})$ und die von $\mathcal{E}$ erzeugte σ–Algebra $\sigma(\mathcal{E})$ überein.

Beweis:
Da jede σ–Algebra auch ein Dynkin–System ist, gilt $\mathfrak{D}(\mathcal{E}) \subseteq \sigma(\mathcal{E})$. Lässt sich umgekehrt nachweisen, dass $\mathfrak{D}(\mathcal{E})$ eine σ–Algebra ist, so folgt auch $\sigma(\mathcal{E}) \subseteq \mathfrak{D}(\mathcal{E})$ und somit $\mathfrak{D}(\mathcal{E}) = \sigma(\mathcal{E})$. Nach Satz 5.12 muss dafür nur noch überprüft werden, ob $\mathfrak{D}(\mathcal{E})$ mit je zwei Mengen A und B auch $A \cap B$ enthält. Betrachte folgendes System für beliebiges $A \in \mathfrak{D}(\mathcal{E})$:

$$\mathfrak{D}_A := \{C \subseteq \Omega \mid A \cap C \in \mathfrak{D}(\mathcal{E})\}.$$

Zeige zunächst, dass $\mathfrak{D}_A$ ein Dynkin–System ist. Mithilfe der $\cap$–Stabilität von $\mathcal{E}$ gilt dann
$\mathcal{E} \subseteq \mathfrak{D}_E$ für alle $E \in \mathcal{E} \Rightarrow \mathfrak{D}(\mathcal{E}) \subseteq \mathfrak{D}_E$,
d.h. $E \cap D \in \mathfrak{D}(\mathcal{E})$ für alle $E \in \mathcal{E}$ und alle $D \in \mathfrak{D}(\mathcal{E})$,
d.h. $\mathfrak{D}(\mathcal{E}) \subseteq \mathfrak{D}_D$ $(D \in \mathfrak{D}(\mathcal{E}))$.
Es bleibt zu zeigen, dass $\mathfrak{D}_A$ ein Dynkinsystem ist:

- Wegen $A \cap \Omega = A \in \mathfrak{D}(\mathcal{E})$ ist $\Omega \in \mathfrak{D}_A$.
- Seien weiter $B, C \in \mathfrak{D}_A$ mit $B \subseteq C$. Dann ist $(A \cap C) \setminus (A \cap B) \in \mathfrak{D}(\mathcal{E})$, weil $A \cap C \supseteq A \cap B$ und $\mathfrak{D}(\mathcal{E})$ Dynkin–System ist. Es gilt aber $(A \cap C) \setminus (A \cap B) = A \cap (C \setminus B)$, so dass $C \setminus B \in \mathfrak{D}_A$.
- Sei nun $(D_n)_{n \in \mathbb{N}}$ eine Folge paarweise fremder Mengen aus $\mathfrak{D}_A$. Da $\mathfrak{D}(\mathcal{E})$ Dynkin–System ist, folgt $\bigcup\limits_{n \in \mathbb{N}} (A \cap D_n) \in \mathfrak{D}(\mathcal{E})$ und wegen $\bigcup\limits_{n \in \mathbb{N}} (A \cap D_n) = A \cap \bigcup\limits_{n \in \mathbb{N}} D_n$ ist deshalb $\bigcup\limits_{n \in \mathbb{N}} D_n \in \mathfrak{D}_A$. ■

5.17 Definition (separabel):
Eine σ–Algebra $\mathfrak{F}$ heißt separabel, wenn es ein abzählbares Teilmengensystem $K \subset \mathfrak{P}(\Omega)$ gibt mit $\sigma(K) = \mathfrak{F}$.

5.18 Satz (Darstellungssatz für Ringe):
Ist S ein Semiring, so ist der von S erzeugte Ring die Klasse K aller Mengen E, die eine endliche Zerlegung der Form

$$E = \bigcup_{i=1}^{n} A_i \ , \quad A_i \in S \ (i = 1, \ldots, n), \ A_i \cap A_j = \varnothing \text{ für } i \neq j,$$

gestatten.

Beweis:
Es ist nur zu zeigen, dass K ein Ring ist. Denn jeder Ring $\mathfrak{R}'$, der S enthält, beinhaltet auch alle Mengen der Form $E = \bigcup_{i=1}^{m} A_i$, $A_i \in S$. Um zu zeigen, dass K ein Ring ist, müssen die Ringeigenschaften nachgewiesen werden.

(i) Es gilt offensichtlich $\varnothing \in K$, da mit $\bigcup_{i=1}^{1} A_i$, $A_i \in S$ automatisch $K \supseteq S$ gilt und $\varnothing \in S$ ist.

(ii) Um $E \setminus D \in K$ zu zeigen, wird die folgende Umformung verwendet:

$$\begin{aligned} E \setminus D &= \left(\bigcup_{i=1}^{m} A_i \right) \setminus \left(\bigcup_{j=1}^{n} B_j \right) = \left(\bigcup_{i=1}^{m} A_i \right) \cap \overline{\bigcup_{j=1}^{n} B_j} = \left(\bigcup_{i=1}^{m} A_i \right) \cap \left(\bigcap_{j=1}^{n} \overline{B}_j \right) \\ &= \bigcup_{i=1}^{m} \bigcap_{j=1}^{n} (A_i \cap \overline{B}_j) = \bigcup_{i=1}^{m} \bigcap_{j=1}^{n} (A_i \setminus B_j), \end{aligned}$$

d.h. $E \setminus D$ besitzt eine endliche Zerlegung in disjunkte Mengen der Form $\bigcap_{j=1}^{n} (A_i \setminus B_j)$, $i = 1, \ldots, m$. Nach Voraussetzung sind aber $A_i, B_j \in S$. Deshalb gilt $A_i \setminus B_j = \bigcup_{\ell=1}^{k} C_\ell$ mit $C_\ell \in S$, weshalb $A_i \setminus B_j$ aus K sein muss. Ist K $\cap$–stabil, so würde $\bigcap_{j=1}^{n} (A_i \setminus B_j) \in K$ für $i = 1, \ldots, m$ folgen. Dann wäre $E \setminus D$ endliche Vereinigung von disjunkten Mengen aus S, d.h. $E \setminus D \in K$.
Die $\cap$–Stabilität von K zeigt man so:
Es seien $E, D \in K$, d.h. es gibt Zerlegungen der Form

$$E = \bigcup_{i=1}^{m} A_i, \ A_i \in S \qquad \text{und} \qquad D = \bigcup_{j=1}^{n} B_j, \ B_j \in S.$$

Mit Hilfe des Distributivgesetzes folgt

$$E \cap D = \left(\bigcup_{i=1}^{m} A_i\right) \cap \left(\bigcup_{j=1}^{n} B_j\right) = \bigcup_{i=1}^{m}\bigcup_{j=1}^{n}(A_i \cap B_j), \quad A_i \cap B_j \in S,$$

d.h. $E \cap D$ besitzt eine Zerlegung in disjunkte Mengen der Form $A_i \cap B_j \in S$. Folglich gilt $E \cap D \in K$, d.h. K ist $\cap$–stabil.

(iii) Die $\cup$-Stabilität lässt sich wie folgt zeigen:
Für $E, D \in K$ ist $E \cup D = (E \backslash D) \cup D$ eine Zerlegung von $E \cup D$ in zwei disjunkte Mengen $E \backslash D \in K$ und $D \in K$. $E \cup D$ ist somit eine endliche Vereinigung von disjunkten Mengen aus S, d.h. $E \cup D \in K$. ■

5.3 Die σ–Algebra der Borelschen Mengen

Im Folgenden sollen noch einmal der Semiring $\mathbb{I}^n := \{(a,b]_{(n)} \mid a, b \in \mathbb{R}^n\}$ aller endlichen, links offenen und rechts abgeschlossenen Intervalle des $\mathbb{R}^n$ sowie die Elemente der von $\mathbb{I}^n$ erzeugten σ–Algebra $\sigma(\mathbb{I}^n)$, die auch σ–Algebra der Borelschen Mengen genannt und mit $\mathfrak{B}^n$ bezeichnet wird, betrachtet werden. Da es sich bei den Ergebnissen von Zufallsexperimenten in der Regel um reelle Zahlen oder reellwertige Vektoren handelt, spielt die σ–Algebra der Borelschen Mengen in der Wahrscheinlichkeitstheorie und deren Anwendungen naturgemäß eine besondere Rolle.
(Siehe auch Lebensdaten von Borel im Anhang D.)

Da die Vereinigung bzw. der Durchschnitt von abzählbar vielen Mengen aus $\mathfrak{B}^n$ wieder ein Element von $\mathfrak{B}^n$ ist, gehören neben den Intervallen $(a,b]_{(n)}$ auch die folgenden Mengen zu $\mathfrak{B}^n$:

$$\begin{aligned}
[a,b]_{(n)} &:= \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_i \leq x_i \leq b_i; i = 1, \ldots, n\} \\
&= \bigcap_{j \in \mathbb{N}} \left(a - \frac{1}{j}, b\right]_{(n)} \in \mathfrak{B}^n, \\
(a,b)_{(n)} &:= \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_i < x_i < b_i; i = 1, \ldots, n\} \\
&= \bigcup_{j \in \mathbb{N}} \left(a, b - \frac{1}{j}\right]_{(n)} \in \mathfrak{B}^n, \\
(-\infty, b]_{(n)} &:= \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid -\infty < x_i \leq b_i; i = 1, \ldots, n\} \\
&= \bigcup_{m \in \mathbb{N}} ((-m, \ldots, -m), b]_{(n)} \in \mathfrak{B}^n, \\
\{b\} &= (a,b]_{(n)} \backslash (a,b)_{(n)} \in \mathfrak{B}^n.
\end{aligned}$$

Es seien

$$K(x,y;r) := \{(\tilde{x}, \tilde{y}) \in \mathbb{R}^2 : (x - \tilde{x})^2 + (y - \tilde{y})^2 < r\}$$

die offene Kreisschreibe um (x,y) mit Radius $r > 0$ und

$$K[x,y;r] := \{(\tilde{x}, \tilde{y}) \in \mathbb{R}^2 : (x - \tilde{x})^2 + (y - \tilde{y})^2 \leq r\}$$

die abgeschlossene Kreisschreibe um (x,y) mit Radius $r > 0$. Beide gehören zu $\mathfrak{B}^2$, denn:

- Es ist
$$K(x,y;r) = \bigcup_{(q_1,q_2)\in(\mathbb{Q}\times\mathbb{Q})\cap K(x,y;r)} (q_1-\delta, q_1+\delta)\times(q_2-\delta,q_2+\delta)$$
mit $\delta := \delta(q_1,q_2,x,y,r) = \frac{1}{\sqrt{2}}\left(r - \sqrt{(q_1-x)^2+(q_2-y)^2}\right)$, wobei $(q_1-\delta,q_1+\delta)\times(q_2-\delta,q_2+\delta)\in\mathfrak{B}^2$ sind.

- Es ist
$$K[x,y;r] = \bigcap_{n=1}^{\infty} K\left(x,y;r+\frac{1}{r}\right).$$

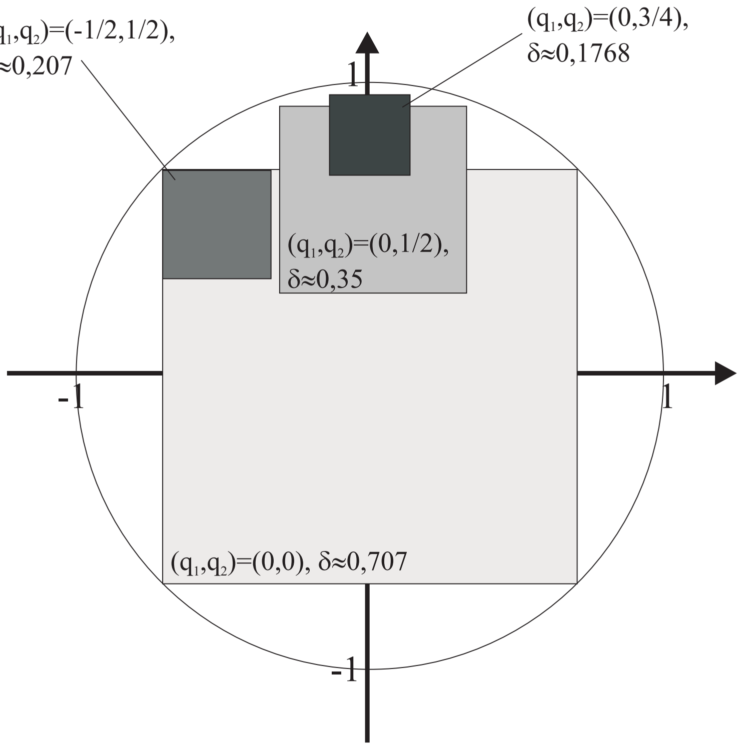


Abbildung 5.1: Veranschaulichung von $K(0,0;1) = \bigcup_{(q_1,q_2)\in(\mathbb{Q}\times\mathbb{Q})\cap K(x,y;r)}(q_1-\delta,q_1+\delta)\times(q_2-\delta,q_2+\delta)$

Auf ähnliche Weise kann gezeigt werden, dass als Erzeugendensystem für $\mathfrak{B}^n$ ebenso die linksseitig abgeschlossenen und rechtsseitig offenen Intervalle des $\mathbb{R}^n$ hätten gewählt werden können. Insbesondere bilden auch die offenen oder die kompakten Mengen des $\mathbb{R}^n$ ein Erzeugendensystem der σ–Algebra der Borelschen Mengen $\mathfrak{B}^n$.

Literatur zu Kapitel 5

Folgende Bücher werden als begleitende Literatur zu diesem Kapitel des Skriptes empfohlen:

- H. Bauer:
  Wahrscheinlichkeitstheorie,
  5. Auflage, Walter de Gruyter, Berlin, 2002.
  ISBN: 3110172364
- W. Behnen, G. Neuhaus:
  Grundkurs Stochastik,
  3. Auflage, Teubner-Verlag, Stuttgart, 1995.
  ISBN: 3930737698
- P. Billingsley:
  Probability and Measure,
  2nd edition, John Wiley and Sons, New York, 1986.
  ISBN: 0471007102
- P. Gänssler und W. Stute:
  Wahrscheinlichkeitstheorie,
  Springer–Verlag, Berlin, 1977.
  ISBN: 3540084185
- E. Henze:
  Einführung in die Maßtheorie,
  Bibl. Institut, Mannheim, 1971.
  ISBN: 341100505X
- R. Mathar/D. Pfeiffer:
  Stochastik für Informatiker,
  Teubner, 1990.
  ISBN: 3519022400
- P. P. Spies:
  Grundlagen stochastischer Modelle,
  Hanser, München, 1982.
  ISBN: 3446137114

Kapitel 6

Mengenfunktionen

Im Folgenden wird der Begriff des Wahrscheinlichkeitsmaßes behandelt. So wie es sich bei den bisherigen Betrachtungen als zweckmäßig erwies, neben der primär interessierenden σ–Algebra auch Mengensysteme mit verwandten Strukturen zu untersuchen, wird es sich auch in diesem Kapitel als nützlich herausstellen, neben Wahrscheinlichkeitsmaßen zunächst allgemeinere Mengenfunktionen zu untersuchen. Im Zusammenhang mit der Konstruktion allgemeiner Wahrscheinlichkeitsmaße spielen die beiden Maßfortsetzungssätze von C. Caratheodory eine zentrale Rolle, die zusammen mit dem Axiomensystem von A.N. Kolmogorov die Grundlage der modernen Wahrscheinlichkeitstheorie bilden.

Schlüsselwörter: Inhalt, Prämaß, Maß, Wahrscheinlichkeitsmaß, Stetigkeit von unten, Stetigkeit von oben, Fortsetzung, 1. und 2. Maß–Fortsetzungssatz, σ–endlich, äußeres Maß.

6.1 Grundbegriffe

In diesem Absachnitt geht es allgemein um Mengenfunktionen und ihre Eigenschaften.

6.1 Definition (nichtnegativ, additiv, σ–additiv, subadditiv, σ–subadditiv):
Es seien S ein Semiring über Ω und $\mu\colon S \to \mathbb{R} \cup \{\infty\}$ eine numerische Funktion.

- *μ heißt <u>nichtnegativ</u>, wenn $\mu(\varnothing) = 0$ und $\mu(A) \geq 0$ für alle $A \in S$ ist.*
- *μ heißt <u>additiv</u>, wenn für alle $A, B \in S$ mit $A \cap B = \varnothing$ und $A \cup B \in S$ gilt:*
$$\mu(A \cup B) = \mu(A) + \mu(B).$$
- *μ heißt <u>σ–additiv</u>, wenn für jede Folge $(A_n)_{n\in\mathbb{N}}$ von paarweise fremden Elementen aus S (d.h. $A_i \cap A_j = \varnothing$ für $i \neq j$) mit $\bigcup_{n\in\mathbb{N}} A_n \in S$ gilt:*
$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n).$$
- *μ heißt <u>subadditiv</u>, wenn für alle $A, B \in S$ mit $A \cup B \in S$ gilt:*
$$\mu(A \cup B) \leq \mu(A) + \mu(B).$$
- *μ heißt <u>σ–subadditiv</u>, wenn für jede Folge $(A_n)_{n\in\mathbb{N}}$ von Elementen aus S mit $\bigcup_{n\in\mathbb{N}} A_n \in S$ gilt:*
$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} \mu(A_n).$$

6.2 Bemerkung:
Die Einschränkung $S \to \mathbb{R}\cup\{\infty\}$ anstelle von $S \to \mathbb{R}\cup\{-\infty, +\infty\}$ wird gemacht, um sinnlose Ausdrücke wie $\infty - \infty$ zu vermeiden.

6.3 Definition (Inhalt, Prämaß und Maß):
Es seien S ein Semiring über Ω und $\mu\colon S \to \mathbb{R} \cup \{\infty\}$ eine numerische Funktion. Dann gilt:

- *μ heißt <u>Inhalt</u>, wenn μ nichtnegativ und additiv ist.*
- *μ heißt <u>Prämaß</u>, wenn μ nichtnegativ und σ–additiv ist.*

Mit Hilfe der Begriffe Inhalt und Prämaß lassen sich nun die zentralen Begriffe der Maßtheorie und der Wahrscheinlichkeitstheorie definieren:

- *μ heißt <u>Maß</u>, wenn μ Prämaß und S eine σ–Algebra ist.*
- *μ heißt <u>Wahrscheinlichkeitsmaß</u>, wenn μ ein Maß ist und $\mu(\Omega) = 1$ gilt.*

6.4 Definition (endlich):

- *Ein Inhalt oder Prämaß μ heißt <u>endlich</u>, wenn $\mu(A) < \infty$ für alle $A \in S$ ist.*
- *Ein Maß μ heißt <u>endlich</u>, falls $\mu(\Omega) < \infty$ ist.*

6.5 Bemerkung:

1. Es seien $\mathfrak{R}$ ein Ring über Ω und $\omega \in \Omega$. Die Abbildung $\mu\colon \mathfrak{R} \to \mathbb{R} \cup \{\infty\}$ sei definiert durch
$$\mu(A) := \begin{cases} 1\,, & \omega \in A \\ 0\,, & \omega \notin A \end{cases}.$$
Dann ist μ ein endliches Prämaß. Ist $\mathfrak{R}$ eine σ–Algebra, so ist μ ein Wahrscheinlichkeitsmaß (sogenanntes Dirac–Maß).

   Veranschaulichung:

   Gegeben sei die σ-Algebra $\mathfrak{R} := \{\varnothing, \{1\}, \{2,3\}, \{1,2,3\}\}$ über $\Omega := \{1,2,3\}$. Wähle $\omega := 1$. Dann ergibt sich $\mu(\varnothing) = 0$, $\mu(\{1\}) = 1$, $\mu(\{2,3\}) = 0$, $\mu(\{1,2,3\}) = 1$.

2. Jedes Prämaß ist ein Inhalt.

3. Es sei $(\mu_n)_{n\in\mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmaßen, die alle auf einer σ–Algebra $\mathfrak{F}$ über Ω definiert sind. Es sei weiter $(\alpha_n)_{n\in\mathbb{N}}$ eine Folge von nichtnegativen reellen Zahlen mit $\sum_{n\in\mathbb{N}} \alpha_n = 1$. Die numerische Funktion $\mu\colon \mathfrak{F} \to \mathbb{R} \cup \{\infty\}$ mit
$$\mu(A) := \sum_{n\in\mathbb{N}} \alpha_n \mu_n(A), \quad \forall\, A \in \mathfrak{F}$$
ist ein Wahrscheinlichkeitsmaß auf $\mathfrak{F}$.

   Beweis der σ–Additivität:
$$\begin{aligned} \mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) &= \sum_{m\in\mathbb{N}} \alpha_m \mu_m \left(\bigcup_{n\in\mathbb{N}} A_n\right) && \text{(Def. von } \mu\text{)} \\ &= \sum_{m\in\mathbb{N}} \alpha_m \sum_{n\in\mathbb{N}} \mu_m(A_n) && (\sigma\text{–Additivität der } \mu_m) \\ &= \sum_{n\in\mathbb{N}} \sum_{m\in\mathbb{N}} \alpha_m\, \mu_m(A_n) && \text{(Umordnungssatz für abs. konv. Reihen)} \\ &= \sum_{n\in\mathbb{N}} \mu(A_n) && \text{(Def. von } \mu\text{).} \end{aligned}$$

4. Es sei $F\colon \mathbb{R} \to \mathbb{R}$ eine monoton wachsende Funktion. Die auf dem Semiring $\mathbb{I}^1$ der links offenen und rechts abgeschlossenen Intervalle $(a,b] \subset \mathbb{R}$, $a \le b$ durch
$$\mu((a,b]) := F(b) - F(a)$$
definierte Mengenfunktion ist ein endlicher Inhalt auf $\mathbb{I}^1$. Denn es gilt:

   (a) $\mu(\varnothing) = \mu((a,a]) = F(a) - F(a) = 0$.

   (b) $\mu((a,b]) = F(b) - F(a) \ge 0$ für $a \le b$ aufgrund der Monotonie von F.

   (c) Es seien $(a,b]$ und $(a',b']$ zwei Intervalle aus $\mathbb{I}^1$ mit $b = a'$. Die Eigenschaft $b = a'$ wird gefordert, um $(a,b] \cup (a',b'] = (a,b'] \in \mathbb{I}^1$ und $(a,b] \cap (a',b'] = \varnothing$ sicherzustellen. Dann gilt:
$$\begin{aligned} \mu((a,b] \cup (a',b']) &= \mu((a,b']) = F(b') - F(a) \\ &= F(a') - F(a) + F(b') - F(a') \\ &= F(b) - F(a) + F(b') - F(a') && \text{(wegen } b = a'\text{)} \\ &= \mu((a,b]) + \mu((a',b']). \end{aligned}$$

6.6 Satz:
Es seien $\mathfrak{R}$ ein Ring über Ω und $\mu\colon \mathfrak{R} \to \mathbb{R} \cup \{\infty\}$ ein Inhalt. Dann gilt:

a) Für alle $A, B \in \mathfrak{R}$ gilt: $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$.

b) μ ist monoton, d.h. $\forall\ A, B \in \mathfrak{R}$ mit $A \subset B$ gilt $\mu(A) \leq \mu(B)$. (Isotonie)

c) Für alle $A, B \in \mathfrak{R}$ mit $A \subset B$ und $\mu(A) < \infty$ gilt $\mu(B \backslash A) = \mu(B) - \mu(A)$. (Subtraktivität)

d) μ ist subadditiv.

e) Ist μ ein Prämaß, dann ist μ σ–subadditiv.

Beweis:
a) Für alle $A, B \in \mathfrak{R}$ gilt:

$$\begin{aligned} A \cup B &= A \cup (B \backslash A), \text{ wobei } A \cap (B \backslash A) = \varnothing \text{ ist,} \\ (A \cap B) \cup (B \backslash A) &= B, \text{ wobei } (A \cap B) \cap (B \backslash A) = \varnothing \text{ ist.} \end{aligned}$$

Damit wird

$$\begin{aligned} \mu(A \cup B) &= \mu(A) + \mu(B \backslash A), \\ \mu(A \cap B) + \mu(B \backslash A) &= \mu(B). \end{aligned}$$

Addition dieser beiden Gleichungen ergibt:

$$\begin{aligned} \mu(A \cup B) + \mu(A \cap B) + \mu(B \backslash A) &= \mu(A) + \mu(B \backslash A) + \mu(B) \text{ bzw.} \\ \mu(A \cup B) + \mu(A \cap B) &= \mu(A) + \mu(B). \end{aligned}$$

b) und c) Für alle $A, B \in \mathfrak{R}$ mit $A \subset B$ gilt: $B = A \cup (B \backslash A)$ und $A \cap (B \backslash A) = \varnothing$. Damit gilt:

$$\mu(B) = \mu(A) + \mu(B \backslash A) \iff \mu(B \backslash A) = \mu(B) - \mu(A) \implies \mu(B) \geq \mu(A).$$

d) Aufgrund von a) gilt:

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

e) Es sei $(A_n)_{n \in \mathbb{N}}$ eine Folge von Elementen aus $\mathfrak{R}$ mit $\bigcup_{n \in \mathbb{N}} A_n \in \mathfrak{R}$. Es werden $B_1 := A_1$ und $B_n := A_n \backslash \bigcup\limits_{m=1}^{n-1} A_m$ für $n \geq 2$ gesetzt. Dann gilt $\bigcup\limits_{n \in \mathbb{N}} B_n = \bigcup\limits_{n \in \mathbb{N}} A_n$, $B_n \subseteq A_n$ und $B_n \in \mathfrak{R}$ für alle n sowie $B_i \cap B_j = \varnothing$ für $i \neq j$. Aus der σ–Additivität von μ folgt jetzt

$$\begin{aligned} \mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= \mu\left(\bigcup_{n \in \mathbb{N}} B_n\right) \\ &= \sum_{n \in \mathbb{N}} \mu(B_n) && \text{(da } \mu \ \sigma\text{–additiv ist)} \\ &\leq \sum_{n \in \mathbb{N}} \mu(A_n) && \text{(da } \mu \text{ monoton ist)} \end{aligned}$$

■

6.7 Satz:
Es seien $\mathfrak{F}$ eine σ–Algebra über Ω und $\mu\colon \mathfrak{F} \to \mathbb{R} \cup \{\infty\}$ ein Maß. Dann gilt:

a) Für jede Folge $(A_n)_{n\in\mathbb{N}}$ von Elementen aus $\mathfrak{F}$ mit $A_n \subseteq A_{n+1}$ für alle $n \in \mathbb{N}$ (sogenannte monoton wachsende Folge) gilt:

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

b) Ist μ endlich, dann gilt für jede Folge $(A_n)_{n\in\mathbb{N}}$ von Elementen aus $\mathfrak{F}$ mit $A_{n+1} \subseteq A_n$ für alle $n \in \mathbb{N}$ (sogenannte monoton fallende Folge):

$$\mu\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

Die unter a) angegebene Eigenschaft von μ bezeichnet man als Stetigkeit von unten, die unter b) als Stetigkeit von oben.

Beweis:
a) O.B.d.A. sei $\mu(A_n) < \infty \;\; \forall\, n \in \mathbb{N}$ (sonst trivial).
Es wird $B_1 := A_1$ und $B_{n+1} = A_{n+1} \backslash A_n$ für $n \geq 1$ gesetzt. Dann gilt: $B_i \cap B_j = \varnothing$ für $i \neq j$ und $\bigcup_{n\in\mathbb{N}} B_n = \bigcup_{n\in\mathbb{N}} A_n$. Hieraus folgt:

$$\begin{aligned}
\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) &= \mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) \\
&= \sum_{n\in\mathbb{N}} \mu(B_n) && (\mu \text{ ist } \sigma\text{–additiv}) \\
&= \mu(A_1) + \sum_{n=1}^{\infty} \mu(A_{n+1} \backslash A_n) \\
&= \mu(A_1) + \lim_{m\to\infty} \sum_{n=1}^{m} \mu(A_{n+1}) - \mu(A_n) && (\text{aufgrund von Satz 6.6 c))} \\
&= \lim_{m\to\infty} \mu(A_m).
\end{aligned}$$

b) Allgemein gilt:

$$\begin{aligned}
\mu\left(\bigcap_{n\in\mathbb{N}} A_n\right) &= \mu\left(\overline{\bigcup_{n\in\mathbb{N}} \overline{A_n}}\right) && (\text{Regeln von de Morgan}) \\
&= \mu\left(\Omega \backslash \bigcup_{n\in\mathbb{N}} \overline{A_n}\right) && (\text{Definition des Komplements}) \\
&= \mu(\Omega) - \mu\left(\bigcup_{n\in\mathbb{N}} \overline{A_n}\right) && (\text{aufgrund von Satz 6.6 c))} \\
&= \mu(\Omega) - \mu\left(\bigcup_{n\in\mathbb{N}} (\Omega \backslash A_n)\right).
\end{aligned}$$

Es gilt aber $\Omega\backslash A_1 \subseteq \Omega\backslash A_2 \subseteq \Omega\backslash A_3 \subseteq \cdots$. Deshalb kann Satz 6.7 a) angewandt werden, mit dem sich

$$\mu\left(\bigcup_{n\in\mathbb{N}} (\Omega\backslash A_n)\right) = \lim_{n\to\infty} \mu(\Omega\backslash A_n) = \mu(\Omega) - \lim_{n\to\infty} \mu(A_n)$$

ergibt. Damit wird

$$\mu\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

■

6.2 Erster Maß–Fortsetzungssatz

Die beiden nachfolgenden Abschnitte sind der Konstuktion allgemeiner Wahrscheinlichkeitsmaße gewidmet. Die Grundlage hierfür bilden die beiden Maßfortsetzungssätze, die von C. Caratheodory stammen und auf folgender Erkenntnis beruhen: Ist S ein durchschnittsstabiles Mengensystem und P ein Wahrscheinlichkeitsmaß auf der von S erzeugten σ–Algebra, dann ist untergewissen Regularitätsbedingungen P durch seine Einschränkung $P|_S$ auf S bereits eindeutig bestimmt. Für die Praxis genügt es folglich, eine geeignete Mengenfunktion P auf S zu definieren und nachzuweisen, dass ihre Erweiterung auf $\mathfrak{F}$ ein Wahrscheinlichkeitsmaß darstellt. Dieser Nachweis erfolgt in zwei Schritten. Man betrachtet zunächst die Erweiterung von P auf den von S erzeugten Ring (erster Fortsetzungssatz) und schließt dann weiter auf die von S erzeugte σ–Algebra (zweiter Fortsetzungssatz).

6.8 Definition (Fortsetzung, Erweiterung, Restriktion, Einschränkung):
Es seien $\mathfrak{M}_1$ und $\mathfrak{M}_2$ zwei Mengensysteme über Ω mit $\mathfrak{M}_1 \subset \mathfrak{M}_2$. Gilt für die beiden Mengenfunktionen $\nu\colon \mathfrak{M}_1 \to \mathbb{R} \cup \{\infty\}$ und $\mu\colon \mathfrak{M}_2 \to \mathbb{R} \cup \{\infty\}$ die Beziehung $\nu(A) = \mu(A)$ für alle $A \in \mathfrak{M}_1$, dann nennt man μ eine Fortsetzung (Erweiterung) von ν auf $\mathfrak{M}_2$ und ν eine Restriktion (Einschränkung) von μ auf $\mathfrak{M}_1$.

6.9 Satz (1. Fortsetzungssatz):
Für jeden Inhalt ν auf einem Semiring S ist

$$\mu(A) = \sum_{i=1}^{n} \nu(A_i) \quad \textit{mit} \quad A := \bigcup_{i=1}^{n} A_i,\ A_i \in S,\ i = 1,\dots,n, \quad \textit{und } A_i \cap A_j = \varnothing \quad \textit{für } i \neq j$$

die einzige Fortsetzung von ν zu einem Inhalt auf $\mathfrak{R} = \mathfrak{R}(S)$. Ist ν ein Prämaß, dann ist auch μ ein Prämaß.

Beweis:
Nach dem Darstellungssatz für Ringe 5.18 kann jede Menge $E \in \mathfrak{R}(S)$ in der Form $E = \bigcup\limits_{i=1}^{p} A_i$, $A_i \in S$ für $i = 1,\dots,p$ und $A_i \cap A_j = \varnothing$ für $i \neq j$ dargestellt werden. Es wird

$$\mu(E) = \sum_{i=1}^{p} \nu(A_i)$$

gesetzt und gezeigt:

(i) μ ist wohldefiniert, d.h. $\mu(E)$ ist unabhängig von der gewählten Zerlegung der Menge E.

(ii) μ ist Inhalt.

(iii) μ ist eindeutig.

(iv) Ist ν Prämaß, so ist auch μ Prämaß.

Zu (i) μ ist wohldefiniert. Sind $E = \bigcup_{i=1}^{p} A_i$ und $E = \bigcup_{j=1}^{q} B_j$ mit $A_i, B_j \in S$ ($i = 1, \dots, p$; $j = 1, \dots, q$) und $A_i \cap A_k = \varnothing$ für $i \neq k$, $B_j \cap B_\ell = \varnothing$ für $j \neq \ell$ zwei endliche Zerlegungen von E, dann ist zu zeigen:

$$\sum_{i=1}^{p} \nu(A_i) \overset{!}{=} \sum_{j=1}^{q} \nu(B_j).$$

Offensichtlich sind

$$A_i = A_i \cap E = A_i \cap \left(\bigcup_{j=1}^{q} B_j \right) = \bigcup_{j=1}^{q} (A_i \cap B_j) \qquad (i = 1, \dots, p)$$
$$B_j = E \cap B_j = \left(\bigcup_{i=1}^{p} A_i \right) \cap B_j = \bigcup_{i=1}^{p} (A_i \cap B_j) \qquad (j = 1, \dots, q)$$

Zerlegungen von A_i und B_j in paarweise fremde Mengen $A_i \cap B_j \in S$ für $i = 1, \dots, p$ und $j = 1, \dots, q$. Es gilt also

$$\begin{aligned}
\sum_{i=1}^{p} \nu(A_i) &= \sum_{i=1}^{p} \nu \left(\bigcup_{j=1}^{q} (A_i \cap B_j) \right) \\
&= \sum_{i=1}^{p} \sum_{j=1}^{q} \nu(A_i \cap B_j) && \text{(da } \nu \text{ additiv ist)} \\
&= \sum_{j=1}^{q} \sum_{i=1}^{p} \nu(A_i \cap B_j) \\
&= \sum_{j=1}^{q} \nu \left(\bigcup_{i=1}^{p} (A_i \cap B_j) \right) \\
&= \sum_{j=1}^{q} \nu(B_j).
\end{aligned}$$

Zu (ii) μ ist ein Inhalt. Hierfür ist zu zeigen, dass μ nichtnegativ und additiv ist. Die Nichtnegativität von μ folgt unmittelbar aus der Definition von μ. Für die Additivität wird $E := E' \cup E''$ mit $E' \cap E'' = \varnothing$ und $E, E', E'' \in \mathfrak{R}(S)$ betrachtet. Es existiert dann Zerlegungen

$$E' = \bigcup_{i=1}^{p} A'_i, \quad E'' = \bigcup_{j=1}^{q} A''_j, \quad A'_i, A''_j \in S, \quad i = 1, \dots, p, \quad j = 1, \dots, q,$$

so dass

$$E = E' \cup E'' = \bigcup_{i=1}^{p} A_i' \cup \bigcup_{j=1}^{q} A_j''$$

gilt. Da wegen $E' \cap E'' = \varnothing$ auch $A_i' \cap A_j'' = \varnothing$ ist, gilt

$$\mu(E) = \sum_{i=1}^{p} \nu(A_i') + \sum_{j=1}^{q} \nu(A_j'') = \mu(E') + \mu(E'').$$

Zu (iii) μ ist eindeutig. Es sei ϑ eine weitere Erweiterung von ν und $E \in \mathfrak{R}(S)$ mit $E := \bigcup_{i=1}^{p} A_i$, dann gilt:

$$\begin{aligned} \vartheta(E) &= \sum_{i=1}^{p} \vartheta(A_i) && (\vartheta \text{ ist Inhalt}) \\ &= \sum_{i=1}^{p} \mu(A_i) && (\mu = \vartheta \text{ auf } S) \\ &= \mu(E). \end{aligned}$$

Zu (iv) Ist ν Prämaß, so ist auch μ Prämaß. Zunächst wird gezeigt: Mit ν ist auch μ σ–additiv. Ist $E := \bigcup_{n \in \mathbb{N}} E_n$ eine Zerlegung von $E \in \mathfrak{R}(S)$ mit $E_n \in \mathfrak{R}(S)$ und $E_n \cap E_m = \varnothing$ für $n \neq m$, so ist zu zeigen:

$$\mu(E) \overset{!}{=} \sum_{n \in \mathbb{N}} \mu(E_n).$$

Aufgrund des Darstellungssatzes für Ringe 5.18 existieren für E und E_n Zerlegungen der Form:

$$\begin{aligned} E &= \bigcup_{i=1}^{p} A_i, \quad A_i \in S,\ A_i \cap A_j = \varnothing \ \text{ für } i \neq j, \\ E_n &= \bigcup_{j=1}^{p_n} B_{nj}, \quad B_{nj} \in S,\ B_{nj} \cap B_{nk} = \varnothing \ \text{ für } j \neq k. \end{aligned}$$

Hieraus folgt:

$$\begin{aligned} E &= \bigcup_{n=1}^{\infty} \bigcup_{j=1}^{p_n} B_{nj}, \\ A_i = A_i \cap E &= \bigcup_{n=1}^{\infty} \bigcup_{j=1}^{p_n} (A_i \cap B_{nj}), \\ B_{nj} = E \cap B_{nj} &= \bigcup_{i=1}^{p} (A_i \cap B_{nj}). \end{aligned}$$

Die Mengen $A_i \cap B_{nj}$ sind paarweise fremd. Folglich gilt:

$$
\begin{aligned}
\mu(E) &= \sum_{i=1}^{p} \nu(A_i) && \text{(Definition der Erweiterung)}\\
&= \sum_{i=1}^{p} \nu\left(\bigcup_{n=1}^{\infty}\bigcup_{j=1}^{p_n}(A_i \cap B_{nj})\right) && \text{(aufgrund der speziellen Zerlegung der } A_i\text{)}\\
&= \sum_{i=1}^{p}\sum_{n=1}^{\infty}\sum_{j=1}^{p_n} \nu(A_i \cap B_{nj}) && (\nu \text{ ist nach Voraussetzung } \sigma\text{–additiv})\\
&= \sum_{n=1}^{\infty}\sum_{j=1}^{p_n}\sum_{i=1}^{p} \nu(A_i \cap B_{nj}) && \text{(Umordnungssatz für abs. konv. Reihen)}\\
&= \sum_{n=1}^{\infty}\sum_{j=1}^{p_n} \nu\left(\bigcup_{i=1}^{p}(A_i \cap B_{nj})\right) && (\nu \text{ ist additiv})\\
&= \sum_{n=1}^{\infty}\sum_{j=1}^{p_n} \nu(B_{nj}) && \text{(aufgrund der speziellen Zerlegung der } B_{nj}\text{)}\\
&= \sum_{n=1}^{\infty}\sum_{j=1}^{p_n} \mu(B_{nj}) && (\mu = \nu \text{ auf } S)\\
&= \sum_{n=1}^{\infty} \mu\left(\bigcup_{j=1}^{p_n} B_{nj}\right) && (\mu \text{ ist additiv})\\
&= \sum_{n=1}^{\infty} \mu(E_n).
\end{aligned}
$$

■

6.3 Zweiter Maß–Fortsetzungssatz

Im ersten Maß–Fortsetzungssatz wurde die Erweiterung eines Prämaßes ν auf einem Semiring S zu einem Prämaß ν auf dem von S erzeugten Ring $\mathfrak{R} = \mathfrak{R}(S)$ behandelt. Der zweite Maßfortsetzungssatz behandelt die Erweiterung eines Prämaßes μ auf S zu einem Maß auf $\sigma(S)$. Für den Beweis, der auf eine Idee von C. Carathéodory zurückgeht, benötigen wir einige Vorbereitungen.

Für beliebiges $U \in \mathfrak{P}(\Omega)$ wird $\widehat{S}(U)$ als das System aller Folgen $A_1, A_2, \ldots \in S$ mit $U \subseteq \bigcup_{n=1}^{\infty} A_n$ definiert. Es sei

$$
\mu^*(U) := \inf\left\{\sum_{n=1}^{\infty} \mu(A_n) \mid (A_n)_{n\in\mathbb{N}} \in \widehat{S}(U)\right\} \quad \text{für alle } U \in \mathfrak{P}(\Omega)\,, \tag{6.1}
$$

hierbei ist $\inf \varnothing := \infty$.

Im folgenden Satz werden zunächst einige wichtige Eigenschaften von μ^* festgehalten.

6.10 Satz:
Die in (6.1) definierte Fortsetzung μ^* des Prämaßes μ auf dem Semiring S über Ω besitzt folgende Eigenschaften:

(1) $\mu^*(U) \geq 0$ *für alle* $U \in \mathfrak{P}(\Omega)$,

(2) $\mu^*(\varnothing) = 0$,

(3) $U_1 \subset U_2 \Rightarrow \mu^*(U_1) \leq \mu^*(U_2)$,

(4) $\mu^*\left(\bigcup_{n=1}^{\infty} U_n\right) \leq \sum_{n=1}^{\infty} \mu^*(U_n)$.

Beweis:

(1) $\mu^*(U) \geq 0$ folgt direkt aus der Definition von μ^*.

(2) $\mu^*(\varnothing) = 0$ folgt mit $(\varnothing, \varnothing, \ldots) \in \widehat{S}(\varnothing)$.

(3) $U_1 \subseteq U_2 \Rightarrow \mu^*(U_1) \leq \mu^*(U_2)$ folgt aus $U_1 \subseteq U_2 \Rightarrow \widehat{S}(U_1) \subseteq \widehat{S}(U_2)$.

(4) Zu zeigen ist $\mu^*\left(\bigcup_{n=1}^{\infty} U_n\right) \leq \sum_{n=1}^{\infty} \mu^*(U_n)$. Es wird $\mu^*(U_n) < \infty$ angenommen. Wähle ein beliebiges, aber festes $\varepsilon > 0$.
Für jedes $n \in \mathbb{N}$ gibt es eine Folge $(A_{n,m})_{n,m=1,2,\ldots} \in \widehat{S}(U)$ mit

$$\sum_{m=1}^{\infty} \mu(A_{n,m}) \leq \mu^*(U_n) + \frac{\varepsilon}{2^n}\,.$$

Die Folge $(A_{n,m})_{n,m=1,2,\ldots}$ liegt in $\widehat{S}(\bigcup_{n=1}^{\infty} U_n)$. Hieraus entsteht

$$\mu^*\left(\bigcup_{n=1}^{\infty} U_n\right) \leq \sum_{m,n} \mu(A_{n,m}) \leq \sum_{n=1}^{\infty}\left(\mu^*(U_n) + \frac{\varepsilon}{2^n}\right) \leq \sum_{n=1}^{\infty} \mu^*(U_n) + \varepsilon.$$

Da $\varepsilon > 0$ beliebig gewählt war, folgt die Aussage. ∎

6.11 Definition (äußeres Maß, induziertes äußeres Maß):

a) Jede Abbildung $\mu^*\colon \mathfrak{P}(\Omega) \to \mathbb{R} \cup \{\infty\}$ *mit den Eigenschaften aus Satz 6.10 heißt ein äußeres Maß auf* Ω.

b) Die durch (6.1) definierte Mengenfunktion μ^* *wird als das vom Prämaß* μ *induzierte äußere Maß bezeichnet.*

Im Folgenden wird die eindeutig bestimmte Fortsetzung des Prämaßes μ auf den Ring $\mathcal{R} := \mathcal{R}(S)$ wieder mit μ bezeichnet (vgl. Satz 6.9).

6.12 Lemma:
Es gilt $\mu^*(B) = \mu(B)$ *für alle* $B \in \mathcal{R}$.

Beweis:
Der Darstellungssatz für Ringe 5.18 besagt, dass jedes $B \in \mathcal{R}$ in der Form $B = \bigcup_{i=1}^{n} C_i, n \in \mathbb{N}, C_i \in S, C_i \cap C_j = \varnothing, i \neq j$ dargestellt werden kann. Daraus folgt:

$$(C_1, C_2, \dots, C_n, \varnothing, \varnothing, \dots) \in \widehat{S}(B) \Rightarrow \mu^*(B) \leq \mu(B).$$

Wenn $\mu^*(B) = \infty$ ist, ist alles gezeigt. Es sei also $\mu^*(B) < \infty \Rightarrow \widehat{S}(B) \neq \varnothing$, und es sei nun $(A_n)_{n\in\mathbb{N}} \in \widehat{S}(B)$. Es folgt: $A_n \in S \subset \mathcal{R} \Rightarrow (A_n \cap B) \in \mathcal{R}$. $\bigcup_{n=1}^{\infty} A_n \supseteq B \Rightarrow B = \bigcup_{n=1}^{\infty}(A_n \cap B)$. Wird nun $D_1 := (A_1 \cap B)$ und $D_n := (A_n \cap B) \setminus \bigcup_{j=1}^{n-1}(A_j \cap B) \in \mathcal{R}$ für $n \geq 2$ gesetzt, so ergibt sich

$$B = \bigcup_{n=1}^{\infty} D_n,\ D_i \cap D_j = \varnothing, i \neq j.$$

Da μ ein Prämaß ist und $D_n \subset A_n$, $n \in \mathbb{N}$, ist, folgt

$$\mu(B) = \sum_{n=1}^{\infty} \mu(D_n) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

Also ist

$$\mu(B) \leq \mu^*(B).$$

■

Existenz

Wenn man zeigen kann, dass jedes von $\mu\colon S \to \mathbb{R} \cup \{\infty\}$ induzierte äußere Maß eingeschränkt auf $\sigma(S)$ ein Maß bildet, hätte man eine Fortsetzung gefunden.

Für das weitere Vorgehen wird der Begriff der μ^*–Messbarkeit eingeführt:

6.13 Definition (μ^*–messbar):
Es sei μ^* ein äußeres Maß auf Ω. $G \in \mathfrak{P}(\Omega)$ heißt $\underline{\mu^*\text{–messbar}}$, falls

$$\mu^*(U) = \mu^*(U \cap G) + \mu^*(U \cap \overline{G}) \text{ für alle } U \in \mathfrak{P}(\Omega) \tag{6.2}$$

gilt.

6.14 Satz:
Es sei μ^* ein äußeres Maß auf Ω. Dann ist das System $\mathfrak{A}$ aller μ^*–messbaren Mengen $G \subset \Omega$ eine σ–Algebra über Ω, und die Restriktion von μ^* auf $\mathfrak{A}$ ist ein Maß.

Beweis:
1. Schritt: $\mathfrak{A}$ ist eine Algebra und μ^* ein Prämaß auf $\mathfrak{A}$.

$\Omega \in \mathfrak{A} : \mu^*(U) = \mu^*(U \cap \Omega) + \mu^*(U \cap \varnothing)$. $G \in \mathfrak{A} \Rightarrow \overline{G} \in \mathfrak{A}$ (wegen (6.2)).

Es seien nun $G, H \in \mathfrak{A}$ und $U \subseteq \Omega$ beliebig:

$$\begin{aligned} \mu^*(U) &= \mu^*(U \cap G) + \mu^*(U \cap \overline{G}) \\ &= \mu^*(U \cap G \cap H) + \mu^*(U \cap G \cap \overline{H}) + \mu^*(U \cap \overline{G} \cap H) + \mu^*(U \cap \overline{G} \cap \overline{H}). \end{aligned} \tag{6.3}$$

Wenn U durch $U \cap (G \cup H)$ in (6.3) ersetzt wird, ergibt sich:

$$\mu^*(U \cap (G \cup H)) = \mu^*(U \cap G \cap H) + \mu^*(U \cap G \cap \overline{H}) + \mu^*(U \cap \overline{G} \cap H). \qquad (6.4)$$

(6.4) in (6.3) eingesetzt ergibt:

$\mu^*(U) = \mu^*(U \cap (G \cup H)) + \mu^*(U \cap \overline{(G \cup H)})$, d.h. $G \cup H \in \mathfrak{A} \Rightarrow \mathfrak{A}$ ist eine Algebra.

Weiter wird gezeigt, dass μ^* σ-additiv auf $\mathfrak{A}$ ist.

$G_1, G_2, \ldots \in \mathfrak{A}$ seien paarweise disjunkt und $U \subseteq \Omega$ beliebig. Nach (6.3) und mit $G = G_1$ und $H = G_2$ folgt:

$$\mu^*(U \cap (G_1 \cup G_2)) = \mu^*(U \cap G_1) + \mu^*(U \cap G_2). \qquad (6.5)$$

Durch vollständige Induktion ergibt sich daraus:

$$\mu^*\left(U \cap \bigcup_{j=1}^{n} G_j\right) = \sum_{j=1}^{n} \mu^*(U \cap G_j) \text{ für alle } n \in \mathbb{N}. \qquad (6.6)$$

Wird in (6.6) $U := \Omega$ gesetzt, so ergibt sich wegen Satz 6.10(3)

$$\mu^*\left(\bigcup_{j=1}^{\infty} G_j\right) \geq \mu^*\left(\bigcup_{j=1}^{n} G_j\right) = \sum_{j=1}^{n} \mu^*(G_j) \quad \text{für alle } n \in \mathbb{N},$$

oder

$$\mu^*\left(\bigcup_{j=1}^{\infty} G_j\right) \geq \sum_{j=1}^{\infty} \mu^*(G_j) \geq \mu^*\left(\bigcup_{j=1}^{\infty} G_j\right)$$

wegen Satz 6.10(4), d.h. μ^* ist ein Prämaß auf $\mathfrak{A}$.

2. Schritt: $\mathfrak{A}$ ist eine σ–Algebra.

Da $\mathfrak{A}$ als Algebra $\cap$–stabil ist, genügt es zu zeigen, dass $\mathfrak{A}$ ein Dynkin-System ist (vgl. Satz 5.12). Es seien also $G_1, G_2, \ldots \in \mathfrak{A}$ paarweise disjunkt, und $U \subseteq \Omega$ beliebig. Setze $G := \bigcup_{j=1}^{\infty} G_j$. Da $\mathfrak{A}$ eine Algebra ist, erhält man wegen (6.2), (6.6) und Satz 6.10(3)

$$\begin{aligned}
\mu^*(U) &= \mu^*\left(U \cap \bigcup_{j=1}^{n} G_j\right) + \mu^*\left(U \cap \overline{\bigcup_{j=1}^{n} G_j}\right) \\
&\geq \mu^*\left(U \cap \bigcup_{j=1}^{n} G_j\right) + \mu^*(U \cap \overline{G}) = \sum_{j=1}^{n} \mu^*(U \cap G_j) + \mu^*(U \cap \overline{G})
\end{aligned}$$

für alle $n \in \mathbb{N}$. Mit Hilfe von Satz 6.10, Eigenschaften (3) und (4), gilt schließlich:

$$\mu^*(U) \geq \sum_{j=1}^{\infty} \mu^*(U \cap G_j) + \mu^*(U \cap \overline{G}) \geq \mu^*(U \cap G) + \mu^*(U \cap \overline{G}) \geq \mu^*(U).$$

Also

$$G \in \mathfrak{A} \Rightarrow \mathfrak{A} \text{ ist ein Dynkin-System .}$$

■

Wenn nun μ^* das von μ induzierte äußere Maß ist, dann muss noch gezeigt werden, dass allle $A \in S$ μ^*–messbar sind. In diesem Falle ist $S \subseteq \mathfrak{U}$. Hieraus folgt $\sigma(S) \subseteq \mathfrak{U}$, und damit ist μ^* eingeschränkt auf $\sigma(S)$ ein Maß, das μ fortsetzt. Im Allgemeinen ist $\mathfrak{U}$ jedoch größer als $\sigma(S)$. In diesem Zusammenhang wird der folgende Satz gezeigt:

6.15 Satz:
Wenn μ^* das von μ induzierte äußere Maß ist, dann sind alle $B \in \mathcal{R} = \mathcal{R}(S)$ μ^*–messbar.

Beweis:
Es seien $B \in \mathcal{R}$ und $U \subset \Omega$ beliebig. Wegen Satz 6.10, Eigenschaften (1) und (4), gilt immer:

$$\mu^*(U) \leq \mu^*(U \cap B) + \mu^*(U \cap \overline{B}).$$

Ist $\widehat{S}(U) = \varnothing$, so folgt $\mu^*(U) = \infty$, und alles ist gezeigt. Es sei nun $(A_n)_{n \in \mathbb{N}} \in \widehat{S}(U)$. Aus $B \in \mathcal{R}$ ergibt sich $B = \bigcup_{i=1}^{n} C_i$, $C_i \in S$, $C_{i_1} \cap C_{i_2} = \varnothing$ für $i_1 \neq i_2$. Aus $A_j \backslash B \in \mathcal{R}$ folgt: $A_j \backslash B = \bigcup_{i=1}^{n_j} D_{j,i}$ mit $D_{j,i} \in S$, $D_{j,i_1} \cap D_{j,i_2} = \varnothing$ für $i_1 \neq i_2$, $j \in \mathbb{N}$. Somit folgt $\Delta_1 := (A_j \cap C_i)_{j \in \mathbb{N}, 1 \leq i \leq n} \in \widehat{S}(U \cap B)$ und $\Delta_2 := (D_{j,i})_{j \in \mathbb{N}, 1 \leq i \leq n_j} \in \widehat{S}(U \backslash B) = \widehat{S}(U \cap \overline{B})$. Da μ additiv auf $\mathcal{R}$ ist, ergibt sich

$$\mu(A_j) = \mu(A_j \cap B) + \mu(A_j \backslash B) = \sum_{i=1}^{n} \mu(A_j \cap C_i) + \sum_{i=1}^{n_j} \mu(D_{j,i})$$

für alle $j \in \mathbb{N}$. Hieraus folgt wiederum

$$\sum_{j=1}^{\infty} \mu(A_j) = \sum_{E_1 \in \Delta_1} \mu(E_1) + \sum_{E_2 \in \Delta_2} \mu(E_2) \leq \mu^*(U \cap B) + \mu^*(U \cap \overline{B})$$

für jedes $(A_j)_{j \in \mathbb{N}} \in \widehat{S}(U)$. Daraus ergibt sich schließlich

$$\mu^*(U) \geq \mu^*(U \cap B) + \mu^*(U \cap \overline{B}).$$

Folglich ist B μ^*–messbar. ■

Eindeutigkeit

6.16 Definition (σ–endlich):
Es seien S ein Semiring über Ω und μ ein Prämaß (Inhalt) auf (Ω, S). $\mathfrak{M} \subseteq S$ sei ein Mengensystem über Ω. μ heißt $\underline{\sigma\text{–endlich}}$ in $\mathfrak{M}$, wenn es Mengen $A_1 \subseteq A_2 \subseteq \ldots \in \mathfrak{M}$ mit $\bigcup_{j=1}^{\infty} A_j = \Omega$ und $\mu(A_j) < \infty$, $j \in \mathbb{N}$, gibt.

6.17 Satz (Eindeutigkeitssatz für Maße):
Es sei $\mathfrak{M}$ ein $\cap$–stabiles System von Teilmengen von Ω. Sind μ_1 und μ_2 zwei Maße auf $\sigma(\mathfrak{M})$, die auf $\mathfrak{M}$ übereinstimmen und dort σ–endlich sind, so stimmen sie auch auf $\sigma(\mathfrak{M})$ überein.

Beweis:
Zu zeigen ist, dass für alle $M \in \sigma(\mathfrak{M})$ gilt: $\mu_1(M) = \mu_2(M)$. Für $E \in \mathfrak{M}$ mit $\mu_1(E) = \mu_2(E) < \infty$ wird $\mathfrak{D}_E := \{D \in \sigma(\mathfrak{M}) \mid \mu_1(E \cap D) = \mu_2(E \cap D)\}$ gesetzt.

Behauptung: $\mathfrak{D}_E$ ist ein Dynkin-System.

1. Wegen $\mu_1(E \cap \Omega) = \mu_1(E) = \mu_2(E) = \mu_2(E \cap \Omega)$ gilt $\Omega \in \mathfrak{D}_E$.
2. Es seien $A, B \in \mathfrak{D}_E$ mit $A \subseteq B$. Mit Satz 6.6 gilt:
$$\begin{aligned} \mu_1(E \cap (B \setminus A)) &= \mu_1((E \cap B) \setminus (E \cap A)) \\ &= \mu_1(E \cap B) - \mu_1(E \cap A) \\ &= \mu_2(E \cap B) - \mu_2(E \cap A) \\ &= \mu_2((E \cap B) \setminus (E \cap A)) \\ &= \mu_2(E \cap (B \setminus A)). \end{aligned}$$
Damit gilt also $\mu_1(E \cap (B \setminus A)) = \mu_2(E \cap (B \setminus A))$ bzw. $B \setminus A \in \mathfrak{D}_E$.
3. Es sei $(D_n)_{n\in\mathbb{N}}$ eine Folge paarweise disjunkter Elemente von $\mathfrak{D}_E$. Dann gilt:
$$\begin{aligned} \mu_1\left(E \cap \bigcup_{n\in\mathbb{N}} D_n\right) &= \mu_1\left(\bigcup_{n\in\mathbb{N}} (E \cap D_n)\right) \\ &= \sum_{n\in\mathbb{N}} \mu_1(E \cap D_n) \\ &= \sum_{n\in\mathbb{N}} \mu_2(E \cap D_n) \\ &= \mu_2\left(E \cap \bigcup_{n\in\mathbb{N}} D_n\right), \end{aligned}$$
also ist $\bigcup_{n\in\mathbb{N}} D_n \in \mathfrak{D}_E$.

Damit ist $\mathfrak{D}_E$ ein Dynkin-System.
Da mit $A, B \in \mathfrak{M}$ auch $A \cap B \in \mathfrak{M}$ gilt, folgt zunächst $\mathfrak{M} \subseteq \mathfrak{D}_E$. Somit gilt für das von $\mathfrak{M}$ erzeugte Dynkin-System $\mathfrak{D}(\mathfrak{M})$ die Beziehung $\mathfrak{D}(\mathfrak{M}) \subseteq \mathfrak{D}_E$. Nach Satz 5.12 folgt deshalb $\mathfrak{D}(\mathfrak{M}) = \mathfrak{D}_E = \sigma(\mathfrak{M})$ und es ergibt sich:
$$\mu_1(E \cap A) = \mu_2(E \cap A) \;\; \forall\, A \in \sigma(\mathfrak{M}) \quad \text{und } \forall\, E \in \mathfrak{M} \quad \text{mit } \mu_1(E) = \mu_2(E).$$
Aufgrund der σ–Endlichkeit von μ_1 und μ_2 existiert eine Folge $(A_n)_{n\in\mathbb{N}}$ von Mengen aus $\mathfrak{M}$ mit $\bigcup\limits_{n\in\mathbb{N}} A_n = \Omega$ und $\mu_1(A_n) = \mu_2(A_n) < \infty \;\; \forall\, n \in \mathbb{N}$. Wie oben ist
$$\mu_1(A_n \cap A) = \mu_2(A_n \cap A) \;\; \forall\, A \in \sigma(\mathfrak{M}),\; n \in \mathbb{N}.$$
Es wird nun $B_1 := A_1,\; B_2 := A_2 \setminus A_1, \ldots, B_n := A_n \setminus (A_1 \cup \ldots \cup A_{n-1})$ gesetzt, so dass $(B_n)_{n\in\mathbb{N}}$ eine Folge von paarweise fremden Mengen aus $\sigma(\mathfrak{M})$ wird. Man beachte, dass $B_n \in \sigma(\mathfrak{M})$; $B_n \subseteq A_n \;\; \forall\, n \in \mathbb{N}$ und $\bigcup\limits_{n=1}^{\infty} B_n = \bigcup\limits_{n=1}^{\infty} A_n = \Omega$ gilt. Damit ergibt sich
$$\mu_1(B_n \cap A) = \mu_1(A_n \cap (B_n \cap A)) = \mu_2(A_n \cap (B_n \cap A)) = \mu_2(B_n \cap A) \;\; \forall\, A \in \sigma(\mathfrak{M}),\; n \in \mathbb{N}.$$
Da $A = \bigcup\limits_{n=1}^{\infty} (B_n \cap A)$ ist, folgt aus der σ–Additivität von μ_1 und μ_2:
$$\mu_1(A) = \sum_{n=1}^{\infty} \mu_1(B_n \cap A) = \sum_{n=1}^{\infty} \mu_2(B_n \cap A) = \mu_2(A) \;\; \forall\, A \in \sigma(\mathfrak{M}). \qquad \blacksquare$$

Damit ist nun der zweite Maßfortsetzungssatz 6.18 bewiesen:

6.18 Satz (2. Fortsetzungssatz):
Es sei μ ein Prämaß auf einem Semiring S über Ω. Dann ist

$$\mu^*(A) := \inf\left\{\sum_{n=1}^{\infty}\mu(A_n) \mid (A_n)_{n\in\mathbb{N}} \in \widehat{S}(A)\right\}, \quad A \in \sigma(S),$$

ein Maß auf $\sigma(S)$, das μ fortsetzt. Ist μ σ–endlich auf S, so ist μ^* die einzige Fortsetzung von μ zu einem Maß auf $\sigma(S)$.

Borel-Maß

Im Hinblick auf die Konstruktion von Wahrscheinlichkeitsmaßen auf der σ–Algebra der Borelschen Mengen $\mathfrak{B}^n$ im nächsten Kapitel wird nun der Begriff des Borel–Maßes eingeführt.

6.19 Definition (Borel–Maße):
Ein Maß μ auf $(\mathbb{R}^n, \mathfrak{B}^n)$, für welches $\mu(K) < \infty$ für jedes kompakte $K \subset \mathbb{R}^n$ ist, heißt ein Borel–Maß auf $(\mathbb{R}^n, \mathfrak{B}^n)$.
(Siehe auch Lebensdaten von Borel im Anhang D.)

6.20 Lemma:
μ ist genau dann ein Borel–Maß auf $(\mathbb{R}^n, \mathfrak{B}^n)$, wenn es auf $\mathbb{I}^n$ endlich ist.

Beweis:
Die Äquivalenz folgt aus dem Satz von Heine-Borel, der besagt, dass jedes kompakte $K \subset \mathbb{R}^n$ eine Teilmenge eines Intervalls $[a, b]$ für geeignet gewählte $a, b \in \mathbb{R}^n$ ist. ■

6.21 Satz:
Jedes Borel-Maß μ auf $(\mathbb{R}^n, \mathfrak{B}^n)$ ist eindeutig durch seine Werte auf $\mathbb{I}^n$ bestimmt.

Beweis:
Es wird Satz 6.17 angewandt: $\mathfrak{B}^n = \sigma(\mathbb{I}^n)$. $\mathbb{I}^n$ ist als Semiring $\cap$–stabil, μ ist endlich auf $\mathbb{I}^n$ und es gilt

$$\bigcup_{n=1}^{\infty}((-n, \ldots, -n), (n, \ldots, n)] = \mathbb{R}^n,$$

d.h. μ ist σ–endlich auf $\mathbb{I}^n$. ■

Literatur zu Kapitel 6

Folgende Bücher werden als begleitende Literatur zu diesem Kapitel des Skriptes empfohlen:

- H. Bauer:
  Wahrscheinlichkeitstheorie,
  5. Auflage, Walter de Gruyter, Berlin, 2002.
  ISBN: 3110172364

- W. Behnen, G. Neuhaus:
  Grundkurs Stochastik,
  3. Auflage, Teubner-Verlag, Stuttgart, 1995.
  ISBN: 3930737698

- P. Billingsley:
  Probability and Measure,
  2nd edition, John Wiley and Sons, New York, 1986.
  ISBN: 0471007102

- P. Gänssler und W. Stute:
  Wahrscheinlichkeitstheorie,
  Springer–Verlag, Berlin, 1977.
  ISBN: 3540084185

- E. Henze:
  Einführung in die Maßtheorie,
  Bibl. Institut, Mannheim, 1971.
  ISBN: 341100505X

- R. Mathar/D. Pfeiffer:
  Stochastik für Informatiker,
  Teubner, 1990.
  ISBN: 3519022400

- P. P. Spies:
  Grundlagen stochastischer Modelle,
  Hanser, München, 1982.
  ISBN: 3446137114

Kapitel 7

Maßdefinierende Funktionen

Als eine erste Anwendung der beiden Maß–Fortsetzungssätze wird die Konstruktion von Maßen und Wahrscheinlichkeitsmaßen auf der σ–Algebra der Borelschen Mengen $(\mathbb{R}, \mathfrak{B})$ behandelt. Grundlegend hierfür sind die Begriffe der maßdefinierenden Funktion und der Verteilungsfunktion. Anschließend werden Beispiele von Wahrscheinlichkeitsmaßen und Verteilungsfunktionen vorgestellt.

Schlüsselwörter: Maßdefinierende Funktion, Verteilungsfunktion, Korrespondenzsatz, Rechenregeln für maßdefinierende Funktionen und Verteilungsfunktionen, Lebesgue–Maß, Exponentialverteilung, Riemann–Dichte, Rechteck–Verteilung, Weibull–Verteilung, Normalverteilung, Standard–Normalverteilung, logarithmische Normalverteilung, χ^2–Verteilung, Cauchy–Verteilung, Gammaverteilung, Erlang–Verteilung, Betaverteilung

7.1 Korrespondenzsatz

Maße und Wahrscheinlichkeitsmaße über $(\mathbb{R}, \mathfrak{B})$ lassen sich durch spezielle reelle Funktionen erzeugen. Dabei sind solche Funktionen von besonderem Interesse, die die Existenz eines Maßes μ über $(\mathbb{R}, \mathfrak{B})$ mit der Eigenschaft

$$\mu((a,b]) = F(b) - F(a)$$

nach sich ziehen.

7.1 Definition (maßdefinierende Funktion, Verteilungsfunktion):
Eine Funktion $F\colon \mathbb{R} \to \mathbb{R}$ heißt maßdefinierende Funktion über $\mathbb{R}$, falls sie monoton nichtfallend und rechtsseitig stetig ist. Eine Funktion $F\colon \mathbb{R} \to \mathbb{R}$ heißt Verteilungsfunktion über $\mathbb{R}$, falls sie monoton nichtfallend, rechtsseitig stetig und normiert, d.h. $\lim\limits_{x\to-\infty} F(x) = 0$ und $\lim\limits_{x\to+\infty} F(x) = 1$, ist.

7.2 Satz:
Zu jeder maßdefinierenden Funktion $F\colon \mathbb{R} \to \mathbb{R}$ gibt es genau ein Maß μ_F über $(\mathbb{R}, \mathfrak{B})$ mit

$$\mu_F((a,b]) \;=\; F(b) - F(a) \qquad \forall\, a, b \in \mathbb{R} \;\; mit \;\; a \le b.$$

Ist F eine Verteilungsfunktion, dann ist μ_F ein Wahrscheinlichkeitsmaß, das mit P_F bezeichnet wird.

Beweis:
Zu zeigen sind:

(1.) Existenz von μ_F.

(2.) Ist F Verteilungsfunktion, so ist μ_F Wahrscheinlichkeitsmaß.

Zu (1.): Um die Existenz von μ_F zu zeigen, betrachten wir den Semiring $\mathbb{I}^1$ der links offenen und rechts abgeschlossenen Intervalle $(a, b]$, $a, b \in \mathbb{R}$ mit $a < b$, auf dem $\nu := \nu((a, b]) := F(b) - F(a)$ bekanntlich einen Inhalt definiert (siehe Beispiel 6.5).

(a) Ist ν σ–additiv und somit ein Prämaß, kann ν aufgrund des ersten Maßfortsetzungssatzes eindeutig zu einem Prämaß Ψ auf dem von $\mathbb{I}^1$ erzeugten Ring $\mathfrak{R}(\mathbb{I}^1)$ fortgesetzt werden.

(b) Ist ν σ–endlich, so existiert aufgrund des zweiten Maßfortsetzungssatzes eine eindeutige Fortsetzung von Ψ zu einem Maß μ_F auf der von $\mathbb{I}^1$ erzeugten σ–Algebra $\sigma(\mathbb{I}^1) = \mathfrak{B}^1$.

Zu (a): Beweis der σ–Additivität.
$(A_n)_{n\in\mathbb{N}}$ bezeichne eine Folge von paarweise fremden Mengen aus $\mathbb{I}^1$ mit der Eigenschaft $A = \bigcup_{n\in\mathbb{N}} A_n \in \mathbb{I}^1$. Mit ψ wird die eindeutige Fortsetzung von ν auf den von $\mathbb{I}^1$ erzeugten Ring $\mathfrak{R}(\mathbb{I}^1)$ bezeichnet. Es werden

$$\nu(A) \;=\; \nu\left(\bigcup_{n=1}^{\infty} A_n\right) \;\ge\; \sum_{n=1}^{\infty} \nu(A_n)$$

und

$$\nu(A) \; = \; \nu\left(\bigcup_{n=1}^{\infty} A_n\right) \; \leq \; \sum_{n=1}^{\infty} \nu(A_n)$$

gezeigt. Aufgrund der Additivität und der Monotonie von ψ gilt:

$$\begin{aligned}
\sum_{n=1}^{m} \nu(A_n) &= \sum_{n=1}^{m} \psi(A_n) && \text{da } \nu \text{ und } \psi \text{ auf } \mathbb{I}^1 \text{ übereinstimmen} \\
&= \psi\left(\bigcup_{n=1}^{m} A_n\right) && \text{da lediglich } \textstyle\bigcup_{n=1}^{m} A_n \in \mathfrak{R}(\mathbb{I}^1) \text{ vorausgesetzt} \\
& && \text{werden kann, muss } \psi \text{ anstelle von } \nu \\
& && \text{herangezogen werden} \\
&\leq \psi\left(\bigcup_{n=1}^{\infty} A_n\right) && \text{aufgrund der Monotonie von } \psi \\
&= \nu\left(\bigcup_{n=1}^{\infty} A_n\right) && \text{da } \textstyle\bigcup_{n=1}^{\infty} A_n \in \mathbb{I}^1 \text{ vorausgesetzt war und} \\
& && \nu \text{ und } \psi \text{ auf } \mathbb{I}^1 \text{ übereinstimmen}
\end{aligned}$$

Hieraus folgt

$$\sum_{n=1}^{\infty} \nu(A_n) \; = \; \lim_{m\to\infty} \sum_{n=1}^{m} \nu(A_n) \; \leq \; \nu\left(\bigcup_{n=1}^{\infty} A_n\right) \; = \; \nu(A).$$

Es bleibt $\nu(\bigcup_{n=1}^{\infty} A_n) \; \leq \; \sum_{n=1}^{\infty} \nu(A_n)$ zu zeigen.
Für $n = 1, 2, \ldots$ sei

$$A := \; (a, b], \qquad A_n := \; (a_n, b_n]$$

gesetzt und für beliebige $\delta > 0$ und $\delta_n > 0$, $n = 1, 2, \ldots$, sei

$$A' := \; (a + \delta, b], \qquad A'_n := \; (a_n, b_n + \delta_n].$$

Offensichtlich gilt:

$$A = \; (a, a + \delta] \cup (a + \delta, b], \qquad A'_n = \; (a_n, b_n] \cup (b_n, b_n + \delta_n]$$

und

$$\begin{aligned}
\nu(A) &= \nu((a, a + \delta]) + \nu((a + \delta, b]), & \nu(A'_n) &= \nu((a_n, b_n]) \\
& & &\quad +\nu((b_n, b_n + \delta_n]) \\
&= F(a + \delta) - F(a) + \nu(A') & &= \nu(A_n) + F(b_n + \delta_n) \\
& & &\quad -F(b_n).
\end{aligned}$$

Da F als rechtsseitig stetig vorausgesetzt war, gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$ und ein $\delta_n > 0$, $n = 1, 2, \ldots$, so dass gilt:

$$\nu(A) \leq \; \nu(A') + \tfrac{\varepsilon}{2}, \qquad \nu(A'_n) \leq \; \nu(A_n) + \tfrac{\varepsilon}{2^{n+1}} \; (n = 1, 2, \ldots). \quad (*)$$

Die Intervalle $(a_n, b_n + \delta_n)$ bilden eine offene Überdeckung des Intervalls $[a + \delta, b]$ und damit auch von $(a + \delta, b]$:

$$A' = (a + \delta, b] \subseteq A^* := [a + \delta, b] \subseteq \bigcup_{n=1}^{\infty} (a_n, b_n + \delta_n) \subseteq \bigcup_{n=1}^{\infty} A'_n. \tag{7.1}$$

Nach dem Überdeckungssatz von Heine–Borel reichen endlich viele der A'_n zur Überdeckung der Menge A' aus:

$$A' \subseteq \bigcup_{n=1}^{\infty} A'_n \subseteq \bigcup_{n=1}^{k} A'_n \quad \text{für ein } k \in \mathbb{N},$$

Damit folgt:

$$\begin{aligned}
\nu(A) &\leq \nu(A') + \frac{\varepsilon}{2} \\
&= \psi(A') + \frac{\varepsilon}{2} && \text{da } \nu = \psi \text{ auf } \mathbb{I}^1 \\
&\leq \psi\left(\bigcup_{n=1}^{k} A'_n\right) + \frac{\varepsilon}{2} && \text{aufgrund von (7.1)} \\
&\leq \sum_{n=1}^{k} \psi(A'_n) + \frac{\varepsilon}{2} && \text{da } \psi \text{ subadditiv ist} \\
&= \sum_{n=1}^{k} \nu(A'_n) + \frac{\varepsilon}{2} && \text{da } \nu = \psi \text{ auf } \mathbb{I}^1 \\
&\leq \sum_{n=1}^{k} \left(\nu(A_n) + \frac{\varepsilon}{2^{n+1}}\right) + \frac{\varepsilon}{2} && \text{aufgrund von } (*) \\
&\leq \sum_{n=1}^{\infty} \nu(A_n) + \varepsilon && \text{da } \sum_{n=1}^{\infty} \frac{\varepsilon}{2^{n+1}} = \frac{\varepsilon}{2}.
\end{aligned}$$

Da $\varepsilon > 0$ beliebig gewählt werden kann, folgt hieraus die Behauptung.

Zu (b): Für die σ–Endlichkeit von ν ist die Existenz einer Mengenfolge $(A_n)_{n\in\mathbb{N}}$ aus $\mathbb{I}^1$ mit $A_1 \subset A_2 \subseteq A_3 \subseteq \ldots$, $\bigcup_{n\in\mathbb{N}} A_n = \mathbb{R}$ und $\nu(A_n) < \infty$ für alle n nachzuweisen. Hierfür wird $A_n = (-n, n]$, $n \in \mathbb{N}$, gewählt. Offensichtlich gilt $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ und $\bigcup_{n\in\mathbb{N}} A_n = \mathbb{R}$. Außerdem ist $\nu((-n, n]) = F(n) - F(-n) < \infty$ für alle n. Es ist somit der zweite Fortsetzungssatz (Satz 6.18) anwendbar, d.h. es gibt ein eindeutig bestimmtes Maß $\mu_F\colon \sigma(\mathbb{I}^1) \to \mathbb{R} \cup \{\infty\}$, das ν fortsetzt und σ-endlich ist.

Zu (2.): Ist F Verteilungsfunktion, so ist μ_F Wahrscheinlichkeitsmaß.
Es wird gezeigt, dass aus den beiden Aussagen $\lim_{x\to-\infty} F(x) = 0$ und $\lim_{x\to+\infty} F(x) = 1$ die Eigenschaft $\mu_F(\mathbb{R}) = 1$ folgt. Dafür wird zunächst bewiesen, dass $F(x) = \mu_F((-\infty, x])$ für alle $x \in \mathbb{R}$ gilt. Da μ_F stetig von unten ist (siehe Satz 6.7a)), folgt:

$$\begin{aligned}
\mu_F((-\infty, x]) = \mu_F\Big(\bigcup_{n\in\mathbb{N}} (-n, x]\Big) &= \lim_{n\to\infty} \mu_F((-n, x]) = \lim_{n\to\infty} (F(x) - F(-n)) \\
&= F(x) - \lim_{n\to\infty} F(-n) = F(x) - 0 = F(x).
\end{aligned}$$

Aufgrund der Darstellung $\mathbb{R} = \bigcup_{n\in\mathbb{N}} (-\infty, n]$ und der Tatsache, dass μ_F stetig von unten ist (siehe Satz 6.7a)), folgt

$$\mu_F(\mathbb{R}) \;=\; \mu_F\Big(\bigcup_{n\in\mathbb{N}} (-\infty, n]\Big) \;=\; \lim_{n\to\infty} \mu_F(-\infty, n] \;=\; \lim_{n\to\infty} F(n) \;=\; 1.$$

■

Die Verteilungsfunktion F wird durch die im folgenden Satz beschriebenen Eigenschaften gekennzeichnet:

7.3 Satz:
Gegeben sei ein Wahrscheinlichkeitsmaß P über $(\mathbb{R}, \mathfrak{B})$ und eine Funktion $F: \mathbb{R} \to \mathbb{R}$ mit der Eigenschaft $F(x) = P((-\infty, x])$ für alle $x \in \mathbb{R}$. Dann gilt:

a) *F ist monoton nichtfallend,*

b) *F ist rechtsseitig stetig,*

c) $\lim\limits_{x \to -\infty} F(x) = 0$ *und* $\lim\limits_{x \to +\infty} F(x) = 1$,

d.h. F ist eine Verteilungsfunktion.

Beweis:
Die Aussage a) folgt unmittelbar aus der Annahme, dass P monoton ist (vgl. Satz 7.2). Für den Nachweis von b) muss noch gezeigt werden, dass für jede monoton fallende Folge $(x_n)_{n\in\mathbb{N}}$ mit $\lim_{n\to\infty} x_n = 0$ auch $\lim_{n\to\infty}[F(x + x_n) - F(x)] = 0$ für alle $x \in \mathbb{R}$ gilt. Da P stetig von oben ist (siehe Satz 6.7b)), gilt aber:

$$\lim_{n\to\infty}[F(x + x_n) - F(x)] = \lim_{n\to\infty} P((x, x + x_n]) = P\Big(\bigcap_{n=1}^{\infty}(x, x + x_n]\Big) = P(\varnothing) = 0.$$

Die Aussage c) lässt sich ebenfalls mit Hilfe von Satz 6.7 beweisen:

$$\begin{aligned} 0 &= P(\varnothing) = P\left(\bigcap_{n\in\mathbb{N}}(-\infty, -n]\right) = \lim_{n\to\infty} P((-\infty, -n]) = \lim_{n\to\infty} F(-n) \\ 1 &= P(\Omega) = P\left(\bigcup_{n\in\mathbb{N}}(-\infty, n]\right) = \lim_{n\to\infty} P((-\infty, n]) = \lim_{n\to\infty} F(n). \end{aligned}$$

■

7.4 Bemerkung:
Die Kernaussage dieses Kapitels liegt in der Existenz einer bijektiven Abbildung von Verteilungsfunktionen über $\mathbb{R}$ und Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathfrak{B})$. Dieser Sachverhalt wird als Korrespondenzsatz bezeichnet.

7.2 Rechenregeln für maßdefinierende Funktionen

7.5 Satz (Rechenregeln für maßdefinierende Funktionen):
Es sei F eine maßdefinierende Funktion über $\mathbb{R}$ und μ_F das korrespondierende Maß auf $(\mathbb{R}, \mathfrak{B})$, dann gilt für alle $a, b \in \mathbb{R}$ mit $a < b$:

a) $\mu_F((a, b]) = F(b) - F(a)$,

b) $\mu_F((a, b)) = F(b - 0) - F(a)$,

c) $\mu_F([a, b]) = F(b) - F(a - 0)$,

d) $\mu_F(\{a\}) = F(a) - F(a - 0)$,

e) $\mu_F([a, b)) = F(b - 0) - F(a - 0)$,

wobei mit $F(x - 0)$ der linksseitige Grenzwert von F an der Stelle x bezeichnet werde (entsprechend $F(x + 0)$).
Ist F eine Verteilungsfunktion über $\mathbb{R}$ und bezeichnet P_F das zugehörige Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathfrak{B})$, dann gilt für alle $x \in \mathbb{R}$:

f) $P_F((-\infty, x]) \;=\; F(x),$

g) $P_F((-\infty, x)) \;=\; F(x-0),$

h) $P_F((x, \infty)) \;=\; 1 - F(x).$

Beweis:

a) Diese Behauptung ist gleichbedeutend mit der Aussage von Satz 7.2.

b)

$$\mu_F((a,b)) = \mu_F\left(\bigcup_{n=1}^{\infty}\left(a, b-\frac{1}{n}\right]\right) \overset{\text{Satz 6.7a)}}{=} \lim_{n\to\infty} \mu_F\left(\left(a, b-\frac{1}{n}\right]\right)$$
$$= \lim_{n\to\infty} F\left(b-\frac{1}{n}\right) - F(a) = F(b-0) - F(a).$$

c)

$$\begin{aligned}\mu_F([a,b]) &= \mu_F\left(\bigcap_{n=1}^{\infty}\left(a-\frac{1}{n}, b\right]\right)\\ &= \lim_{n\to\infty} \mu_F\left(\left(a-\frac{1}{n}, b\right]\right) \qquad (\mu_F \text{ ist stetig von oben (siehe Satz 6.7b)))}\\ &= \lim_{n\to\infty}\left(F(b) - F\left(a-\frac{1}{n}\right)\right)\\ &= F(b) - F(a-0).\end{aligned}$$

d)

$$\mu_F(\{a\}) = \mu_F\left(\bigcap_{n=1}^{\infty}\left(a-\frac{1}{n}, a\right]\right) = \lim_{n\to\infty} \mu_F\left(\left(a-\frac{1}{n}, a\right]\right) = \lim_{n\to\infty}\left(F(a) - F\left(a-\frac{1}{n}\right)\right)$$
$$= F(a) - F(a-0).$$

e) $\mu_F([a,b)) = \mu_F((a,b) \cup \{a\}) = \mu_F((a,b)) + \mu_F(\{a\}) = F(b-0) - F(a-0).$

f) Wurde bereits unter (iv) im Beweis von Satz 7.2 gezeigt.

g) $P_F((-\infty, x)) = P_F((-\infty, x]\backslash\{x\}) = P_F((-\infty, x]) - P_F(\{x\}) = F(x) - (F(x) - F(x-0)) = F(x-0).$

h) $P_F((x,\infty)) = P_F(\mathbb{R}\backslash(-\infty, x]) = P_F(\mathbb{R}) - P_F((-\infty, x]) = 1 - F(x).$ ■

7.6 Beispiel:
Die Funktion

$$F(x) := \begin{cases} 0\;, & x < 0, \\ \frac{1}{8}\;, & 0 \le x < 1, \\ \frac{4}{8}\;, & 1 \le x < 2, \\ \frac{7}{8}\;, & 2 \le x < 3, \\ 1\;, & 3 \le x \end{cases}$$

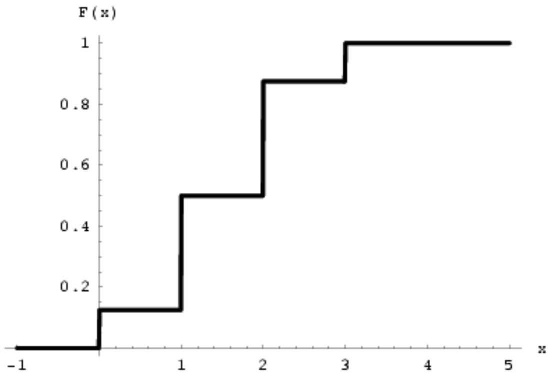


Abbildung 7.1: Darstellung der Verteilungsfunktion $F(x)$

definiert eine Verteilungsfunktion über $\mathbb{R}$.
Man verifiziert leicht:

$$\begin{aligned}
P_F((-1,5]) &= F(5) - F(-1) = 1 - 0 = 1.\\
P_F((0.5,2]) &= F(2) - F(0.5) = \frac{7}{8} - \frac{1}{8} = \frac{3}{4}.\\
P_F([1.5,3)) &= F(3-0) - F(1.5-0) = \frac{7}{8} - \frac{4}{8} = \frac{3}{8}.\\
P_F((0,1)) &= F(1-0) - F(0) = \frac{1}{8} - \frac{1}{8} = 0.\\
P_F(\{2\}) &= F(2) - F(2-0) = \frac{7}{8} - \frac{4}{8} = \frac{3}{8}.\\
P_F((-\infty,1.5]) &= F(1.5) = \frac{4}{8} = \frac{1}{2}.\\
P_F((1,\infty)) &= 1 - F(1) = \frac{4}{8} = \frac{1}{2}.
\end{aligned}$$

(Siehe auch Mathematica-Notebook zu diesem Beispiel und PowerPoint-Präsentation zu diesem Beispiel.)

7.3 Beispiele für maßdefinierende Funktionen

Wurden die diskreten Wahrscheinlichkeitsmaße in Kapitel 2 durch die Summe ihrer Wahrscheinlichkeiten beschrieben, so kennzeichnet die stetigen Wahrscheinlichkeitsmaße auf $\mathfrak{B}$ eine Dichte bezüglich des Lebesgue Maßes.

Das Lebesgue-Maß λ auf $(\mathbb{R}, \mathfrak{B})$ ist dabei das mit der maßdefinierenden Funktion

$$F(x) := x \qquad \forall\, x \in \mathbb{R}$$

korrespondierende Maß $\lambda = \lambda_F$ über $\mathfrak{B}$. Es ist die Fortsetzung des elementargeometrischen Inhalts

$$\lambda((a,b]) = F(b) - F(a) = b - a, \qquad \forall\, a,b \in \mathbb{R},\ a \leq b.$$

(Siehe auch Lebensdaten von Lebesgue im Anhang D.)

7.7 Definition (Dichte):
Man bezeichnet eine uneigentlich Riemann-integrierbare Funktion $f\colon \mathbb{R} \to \mathbb{R}^+$ mit der Eigenschaft

$$\int_{\mathbb{R}} f(t)\ dt = 1$$

als Dichte. (Siehe auch Lebensdaten von Riemann im Anhang D.)

Um einen Zusammenhang zwischen dem Wahrscheinlichkeitsmaß P und der Dichte f herzustellen, wird die Verteilungsfunktion F betrachtet:
Da die Verteilungsfunktion monoton ist, ist sie (λ– fast überall) differenzierbar (siehe Definition 9.35). Setze

$$f := \frac{d}{dx} F.$$

Mit den Ergebnissen der vorhergehenden Kapitel folgt dann

$$P_F((-\infty, x]) = F(x) = \int_{-\infty}^{x} f(t)\ dt, \quad x \in \mathbb{R}$$

und

$$P_F((a, b]) = F(b) - F(a) = \int_a^b f(t)\ dt, \quad a, b \in \mathbb{R},\ a \leq b.$$

Es lässt sich zeigen, dass die Ableitung der Verteilungsfunktion eine Dichte ist. Damit kann das nachfolgende Ergebnis formuliert werden:

Es sei $f\colon \mathbb{R} \to \mathbb{R}^+$ eine Dichte. Dann wird durch

$$F(x) := \int_{-\infty}^{x} f(t)\ dt, \quad x \in \mathbb{R}$$

eine Verteilungsfunktion F über $\mathbb{R}$ und damit ein Wahrscheinlichkeitsmaß P_F auf $(\mathbb{R}, \mathfrak{B})$ definiert. Die Funktion f wird auch Dichte der Wahrscheinlichkeitsverteilung P_F genannt.

Es werden nun einige Beispiele für maßdefinierende Funktionen bzw. stetige Wahrscheinlichkeitsmaße und ihre korrespondierenden Verteilungsfunktionen und Dichten angegeben:

Die Rechteckverteilung

Für jedes Paar $a, b \in \mathbb{R}$ mit $a < b$ wird durch

$$f(x) := \begin{cases} \frac{1}{b-a} & , \quad x \in [a, b] \\ 0 & , \quad x \notin [a, b], \end{cases}$$

eine Riemann–Dichte über $\mathbb{R}$ definiert. Das zugehörige Wahrscheinlichkeitsmaß heißt Rechteck(a, b)–Verteilung oder Gleichverteilung auf $[a, b]$, kurz $R(a, b)$. Die zugehörige Verteilungsfunktion lautet:

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

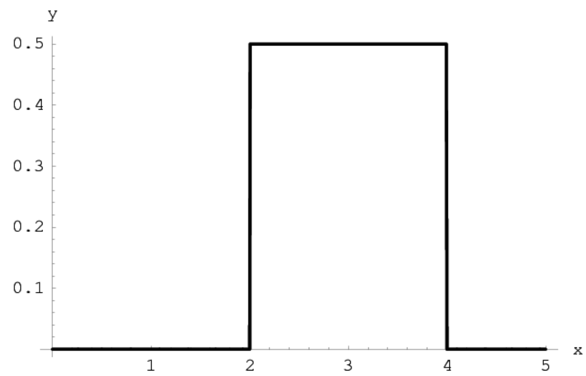


Abbildung 7.2: Dichtefunktion der Rechteckverteilung mit $a := 2$ und $b := 4$.

Es ist sofort ersichtlich, dass

$$\int_{\mathbb{R}} f(x)\ dx = \int_a^b \frac{1}{b-a}\ dx = \frac{1}{b-a} \int_a^b 1\ dx = 1$$

gilt, sowie dass $F(-\infty) = 0$ und $F(\infty) = 1$ ist.

Die Rechteck–Verteilung spielt bei der Erzeugung von Zufallszahlen und der Simulation stochastischer Prozesse eine wichtige Rolle.

Die Exponential–Verteilung

Es sei $\lambda > 0$. Die Funktion $F\colon \mathbb{R} \to \mathbb{R}$ mit

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

definiert eine Verteilungsfunktion über $\mathbb{R}$. Man nennt F Exponentialverteilung mit dem Parameter λ. Für die Exponentialverteilung mit dem Parameter λ verwenden wir das Symbol $\mathrm{Exp}(\lambda)$.

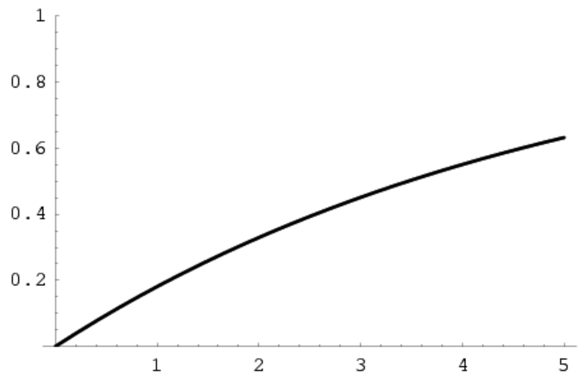


Abbildung 7.3: Verteilungsfunktion der Exponentialverteilung mit $\lambda := 0.2$.

Man beachte, dass F stetig und an allen Stellen $x \neq 0$ differenzierbar ist:

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

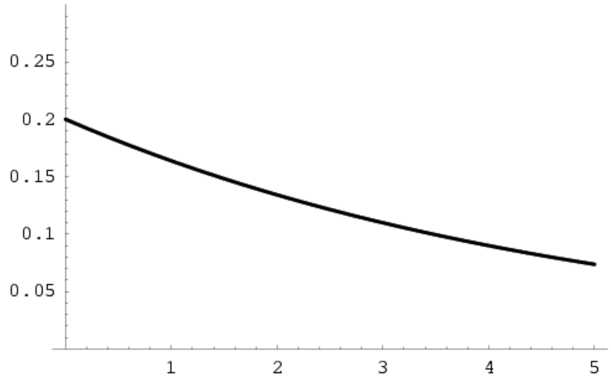


Abbildung 7.4: Dichtefunktion der Exponentialverteilung mit $\lambda := 0.2$.

Der Nachweis von $\int_{\mathbb{R}} f(x)\ dx = 1$ erfolgt durch Nachrechnen:

$$\int_{\mathbb{R}} f(x)\ dx = \int_0^\infty \lambda e^{-\lambda x}\ dx = \lambda \left[-\frac{1}{\lambda} e^{-\lambda x} \right]_0^\infty = 0 - (-1) = 1.$$

Diese Verteilung eignet sich besonders zur Modellierung von Lebensdauern, wie etwa dem Ausfallverhalten einer Maschine oder der Dauer eines Telefonats im Callcenter.

Weibull–Verteilung

Für $\lambda, \beta > 0$ definiert

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

eine Wahrscheinlichkeitsdichte über $\mathbb{R}$.

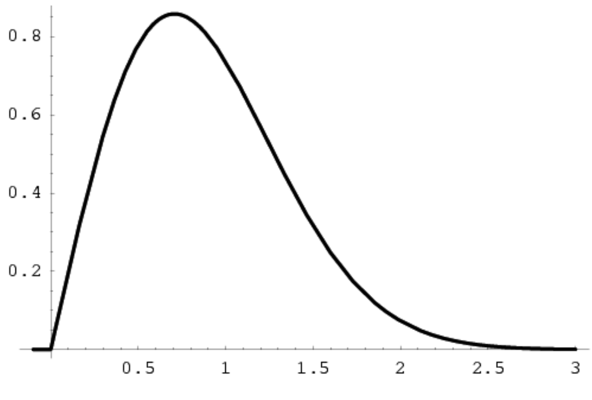


Abbildung 7.5: Dichtefunktion der Weibullverteilung mit $\lambda := 1$ und $\beta := 2$.

Integration liefert zunächst $\int f(x)\ dx = -e^{-\lambda x^\beta} + c =: F(x)$ für $x \geq 0$. Das Einsetzen der Bedingung $F(\infty) = 1$ ergibt $c = 1$. Daher lautet die zu der Dichte $f(x)$ gehörende Verteilungsfunktion:

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

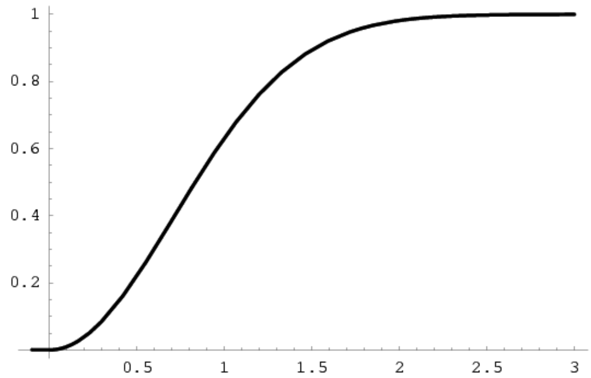


Abbildung 7.6: Verteilungsfunktion der Weibullverteilung mit $\lambda := 1$ und $\beta := 2$.

Damit ergibt sich auch für die Bedingung $\int_{\mathbb{R}} f(x)\ dx = 1$:

$$\int_{\mathbb{R}} f(x)\ dx = \int_0^{\infty} \lambda \cdot \beta \cdot x^{\beta-1} \cdot e^{-\lambda x^{\beta}}\ dx = \left[-e^{-\lambda x^{\beta}}\right]_0^{\infty} = 0 - (-1) = 1.$$

Das zu der Verteilungsfunktion korrespondierende Wahrscheinlichkeitsmaß P_F über $(\mathbb{R}, \mathfrak{B})$ heißt Weibull–Verteilung mit den Parametern λ, β. Für den Fall $\beta = 1$ ergibt sich die Exponentialverteilung.

Die Weibull–Verteilung findet Anwendung in der Zuverlässigkeitstheorie, etwa bei der Modellierung von Lebensdauern bei Geräten mit Abnutzungserscheinungen.

Die Normalverteilung

Die Funktion

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

wird als Gaußsche Glockenkurve mit den Parametern $\mu \in \mathbb{R}$ und $\sigma > 0$ bezeichnet und definiert eine Wahrscheinlichkeitsdichte über $\mathbb{R}$. Die korrespondierende Verteilungsfunktion lautet

$$F(x) := \int\limits_{-\infty}^{x} f(t)dt = \int\limits_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt, \quad x \in \mathbb{R}.$$

Die zu $F(x)$ gehörende Verteilung P_F auf $(\mathbb{R}, \mathfrak{B})$ wird Normalverteilung mit den Parametern μ und σ genannt. Im Fall $\mu = 0$ und $\sigma = 1$ spricht man von der sogenannten Standard–Normalverteilung. Für die Normalverteilung verwendet man das Symbol $\mathcal{N}(\mu, \sigma)$. Im Fall der Standard–Normalverteilung verwendet man anstelle von $F(x)$ das Symbol $\Phi(x)$ und anstelle von $f(x)$ das Symbol $\varphi(x)$.

Um die Werte der Normalverteilung zu berechnen, genügt es, die Standard–Normalverteilung

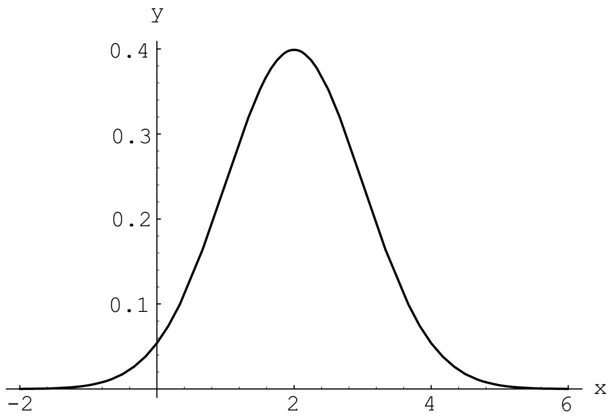


Abbildung 7.7: Dichtefunktion der Normalverteilung mit Lokalisationsparameter $\mu := 2$ und Streuparameter $\sigma := 1$.

zu kennen, denn mit der Substitution $y = (t - \mu)/\sigma$ und $\sigma \cdot dy = dt$ erhält man

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t)\ dt \ = \ \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2}\ dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-y^2/2}\ dy \ = \ \int_{-\infty}^{\frac{x-\mu}{\sigma}} \varphi(y)\ dy \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R}. \end{aligned}$$

Dies ist der Grund, warum in Statistik–Büchern lediglich die Standard–Normalverteilung tabelliert ist. Für den Nachweis von

$$\int_{-\infty}^{+\infty} \varphi(x)\ dx = 1$$

benutzt man die Beziehung

$$\begin{aligned} \left(\sqrt{2\pi}\int_{-\infty}^{+\infty} \varphi(x)\ dx\right)^2 &= \left(\int_{-\infty}^{+\infty} \exp(-t^2/2)\ dt\right)^2 \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp(-x^2/2)\exp(-y^2/2)\ dx\ dy \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2+y^2}{2}\right)\ dx\ dy, \end{aligned}$$

die mit Hilfe von Polarkoordinaten, d.h. mit Hilfe der Substitution $dx\ dy = r\ d\vartheta\ dr$, in

$$\int_0^{2\pi}\int_0^{\infty} \exp(-r^2/2)r\ dr\ d\vartheta \ = \ 2\pi\left[-\exp(-r^2/2)\right]\Big|_0^{\infty} \ = \ 2\pi$$

überführt werden kann. Damit gilt also:

$$\left(\sqrt{2\pi}\int_{-\infty}^{+\infty}\varphi(x)\ dx\right)^2 = 2\pi \quad \text{bzw.} \quad \int_{-\infty}^{+\infty}\varphi(x)\ dx = 1.$$

Die Normalverteilung erhält ihre Bedeutung hauptsächlich aus dem „zentralen Grenzwertsatz", der besagt, dass Summen aus unabhängigen aber identisch verteilten Zufallsvariablen gegen die Normalverteilung konvergieren (vgl. Kapitel 14.3 in Stochastik II).

Logarithmische Normalverteilung

Für $\mu \in \mathbb{R}$ und $\sigma > 0$ definiert auch

$$f(x) := \begin{cases} \dfrac{1}{\sigma x\sqrt{2\pi}}e^{-(\log(x)-\mu)^2/2\sigma^2} & ,\ x > 0 \\ 0 & ,\ x \le 0 \end{cases}$$

eine Wahrscheinlichkeitsdichte über $\mathbb{R}$. Das zugehörige Wahrscheinlichkeitsmaß über $(\mathbb{R}, \mathfrak{B})$ heißt logarithmische Normalverteilung.
Die logarithmische Normalverteilung kann auf die Standardnormalverteilung zurückgeführt werden. Mit der Substitution $y = (\log(t) - \mu)/\sigma$ und $dy = \sigma t \cdot dt$ ergibt sich:

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t)\ dt \ = \ \int_0^x \frac{1}{\sigma t\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^2}\ dt \\ &= \int_0^{\frac{\log(x)-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\ dy \\ &= \Phi\left(\frac{\log(x)-\mu}{\sigma}\right), \quad x \in \mathbb{R}. \end{aligned}$$

Damit folgt auch sofort $\int_{\mathbb{R}} f(x)\ dx = F(\infty) = \lim_{x\to\infty}\Phi\left(\frac{\log(x)-\mu}{\sigma}\right) = 1$.
Die logarithmische Normalverteilung wird als Modellverteilung bei Lebensdauer– und Festigkeitsproblemen eingesetzt.

Die Cauchy–Verteilung

Für $\lambda > 0$ und $\mu \in \mathbb{R}$ definiert

$$f(x) := \frac{1}{\pi}\cdot\frac{\lambda}{\lambda^2 + (x-\mu)^2}, \quad x \in \mathbb{R},$$

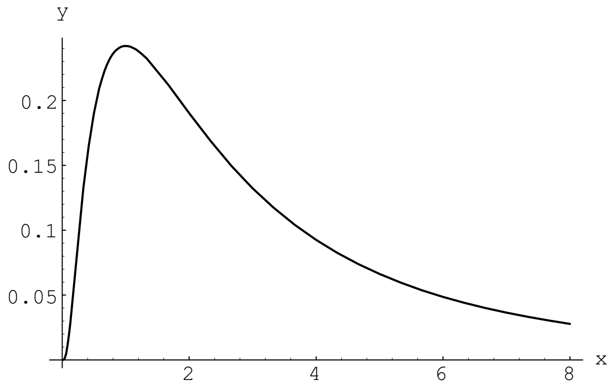


Abbildung 7.8: Dichtefunktion der logarithmischen Normalverteilung mit Parametern $\mu := 1$ und $\sigma := 1$.

eine Wahrscheinlichkeitsdichte über $\mathbb{R}$, denn es ist

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^{\infty} \frac{1}{\lambda\pi} \cdot \frac{1}{1+\left(\frac{x-\mu}{\lambda}\right)^2}dx \qquad \left(z := \frac{x-\mu}{\lambda}, \quad dz = \frac{1}{\lambda}dx\right) \\
&= \int_{-\infty}^{\infty} \frac{1}{\pi}\frac{1}{1+z^2}dz \\
&= \frac{1}{\pi}\left(\int_{-\infty}^{0} \frac{1}{1+z^2}dz + \int_{0}^{\infty} \frac{1}{1+z^2}dz\right) \\
&= \frac{1}{\pi}\left(\lim_{a\to-\infty}\int_{a}^{0} \frac{1}{1+z^2}dz + \lim_{b\to+\infty}\int_{0}^{b} \frac{1}{1+z^2}dz\right) \\
&= \frac{1}{\pi}\left(\lim_{a\to-\infty}(-\arctan a) + \lim_{b\to+\infty}(\arctan b)\right) \\
&= \frac{1}{\pi}\left(\frac{\pi}{2}+\frac{\pi}{2}\right) \;=\; 1.
\end{aligned}$$

Die zu f gehörende Verteilungfunktion heißt Cauchy–Verteilung.

Die Gammaverteilung

Es seien $b, p \in \mathbb{R}^+$. Das zur Dichte

$$f(x) := \begin{cases} \dfrac{b^p}{\Gamma(p)} x^{p-1}e^{-bx} & , \ x \geq 0 \\ 0 & , \ x < 0 \end{cases}$$

gehörende Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathfrak{B})$ heißt Gammaverteilung mit den Parametern b und p, kurz Gamma(b, p). Dabei sei $\Gamma(p)$ die Gamma-Funktion an der Stelle p:

$$\Gamma(p) := \int_0^\infty x^{p-1}e^{-x}dx.$$

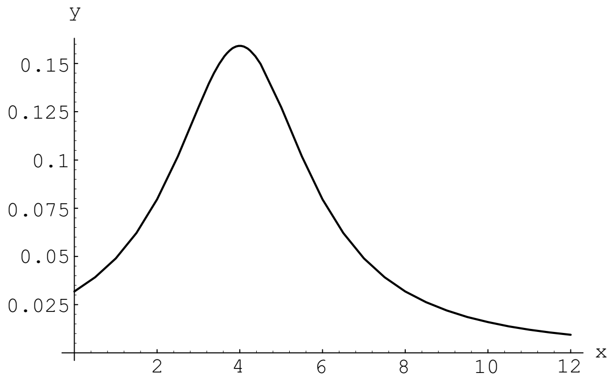


Abbildung 7.9: Dichtefunktion der Cauchy–Verteilung mit Lokalisationsparameter $\mu := 4$ und Streuparameter $\lambda := 2$.

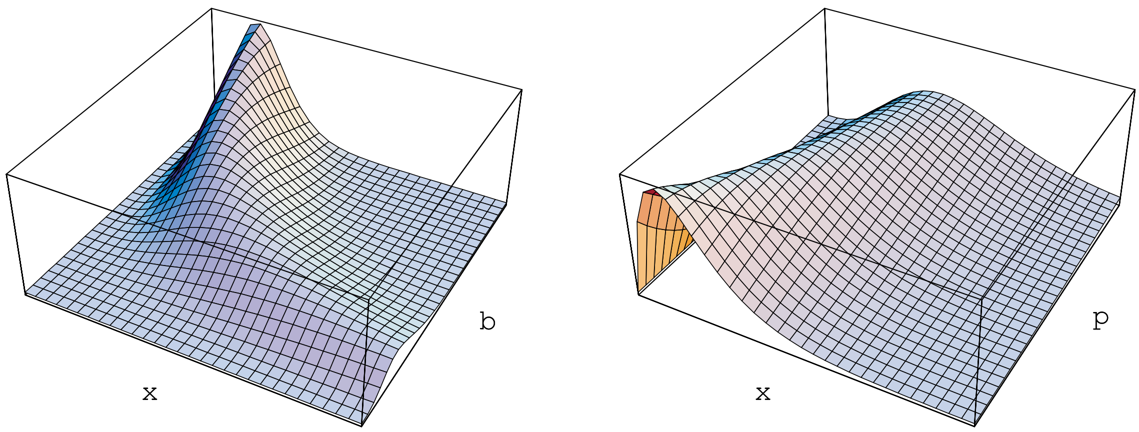


Abbildung 7.10: Dichtefunktion der Gammaverteilung mit variablem Parameter b bei konstantem p (links) und entsprechend umgekehrt (rechts).

Die Tatsache, dass $\int_{-\infty}^{+\infty} f(x)\ dx = 1$ ist, folgt aus der Beziehung

$$\begin{aligned}
\Gamma(p) \cdot \int_{-\infty}^{\infty} f(x)\ dx &= \int_{0}^{\infty} b^p x^{p-1} e^{-bx}\ dx \\
&= \int_{0}^{\infty} (bx)^{p-1} e^{-bx} b\ dx \quad (z := bx \text{ und } dz = b\ dx \text{ gesetzt}) \\
&= \int_{0}^{\infty} z^{p-1} e^{-z}\ dz \ = \ \Gamma(p).
\end{aligned}$$

Für die Gammaverteilung wird das Kürzel $Gamma(b, p)$ verwendet. Die Gammaverteilung wird unter anderem als Modellverteilung in der Zuverlässigkeitstheorie und der Warteschlangentheorie verwendet. Als Spezialfälle der Gammaverteilung ergeben sich die χ^2–Verteilung und die Erlang-Verteilung (siehe unten).

Die χ^2–Verteilung

Das Wahrscheinlichkeitsmaß P mit der Dichte

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

heißt χ^2–Verteilung mit n Freiheitsgraden, $n \in \mathbb{N}$. Die χ^2–Verteilung ergibt sich aus der Gammaverteilung, indem $p := \frac{n}{2}$ und $b := \frac{1}{2}$ gesetzt wird. (Damit folgt auch sofort $\int_{\mathbb{R}} f(x)\, dx = 1$.)

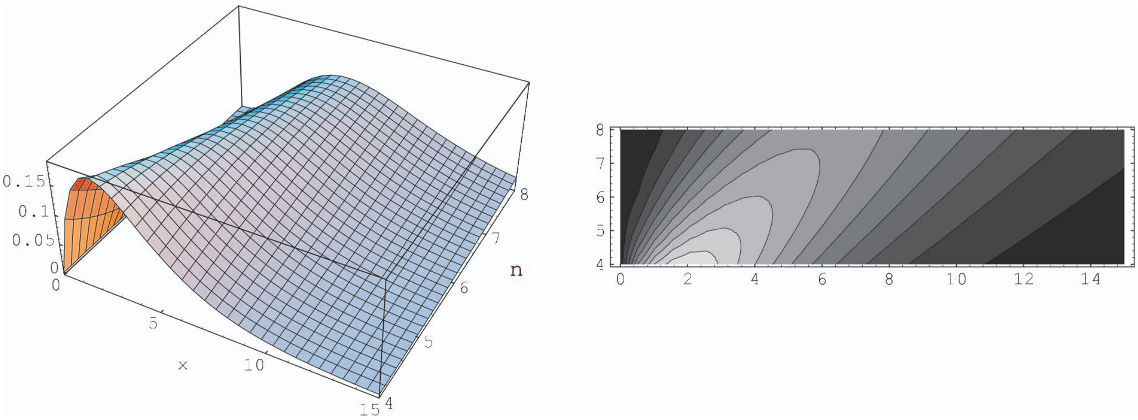


Abbildung 7.11: Dichtefunktion der χ^2–Verteilung mit variablem Freiheitsgrad n.

Die χ^2–Verteilung spielt eine zentrale Rolle in der mathematischen Statistik.

Die Erlang–Verteilung

Es seien $n \in \mathbb{N}$ und $b \in \mathbb{R}^+$. Das zur Dichte

$$f(x) := \begin{cases} \frac{b^n}{(n-1)!} x^{n-1} e^{-bx} & , \ x \geq 0 \\ 0 & , \ x < 0 \end{cases}$$

gehörende Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathfrak{B})$ heißt Erlang-Verteilung mit den Parametern b und n, kurz $Erlang(b, n)$. Für diesen Spezialfall der Gammaverteilung lässt sich die zugehörige Verteilungsfunktion $F(x)$ in geschlossener Form darstellen:

$$F(x) := \begin{cases} 1 - e^{-bx} \cdot \displaystyle\sum_{k=0}^{n-1} \frac{(bx)^k}{k!} & , \ x \geq 0 \\ 0 & , \ x < 0. \end{cases}$$

Die Behauptung lässt sich durch Differenzieren leicht verifizieren:

$$\begin{aligned} F^{'}(x) &= -\left[-b \cdot e^{-bx} \cdot \sum_{k=0}^{n-1} \frac{(bx)^k}{k!} + e^{-bx} \cdot b \cdot \sum_{k=1}^{n-1} \frac{(bx)^{k-1}}{(k-1)!} \right] \\ &= b \cdot e^{-bx} \cdot \frac{(bx)^{n-1}}{(n-1)!} = \frac{b^n}{(n-1)!} x^{n-1} \cdot e^{-bx} = f(x), \quad x \geq 0. \end{aligned}$$

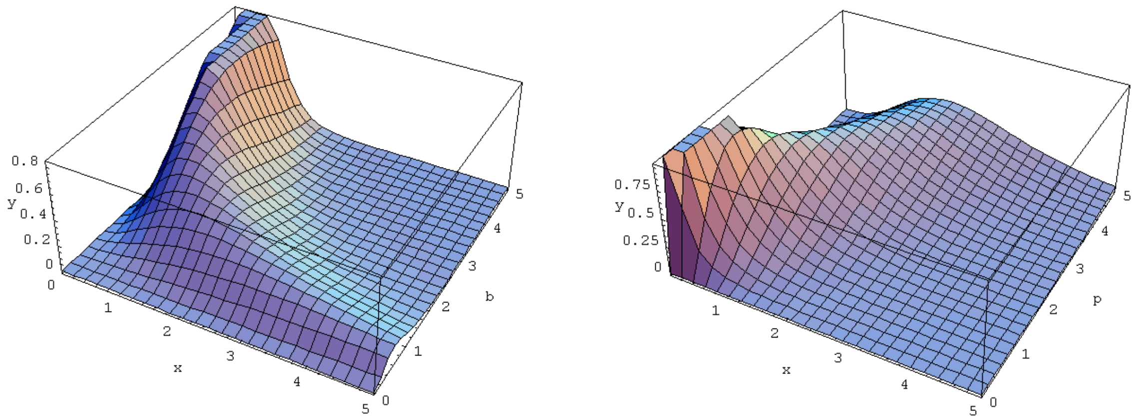


Abbildung 7.12: Dichtefunktion der Erlangverteilung mit variablem Parameter b bei konstantem $p = 3$ (links) und entsprechend umgekehrt (rechts) mit $b = 2$.

Da die Erlang–Verteilung ein Spezialfall der Gammaverteilung ist, ist somit $\int_{\mathbb{R}} f(x)\ dx = 1$ sichergestellt.
Die Erlang–Verteilung verdankt ihren Namen dem dänischen Mathematiker A.K. Erlang, der 1908 Mitarbeiter der Copenhagen Telephone Company wurde und mit seinen Arbeiten zur Leistungsbewertung von Fernsprechvermittlungssystemen den Grundstein für die Warteschlangentheorie legte.

Die Betaverteilung

Das Wahrscheinlichkeitsmaß über $(\mathbb{R}, \mathfrak{B})$ mit der Riemann–Dichte

$$f(x) := \begin{cases} \dfrac{(b-a)^{1-p-q}}{B(p,q)}(x-a)^{p-1}(b-x)^{q-1} & , \ x \in (a,b) \\ 0 & , \ x \notin (a,b) \end{cases}$$

mit $a, b \in \mathbb{R}$, $a < b$ und $p, q > 0$ heißt Betaverteilung 1. Art über dem Intervall (a, b). Der Ausdruck

$$B(p,q) := \int_0^1 t^{p-1}(1-t)^{q-1}\ dt$$

stellt dabei die Betafunktion dar. Für die Betaverteilung mit den Parametern p und q wird das Symbol Beta(p, q) verwendet.
Die Tatsache, dass $\int_{\mathbb{R}} f(x)\ dx = 1$ gilt, folgt durch nachrechnen; mit der Substitution $x = (b-a)t + a$ bzw. $t = \frac{x-a}{b-a}$ und $\frac{dx}{dt} = b - a$ ergibt sich:

$$\begin{aligned} \int_{\mathbb{R}} f(x)\ dx &= \frac{(b-a)^{1-p-q}}{B(p,q)} \int_a^b (x-a)^{p-1}(b-x)^{q-1}\ dx \\ &= \frac{(b-a)^{1-p-q}}{B(p,q)} \int_0^1 (b-a)^{p-1}t^{p-1}(b-a-(b-a)t)^{q-1} \cdot (b-a)\ dt \\ &= \frac{(b-a)^{1-p-q}}{B(p,q)} \cdot (b-a)^{p+q-1} \underbrace{\int_0^1 t^{p-1}(1-t)^{q-1}\ dt}_{=B(p,q)} \ = \ 1. \end{aligned}$$

Die Betaverteilung hat Anwendungen in der Netzplantechnik, in der sie zur Modellierung von Übergangszeiten verwendet wird.

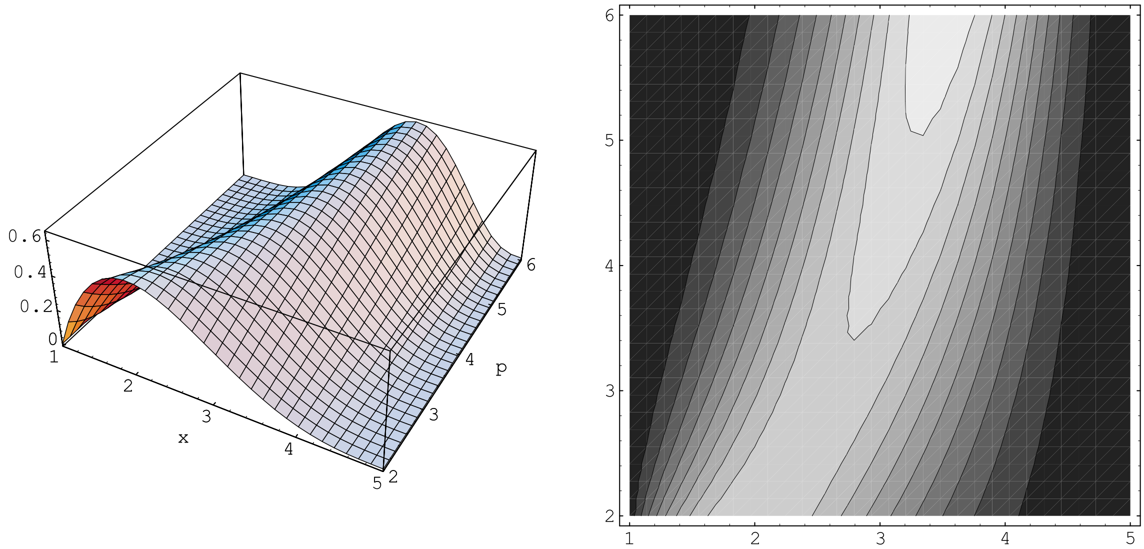


Abbildung 7.13: Dichtefunktion der Betaverteilung mit Parameter $q := 4$ und variablem p.

Literatur zu Kapitel 7

Folgende Bücher werden als begleitende Literatur zu diesem Kapitel des Skriptes empfohlen:

- H. Bauer:
  Wahrscheinlichkeitstheorie,
  5. Auflage, Walter de Gruyter, Berlin, 2002.
  ISBN: 3110172364

- W. Behnen, G. Neuhaus:
  Grundkurs Stochastik,
  3. Auflage, Teubner-Verlag, Stuttgart, 1995.
  ISBN: 3930737698

- P. Billingsley:
  Probability and Measure,
  2nd edition, John Wiley and Sons, New York, 1986.
  ISBN: 0471007102

- P. Gänssler und W. Stute:
  Wahrscheinlichkeitstheorie,
  Springer–Verlag, Berlin, 1977.
  ISBN: 3540084185

- E. Henze:
  Einführung in die Maßtheorie,

Bibl. Institut, Mannheim, 1971.
ISBN: 341100505X

- R. Mathar/D. Pfeiffer:
  Stochastik für Informatiker,
  Teubner, 1990.
  ISBN: 3519022400

- P. P. Spies:
  Grundlagen stochastischer Modelle,
  Hanser, München, 1982.
  ISBN: 3446137114

Kapitel 8

Messbare Abbildungen

In diesem Kapitel wird der Begriff der Zufallsvariablen auf allgemeine Wahrscheinlichkeitsräume ausgedehnt. Um die damit zusammenhängenden Fragen beantworten zu können, wird auf den Begriff der messbaren Abbildung aus der Maßtheorie zurückgegriffen.

Schlüsselwörter: Umkehrabbildung, Urbild, operationstreu, messbare Abbildung, Indikatorvariable, Bildmaß, Zufallsvariable, Zufallsgröße, Zufallsvektor, Verteilungsfunktion einer Zufallsvariable

8.1 Messbare Abbildungen und Bildmaße

Bei manchen Fragestellungen erweist es sich als zweckmäßig, einen gegebenen Wahrscheinlichkeitsraum $(\Omega, \mathfrak{F}, P)$ durch einen modifizierten Wahrscheinlichkeitsraum $(\Omega', \mathfrak{F}', P')$ zu ersetzen, wobei Ω' aus Ω durch eine Abbildung $T\colon \Omega \to \Omega'$ hervorgeht. In diesem Zusammenhang stellt sich die Frage, wie P' von P abhängt bzw. wie T beschaffen sein muss, um P' direkt aus P gewinnen zu können.

8.1 Definition (Urbild):

a) Es sei $T\colon \Omega \to \Omega'$ eine beliebige Abbildung von Ω in Ω'. Dann ist die zu T gehörige Umkehrabbildung $T^{-1}\colon \mathfrak{P}(\Omega') \to \mathfrak{P}(\Omega)$ definiert durch

$$T^{-1}(A') := \{\omega \in \Omega \mid T(\omega) \in A'\} \qquad \forall\ A' \in \mathfrak{P}\big(\Omega'\big)\,.$$

$T^{-1}(A')$ heißt das <u>Urbild</u> von A' unter T.

b) Ist $\mathfrak{M}'$ ein Mengensystem über Ω', so heißt das Mengensystem

$$T^{-1}(\mathfrak{M}') := \{T^{-1}(A') \mid A' \in \mathfrak{M}'\}$$

das Urbild von $\mathfrak{M}'$ unter T.

8.2 Satz (Eigenschaften von T):
Es sei $T\colon \Omega \to \Omega'$ eine Abbildung von Ω in Ω' und es seien ferner $A', A'_1, A'_2, \ldots \in \mathfrak{P}(\Omega')$. Dann gilt:

(i) $T^{-1}(\varnothing) = \varnothing$,

(ii) $T^{-1}(\overline{A'}) = \overline{T^{-1}(A')}$,

(iii) $\displaystyle\bigcup_{i\in I} T^{-1}(A'_i) = T^{-1}\left(\bigcup_{i\in I} A'_i\right)$,

(iv) $\displaystyle\bigcap_{i\in I} T^{-1}(A'_i) = T^{-1}\left(\bigcap_{i\in I} A'_i\right)$.

Man sagt, T ist <u>operationstreu</u>.

Beweis:

(i) Folgt direkt aus der Definition.

(ii) $\omega \in T^{-1}(\overline{A}') \iff T(\omega) \in \overline{A}' \iff T(\omega) \notin A' \iff \omega \notin T^{-1}(A') \iff \omega \in \overline{T^{-1}(A')}$.

(iii) $\displaystyle\omega \in \bigcup_{i\in I} T^{-1}(A'_i) \iff \exists\ i \in I : T(\omega) \in A'_i \iff T(\omega) \in \bigcup_{i\in I} A'_i \iff \omega \in T^{-1}\left(\bigcup_{i\in I} A'_i\right)$.

(iv) $\displaystyle\omega \in \bigcap_{i\in I} T^{-1}(A'_i) \iff \forall\ i \in I : T(\omega) \in A'_i \iff T(\omega) \in \bigcap_{i\in I} A'_i \iff \omega \in T^{-1}\left(\bigcap_{i\in I} A'_i\right)$.

■

Folgerungen:

(v) $T^{-1}(\Omega') = T^{-1}(\overline{\varnothing}) \overset{\text{(ii)}}{=} \overline{T^{-1}(\varnothing)} \overset{\text{(i)}}{=} \overline{\varnothing} = \Omega$.

(vi) $A' \subseteq B' \Longrightarrow T^{-1}(A') \subseteq T^{-1}(B')$.

(vii) $T^{-1}(A'_1 \backslash A'_2) = T^{-1}(A'_1) \backslash T^{-1}(A'_2)$.

(viii) $T^{-1}(A'_1) \cap T^{-1}(A'_2) = \varnothing$ falls $A'_1 \cap A'_2 = \varnothing$.

8.3 Satz:
Es sei $T\colon \Omega \to \Omega'$ eine beliebige Abbildung. Dann gilt:

a) Das Urbild $T^{-1}(\mathfrak{F}')$ einer σ–Algebra $\mathfrak{F}'$ über Ω' ist eine σ–Algebra über Ω.

b) Ist $\mathfrak{F}$ eine σ–Algebra über Ω, so ist das System

$$\mathfrak{S} := \{A' \subset \Omega' \mid T^{-1}(A') \in \mathfrak{F}\}$$

eine σ–Algebra über Ω'.

Beweis:
Der Beweis ergibt sich unmittelbar aus Satz 8.2. ■

Damit lässt sich nun der Begriff des Messraums einführen:

8.4 Definition (Messraum, messbar):
Sind Ω eine Menge und $\mathfrak{F}$ eine σ–Algebra über Ω, so nennt man das Paar $(\Omega, \mathfrak{F})$ Messraum. Es seien $(\Omega, \mathfrak{F})$ und $(\Omega', \mathfrak{F}')$ zwei Messräume. Eine Abbildung $T\colon \Omega \to \Omega'$ heißt $\mathfrak{F}$–$\mathfrak{F}'$–messbar, wenn $T^{-1}(\mathfrak{F}') \subseteq \mathfrak{F}$ ist.

8.5 Beispiel:

1. Es seien $(\Omega, \mathfrak{F})$ und $(\Omega', \mathfrak{F}')$ zwei beliebige Messräume. Die konstante Abbildung $T\colon \Omega \to \Omega'$ mit $T(\omega) := c$ für alle $\omega \in \Omega$ ist messbar, denn es gilt:

$$T^{-1}(\mathfrak{F}') = \begin{cases} \Omega \in \mathfrak{F} & \text{, falls } c \in \mathfrak{F}' \\ \varnothing \in \mathfrak{F} & \text{, falls } c \notin \mathfrak{F}'. \end{cases}$$

2. Es seien $(\Omega, \mathfrak{F})$ und $(\Omega', \mathfrak{F}')$ zwei Messräume und $A \subseteq \Omega$. Die Indikatorvariable

$$T(\omega) \;=\; I_A(\omega) \;:=\; \begin{cases} 1 & , \omega \in A \\ 0 & , \omega \notin A \end{cases}$$

ist genau dann $\mathfrak{F}$–$\mathfrak{F}'$–messbar, wenn $A \in \mathfrak{F}$ ist. Denn es gilt:

$$T^{-1}(A') \;=\; \begin{cases} \varnothing & , \text{ für } 0 \notin A',\ 1 \notin A' \\ A & , \text{ für } 0 \notin A',\ 1 \in A' \\ \overline{A} & , \text{ für } 0 \in A',\ 1 \notin A' \\ \Omega & , \text{ für } 0 \in A',\ 1 \in A'. \end{cases}$$

Mit Hilfe messbarer Abbildungen können Maße auf andere Messräume übertragen werden, wie aus dem nachfolgenden Satz hervorgeht.

8.6 Satz (Bild, Bildmaß):
Es seien $T\colon (\Omega, \mathfrak{F}) \to (\Omega', \mathfrak{F}')$ *eine messbare Abbildung und* μ *ein Maß auf* $(\Omega, \mathfrak{F})$.

a) *Mit der Festlegung* $\mu'(A') := \mu(T^{-1}(A'))$ *für alle* $A' \in \mathfrak{F}'$ *wird ein Maß* μ' *auf* $(\Omega', \mathfrak{F}')$ *definiert.*

b) *Ist* μ *ein Wahrscheinlichkeitsmaß auf* $(\Omega, \mathfrak{F})$, *dann ist auch* μ' *ein Wahrscheinlichkeitsmaß auf* $(\Omega', \mathfrak{F}')$.

μ' *heißt das* *Bild* *oder* *Bildmaß* *von* μ *unter* T.

Beweis:
zu a) Es gilt (siehe Satz 8.2):

$$\begin{aligned} \mu'(A') &= \mu(T^{-1}(A')) \geq 0 \qquad \forall\, A' \in \mathfrak{F}' \\ \mu'(\varnothing) &= \mu(T^{-1}(\varnothing)) = \mu(\varnothing) = 0. \end{aligned}$$

Aufgrund der σ–Additivität von μ gilt für jede Folge paarweise fremder Mengen $(A'_i)_{i\in\mathbb{N}}$ aus $\mathfrak{F}'$:

$$\begin{aligned} \mu'\left(\bigcup_{i\in\mathbb{N}} A'_i\right) &= \mu\left(T^{-1}\left(\bigcup_{i\in\mathbb{N}} A'_i\right)\right) \\ &= \mu\left(\bigcup_{i\in\mathbb{N}} T^{-1}(A'_i)\right) \quad \text{(aufgrund von Satz 8.2 (iii))} \\ &= \sum_{i\in\mathbb{N}} \mu(T^{-1}(A'_i)) \quad (\mu \text{ ist } \sigma\text{–additiv}) \\ &= \sum_{i\in\mathbb{N}} \mu'(A'_i), \end{aligned}$$

d.h. auch μ' ist σ-additiv.

zu b) $\mu'(\Omega') = \mu(T^{-1}(\Omega')) = \mu(\Omega) = 1$ (Satz 8.2).

■

Mit Hilfe dieser Vorüberlegung kann der Begriff der Zufallsvariablen im allgemeinen Fall eingeführt werden.

8.7 Definition ((reelle) Zufallsvariable, Zufallsgröße, Zufallsvektor):
Es seien $(\Omega, \mathfrak{F}, P)$ *ein Wahrscheinlichkeitsraum und* $(\Omega', \mathfrak{F}')$ *ein Messraum, dann heißt jede* $\mathfrak{F}$–$\mathfrak{F}'$*–messbare Abbildung* $X\colon (\Omega, \mathfrak{F}) \to (\Omega', \mathfrak{F}')$ *Zufallsvariable.*
Ist speziell $(\Omega', \mathfrak{F}') := (\mathbb{R}, \mathfrak{B})$ *bzw.* $:= (\mathbb{R}^n, \mathfrak{B}^n)$, $n > 1$, *so heißt* X *eine reelle Zufallsvariable (Zufallsgröße) bzw. ein reeller* n*-dimensionaler Zufallsvektor.*

8.8 Bemerkung:
Es seien X eine reelle Zufallsvariable und P_X das Bildmaß von P unter X. Die Verteilungsfunktion von P_X sei F_X. Für $P(\{\omega \in \Omega \mid X(\omega) \leq x\})$ wird auch $P(X \leq x)$ geschrieben. Aufgrund des Zusammenhangs von X und P_X gilt:

$$F_X(x) \;=\; P_X((-\infty, x]) \;=\; P(X^{-1}((-\infty, x])) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}) = P(X \leq x),$$

d.h. es ist $F_X(x) = P(X \leq x)$. In diesem Fall sagt man, dass X nach F_X verteilt ist bzw. dass X die Verteilungsfunktion F_X besitzt. Ist F die Verteilungsfunktion von X, so wird auch $X \stackrel{d}{=} F$ („d" von englisch „distribution") geschrieben, wobei für F in der Regel das spezifische Verteilungssymbol verwendet wird, z.B. $X \stackrel{d}{=} \mathrm{Exp}(\lambda)$.

8.2 Kriterien für Messbarkeit

Für Anwendungen ist es wichtig, über einfache Kriterien für die Messbarkeit von Abbildungen zu verfügen. Solche Kriterien können aus dem folgenden Satz abgeleitet werden.

8.9 Satz:
Es seien $(\Omega, \mathfrak{F})$ und $(\Omega', \mathfrak{F}')$ zwei Meßräume und E' ein Erzeugendensystem von $\mathfrak{F}'$, d.h. $\mathfrak{F}' = \sigma(E')$. Die Abbildung $T\colon (\Omega, \mathfrak{F}) \to (\Omega', \mathfrak{F}')$ ist genau dann messbar, wenn gilt:

$$T^{-1}(E') \subseteq \mathfrak{F}.$$

Beweis:
Ist T messbar, so gilt $T^{-1}(A') \in \mathfrak{F}$ für alle $A' \in \mathfrak{F}'$, also auch

$$T^{-1}(E') = \{T^{-1}(A') \mid A' \in E'\} \subseteq \mathfrak{F}.$$

Für die umgekehrte Richtung wird das System $S' := \{A' \subseteq \Omega' \mid T^{-1}(A') \in \mathfrak{F}\}$ betrachtet. Zu zeigen ist: $S' \supseteq \mathfrak{F}'$. Nach Satz 8.3 b) ist S' eine σ–Algebra über Ω'. Aufgrund der Voraussetzung gilt $E' \subseteq S'$. Hieraus folgt $\sigma(E') = \mathfrak{F}' \subseteq S'$, da $\mathfrak{F}'$ die kleinste σ-Algebra ist, die E' enthält, d.h. $T^{-1}(A') \in \mathfrak{F}$ für alle $A' \in \mathfrak{F}'$. ■

8.10 Bemerkung:
Es sei $(\Omega, \mathfrak{F}, \mu)$ ein Maßraum. Im Folgenden werden lediglich Abbildungen $f\colon \Omega \to \overline{\mathbb{R}}$ untersucht, deren Werte in den erweiterten reellen Zahlen $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ liegen. Zu $\overline{\mathbb{R}}$ gehört die erweiterte σ–Algebra der Borelschen Mengen

$$\overline{\mathfrak{B}} \;:=\; \{B_0,\; B_0 \cup \{-\infty\},\; B_0 \cup \{+\infty\},\; B_0 \cup \{-\infty, +\infty\} \mid B_0 \in \mathfrak{B}\}.$$

Man nennt $\mathfrak{F}$–$\overline{\mathfrak{B}}$–messbare Abbildungen auch <u>$\mathfrak{F}$–messbare numerische Funktionen</u>.

Die nachstehenden Sätze vereinfachen die Analyse der Messbarkeit von Abbildungen.

8.11 Satz:
Eine numerische Funktion $f\colon \Omega \to \overline{\mathbb{R}}$ ist genau dann $\mathfrak{F}$–$\overline{\mathfrak{B}}$–messbar, wenn für alle $c \in \mathbb{R}$ gilt:

a) $\{\omega \mid f(\omega) < c\} \in \mathfrak{F}$

bzw. äquivalent dazu

b) $\{\omega \mid f(\omega) \leq c\} \in \mathfrak{F}$,

c) $\{\omega \mid f(\omega) > c\} \in \mathfrak{F}$,

d) $\{\omega \mid f(\omega) \geq c\} \in \mathfrak{F}$.

Beweis:

a) Überträgt man die Messbarkeitsanforderung aus Satz 8.9 auf die Mengensysteme $\mathfrak{F} := \mathfrak{B}$ und $\mathfrak{F}' := \overline{\mathfrak{B}}$, so muss für den Nachweis von a) gezeigt werden, dass die Intervalle $[-\infty, c)$ die σ–Algebra $\overline{\mathfrak{B}}$ erzeugen. Aus $[c-n, c) \in \mathfrak{B}\ \ \forall\, n \in \mathbb{N}$ folgt $(-\infty, c) = \bigcup\limits_{n=1}^{\infty} [c-n, c) \in \mathfrak{B}$. Wegen $\mathfrak{B} \subset \overline{\mathfrak{B}}$ und wegen $\{-\infty\} \in \overline{\mathfrak{B}}$ ist deshalb auch $[-\infty, c) \in \overline{\mathfrak{B}}$. Sei nun $\widetilde{\mathfrak{B}}$ die von den Intervallen $[-\infty, c),\ c \in \mathbb{R}$, erzeugte σ–Algebra. Aufgrund der Vorbemerkung gilt $\widetilde{\mathfrak{B}} \subset \overline{\mathfrak{B}}$. Um $\overline{\mathfrak{B}} \subset \widetilde{\mathfrak{B}}$ zu zeigen, betrachten wir die Intervalle $[a, b)$. Wegen $[a, b) = [-\infty, b) \setminus [-\infty, a),\ a \leq b$ gilt $[a, b) \in \widetilde{\mathfrak{B}}$ bzw. $\mathbb{I}^1 \subset \widetilde{\mathfrak{B}}$. Wegen $[a, b] = \bigcap\limits_{n=1}^{\infty} \left[a, b + \frac{1}{n}\right) \in \widetilde{\mathfrak{B}}$ und damit $[-\infty, b] \in \widetilde{\mathfrak{B}}$ für alle $b \in \mathbb{R}$ sind auch $\{-\infty\} = \bigcap\limits_{n=1}^{\infty} [-\infty, -n) \in \widetilde{\mathfrak{B}}$ und $\{+\infty\} = \bigcap\limits_{n=1}^{\infty} \overline{[-\infty, n]} \in \widetilde{\mathfrak{B}}$, womit $\overline{\mathfrak{B}} \subset \widetilde{\mathfrak{B}}$ gezeigt ist. Insgesamt ergibt sich $\widetilde{\mathfrak{B}} = \overline{\mathfrak{B}}$. Und wir schließen weiter:

b) $\{\omega \mid f(\omega) \leq c\} = \bigcap\limits_{n=1}^{\infty} \left\{\omega \mid f(\omega) < c + \frac{1}{n}\right\} \in \mathfrak{F}$,

c) $\{\omega \mid f(\omega) > c\} = \overline{\{\omega \mid f(\omega) \leq c\}} \in \mathfrak{F}$,

d) $\{\omega \mid f(\omega) \geq c\} = \overline{\{\omega \mid f(\omega) < c\}} \in \mathfrak{F}$. ■

8.12 Satz:
Die beiden Funktionen $f\colon \Omega \to \overline{\mathbb{R}}$ *und* $g\colon \Omega \to \overline{\mathbb{R}}$ *seien* $\mathfrak{F}$–$\overline{\mathfrak{B}}$*–messbar, dann gilt:*

a) $\{\omega \mid f(\omega) < g(\omega)\} \in \mathfrak{F}$,

b) $\{\omega \mid f(\omega) \leq g(\omega)\} \in \mathfrak{F}$,

c) $\{\omega \mid f(\omega) = g(\omega)\} \in \mathfrak{F}$.

Beweis: Übung.

8.13 Satz:
Es seien f*,* g *und* f_n*,* $n \in \mathbb{N}$*,* $\mathfrak{F}$–$\overline{\mathfrak{B}}$*–messbare Abbildungen von* Ω *nach* $\overline{\mathbb{R}}$ *und* $c \in \mathbb{R}$*. Dann sind auch folgende Abbildungen* $\mathfrak{F}$–$\overline{\mathfrak{B}}$*–messbar:*

a) $f + g$ *(wobei die Summe überall definiert sein muss),*

b) $c \cdot f$,

c) $f \cdot g$,

d) $\frac{1}{g}$ *mit* $\frac{1}{g(\omega)} = +\infty$ *falls* $g(\omega) = 0$,

e) $\sup\limits_{n\in\mathbb{N}} f_n$, $\inf\limits_{n\in\mathbb{N}} f_n$,

f) $max(f, g)$, $min(f, g)$, $f^+ := max(f, 0)$ *und* $f^- := -\,min(f, 0)$,

g) $\liminf\limits_{n\to\infty} f_n$, $\limsup\limits_{n\to\infty} f_n$,

h) $\lim\limits_{n\to\infty} f_n$, *falls* $\{f_n(\omega)\}$ *für alle* ω *konvergiert.*

Beweis:

a) Aufgrund der Messbarkeit von f gilt stets $\{\omega \mid f(\omega) < a\} \in \mathfrak{F}$ für alle $a \in \mathbb{R}$. Entsprechendes gilt für $g, f_1, f_2, \ldots$. Aus

$$\{\omega \mid f(\omega) + g(\omega) < a\} = \bigcup_{r\in\mathbb{Q}} (\{\omega \mid f(\omega) < r\} \cap \{\omega \mid g(\omega) < a - r\})$$

folgt zusammen mit Satz 8.11, dass $f + g$ (und analog auch $f - g$) messbar sind.

b) Es ist

$$\{\omega \mid cf(\omega) < a\} = \begin{cases} \{\omega \mid f(\omega) < \frac{a}{c}\} & , \ c > 0 \\ \{\omega \mid f(\omega) > \frac{a}{c}\} & , \ c < 0. \end{cases}$$

Ist $c = 0$, dann ist $cf = 0$ eine konstante Funktion, die messbar ist.

c) Zunächst wird gezeigt, dass f^2 messbar ist:

$$\begin{aligned} \{\omega \mid f^2(\omega) < a\} &= \begin{cases} \varnothing & , \ a \le 0 \\ \{\omega \mid -\sqrt{a} < f(\omega) < \sqrt{a}\} & , \ a > 0 \end{cases} \\ &= \begin{cases} \varnothing \in \mathfrak{F} & , \ a \le 0 \\ \{\omega \mid f(\omega) < \sqrt{a}\} \cap \{\omega \mid f(\omega) > -\sqrt{a}\} \in \mathfrak{F} & , \ a > 0. \end{cases} \end{aligned}$$

Da $f \cdot g$ für reellwertige f und g als

$$f \cdot g = \frac{1}{4}((f + g)^2 - (f - g)^2)$$

dargestellt werden kann, ist in diesem Fall auch $f \cdot g$ messbar. Im Falle numerischer Funktionen f und g sieht man, dass die Mengen $A_1 := \{\omega \mid f(\omega) \cdot g(\omega) = \infty\}$, $A_2 := \{\omega \mid f(\omega) \cdot g(\omega) = -\infty\}$, $A_3 := \{\omega \mid f(\omega) \cdot g(\omega) = 0\}$ und $A_4 := \overline{A_1 \cup A_2 \cup A_3}$ in $\mathfrak{F}$ liegen. Die Restriktionen f^* und g^* von f und g auf A_4 sind $A_4 \cap \mathfrak{F} - \overline{\mathfrak{B}}$–messbar und reellwertig. Das Produkt $f^* \cdot g^*$ ist daher ebenfalls $A_4 \cap \mathfrak{F} - \overline{\mathfrak{B}}$–messbar. Folglich ist $f \cdot g$ $\mathfrak{B} - \overline{\mathfrak{B}}$–messbar.

d) Es sei $g(\omega) \neq 0$ für alle $\omega \in \Omega$. Offensichtlich ist für $a > 0$

$$\begin{aligned} \left\{\omega \mid \frac{1}{g(\omega)} < a\right\} &= \left(\{\omega \mid g(\omega) > 0\} \cap \left\{\omega \mid g(\omega) > \frac{1}{a}\right\}\right) \cup \\ &\cup \left(\{\omega \mid g(\omega) < 0\} \cap \left\{\omega \mid g(\omega) < \frac{1}{a}\right\}\right) \in \mathfrak{F}. \end{aligned}$$

Damit ist nicht nur $1/g$, sondern wegen b) auch $f/g = f \cdot (1/g)$ messbar. Für $a < 0$ schließt man analog.

e) Es wird $f(\omega) := \sup_{n\in\mathbb{N}} f_n(\omega)$ und $g(\omega) = \inf_{n\in\mathbb{N}} f_n(\omega)$ gesetzt. Damit wird

$$\begin{aligned}
\{\omega \mid f(\omega) < a\} &= \bigcap_{n\in\mathbb{N}} \{\omega \mid f_n(\omega) < a\} \in \mathfrak{F}, \\
\{\omega \mid g(\omega) > a\} &= \bigcap_{n\in\mathbb{N}} \{\omega \mid f_n(\omega) > a\} \in \mathfrak{F}.
\end{aligned}$$

f) Die Behauptungen folgen aus den Zusammenhängen:

$$\begin{aligned}
\{\omega \mid \max(f,g)(\omega) < a\} &= \{\omega \mid f(\omega) < a\} \cap \{\omega \mid g(\omega) < a\}, \\
\{\omega \mid \min(f,g)(\omega) > a\} &= \{\omega \mid f(\omega) > a\} \cap \{\omega \mid g(\omega) > a\}.
\end{aligned}$$

g) Diese Aussage ist eine unmittelbare Konsequenz aus d) und e), denn es gilt:

$$\begin{aligned}
\limsup_{n\to\infty} f_n(\omega) &= \inf_{k\in\mathbb{N}} \sup_{n\geq k} f_n(\omega), \\
\liminf_{n\to\infty} f_n(\omega) &= \sup_{k\in\mathbb{N}} \inf_{n\geq k} f_n(\omega).
\end{aligned}$$

h) Es gilt:

$$\limsup_{n\to\infty} f_n(\omega) = \liminf_{n\to\infty} f_n(\omega) = \lim_{n\to\infty} f_n(\omega).$$

■

8.14 Satz:
Für $i \in \{1,2,3\}$ seien $(\Omega_i, \mathfrak{F}_i)$ Messräume. Die Abbildung $f\colon \Omega_1 \to \Omega_2$ sei $\mathfrak{F}_1$–$\mathfrak{F}_2$–messbar und die Abbildung $g\colon \Omega_2 \to \Omega_3$ sei $\mathfrak{F}_2$–$\mathfrak{F}_3$–messbar. Dann ist $g \circ f\colon \Omega_1 \to \Omega_3$ eine $\mathfrak{F}_1$–$\mathfrak{F}_3$–messbare Abbildung.

Beweis:
Die $\mathfrak{F}_1$–$\mathfrak{F}_3$–Messbarkeit von $g \circ f$ folgt aus der Tatsache, dass $(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A))$ für alle $A \in \mathfrak{F}_3$ gilt. ■

8.15 Satz:
Jede stetige Abbildung $T : \mathbb{R}^n \to \mathbb{R}^m$, $n, m \in \mathbb{N}$, ist $\mathfrak{B}^n$–$\mathfrak{B}^m$–messbar.

Beweis:
Es wird Satz 8.9 angewandt: Bekanntlich bilden die offenen Mengen des $\mathbb{R}^n$ ein Erzeugendensystem von $\mathfrak{B}^n$ (Übung). Bei einer stetigen Abbildung ist das Urbild einer offenen Menge wieder eine offene Menge. Die offenen Mengen liegen aber in $\mathfrak{B}^m$. ■

Literatur zu Kapitel 8

Folgende Bücher werden als begleitende Literatur zu diesem Kapitel des Skriptes empfohlen:

- H. Bauer:
  Wahrscheinlichkeitstheorie,
  5. Auflage, Walter de Gruyter, Berlin, 2002.
  ISBN: 3110172364

- W. Behnen, G. Neuhaus:
  Grundkurs Stochastik,
  3. Auflage, Teubner-Verlag, Stuttgart, 1995.
  ISBN: 3930737698

- P. Billingsley:
  Probability and Measure,
  2nd edition, John Wiley and Sons, New York, 1986.
  ISBN: 0471007102

- P. Gänssler und W. Stute:
  Wahrscheinlichkeitstheorie,
  Springer–Verlag, Berlin, 1977.
  ISBN: 3540084185

- E. Henze:
  Einführung in die Maßtheorie,
  Bibl. Institut, Mannheim, 1971.
  ISBN: 341100505X

- R. Mathar/D. Pfeiffer:
  Stochastik für Informatiker,
  Teubner, 1990.
  ISBN: 3519022400

- P. P. Spies:
  Grundlagen stochastischer Modelle,
  Hanser, München, 1982.
  ISBN: 3446137114

Kapitel 9

Integration

In diesem Kapitel wird der aus der Analysis bekannte Begriff des Lebesgue-Integrals verallgemeinert und für die Definition des Erwartungswertes allgemeiner Zufallsvariablen herangezogen.

Schlüsselwörter: messbare numerische Funktion, Elementarfunktion, μ–Integral, Positiv– und Negativteil einer messbaren numerischen Funktion, Erwartungswert und Varianz einer reellen Zufallsvariable, Transformationssatz für Erwartungswerte, k–tes Moment, k–tes zentrales Moment, Variationskoeffizient.

9.1 Vorbemerkungen

In Analogie zum Begriff des Erwartungswertes für diskrete Zufallsvariablen würde es naheliegen, für eine reellwertige Zufallsvariable X mit Dichte f zu definieren:

$$\mathrm{E}[X] := \int_{\mathbb{R}} x \cdot f(x)\ dx,$$

sofern

$$\int_{\mathbb{R}} |x| \cdot f(x)\ dx < \infty.$$

Für eine mit dem Parameter λ exponentiell verteilte Zufallsvariable würde sich auf diese Weise z.B.

$$\begin{aligned}
\mathrm{E}[X] &= \int_{\mathbb{R}} x \cdot f(x)\ dx \\
&= \int_{\mathbb{R}^+} x \cdot \lambda \cdot e^{-\lambda x}\ dx \\
&\quad \text{und nach partieller Integration:} \\
&= \left(\lambda x \cdot \left(-\frac{1}{\lambda} \right) e^{-\lambda x} \right) \Bigg|_0^\infty - \lambda \int_0^\infty \left(-\frac{1}{\lambda} \right) e^{-\lambda x}\ dx \\
&= -x e^{-\lambda x} \Big|_0^\infty - \frac{1}{\lambda} e^{-\lambda x} \Big|_0^\infty \\
&= 0 - \left(0 - \frac{1}{\lambda} \right) = \frac{1}{\lambda}
\end{aligned}$$

ergeben. Allerdings führen schon einfache praktische Fragestellungen auf Zufallsvariablen, die weder diskret noch stetig sind. Als Beispiel betrachten wir die Wartezeitverteilung an einem Fahrkartenschalter. Aus eigener Beobachtung weiß man, dass das System selbst bei hoher Auslastung des Bedieners immer wieder einmal in einen leeren Zustand zurückkehrt, so dass ein neu ankommender Kunde überhaupt nicht warten muss. Folglich besitzt die Wartezeitverteilung $F(w)$, die angibt wie groß die Wahrscheinlichkeit ist, eine Zeit kleiner als w zu warten, an der Stelle $w = 0$ einen Sprung der Höhe p. Dabei bedeutet p die Wahrscheinlichkeit, dass der Bediener frei ist. Da $p > 0$ ist, besitzt $F(w)$ keine Dichte. Deswegen ist es notwendig, einen allgemeineren Integralbegriff einzuführen.

Vorgehensweise in den nächsten Abschnitten:
Ziel des Kapitels ist es, den Begriff des Erwartungswertes aus dem maßtheoretischen Begriff des μ–Integrals für messbare numerische Funktionen zu entwickeln.
Dies geschieht in drei Schritten:

1. Definiere das μ–Integral für Elementarfunktionen (Kapitel 9.2).
2. Erweitere den Begriff des μ–Integrals für Elementarfunktionen durch Grenzübergang auf den Begriff des μ–Integrals für nichtnegative messbare Funktionen (Kapitel 9.3).

3. Erweitere den Begriff des μ–Integrals auf messbare numerische Funktionen durch Betrachtung ihrer Positiv– und Negativ–Teile (Kapitel 9.4).

Damit lässt sich in Kapitel 9.5 der Begriff des Erwartungswerts definieren.

9.2 Das µ–Integral von Elementarfunktionen

9.1 Definition (Elementarfunktion):
Es sei $(\Omega, \mathfrak{F})$ ein Meßraum. Eine reellwertige Funktion $e \colon \Omega \to \mathbb{R}$ heißt Elementarfunktion, wenn sie $\mathfrak{F}$–$\mathfrak{B}$–messbar ist und nur endlich viele Werte annimmt.

9.2 Satz:
Ist $e(\omega)$ eine Elementarfunktion, dann existieren eine Partition $(A_i)_{i=1}^n$ von Ω und reelle Zahlen $\alpha_i \in \mathbb{R}$, $i = 1, \ldots, n$, so dass mit den Indikatorvariablen I_{A_i}, $i = 1, \ldots, n$, gilt:

$$e(\omega) = \sum_{i=1}^{n} \alpha_i I_{A_i}(\omega). \tag{9.1}$$

Beweis:
Besitzt $e(\omega)$ die n verschiedenen Werte x_i, so setzen wir $\alpha_i := x_i$, $i = 1, \ldots, n$. Die Träger (Urbilder) der Werte α_i

$$A_i := e^{-1}(\alpha_i) = \{\omega \in \Omega \mid e(\omega) = \alpha_i\}, \quad i = 1, \ldots, n$$

bilden ein disjunktes System von Teilmengen aus Ω. Es ist $A_i \in \mathfrak{F}$, da die Mengen $\{\alpha_i\} \in \mathfrak{B}$ sind und $e(\omega)$ messbar ist. Es ist $e(\omega) = \alpha_i$, falls $\omega \in A_i$. Da die A_i disjunkt sind, kann man dafür auch

$$e(\omega) = \sum_{i=1}^{n} \alpha_i I_{A_i}(\omega) = \begin{cases} \alpha_1 \ , & \omega \in A_1 \\ \alpha_2 \ , & \omega \in A_2 \\ \vdots & \\ \alpha_n \ , & \omega \in A_n \end{cases}$$

schreiben. ∎

Die unter (9.1) angegebene Darstellung von e heißt eine Normaldarstellung von e.

9.3 Satz:
Summe, Differenz und Produkt von Elementarfunktionen sind wieder Elementarfunktionen. Ist e elementar und $\lambda \in \mathbb{R}$, dann ist auch λe elementar.

Beweis:
Die Messbarkeit der zusammengesetzten Funktionen folgt aus Satz 8.13. Offensichtlich bleibt bei den genannten Operationen auch die Endlichkeit der Wertebereiche erhalten. ∎

9.4 Definition (integrierbar, integrabel):
Es sei $(\Omega, \mathfrak{F}, \mu)$ ein Maßraum. Eine Elementarfunktion $e(\omega) := \sum_{i=1}^{n} \alpha_i I_{A_i}(\omega)$ heißt integrabel oder integrierbar, wenn $\mu(A_i) < \infty$ für alle $i = 1, \ldots, n$ mit $\alpha_i \neq 0$ gilt.

9.5 Definition (bestimmtes Integral, μ-Integral):
Es sei $(\Omega, \mathfrak{F}, \mu)$ ein Maßraum. Weiter sei $e\colon \Omega \longrightarrow \mathbb{R}$ eine nichtnegative Elementarfunktion mit einer Normaldarstellung $e(\omega) := \sum_{i=1}^{n} \alpha_i I_{A_i}(\omega)$. Dann heißt

$$\int\limits_{\Omega} e(\omega)\ d\mu(\omega) \ := \ \int e\ d\mu \ := \ \sum_{i=1}^{n} \alpha_i \mu(A_i)$$

das <u>bestimmte Integral</u> oder <u>μ–Integral</u> von e über Ω.

9.6 Satz:
Das bestimmte Integral über Ω hängt nicht von der speziellen Wahl der Normaldarstellung ab.

Beweis:
Es seien $e(\omega) = \sum_{i=1}^{n} \alpha_i I_{A_i}$ und $e(\omega) = \sum_{j=1}^{m} \beta_j I_{B_j}(\omega)$ zwei Normaldarstellungen von e. Dann ist

$$\begin{aligned} A_i &= A_i \cap \Omega = A_i \cap \bigcup_{j=1}^{m} B_j = \bigcup_{j=1}^{m} (A_i \cap B_j)\,, \\ B_j &= B_j \cap \Omega = B_j \cap \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} (A_i \cap B_j)\,. \end{aligned}$$

Also

$$\begin{aligned} I_{A_i}(\omega) &= I_{\bigcup_{j=1}^{m}(A_i \cap B_j)}(\omega) = \sum_{j=1}^{m} I_{A_i \cap B_j}(\omega)\,, \\ I_{B_j}(\omega) &= I_{\bigcup_{i=1}^{n}(A_i \cap B_j)}(\omega) = \sum_{i=1}^{n} I_{A_i \cap B_j}(\omega)\,. \end{aligned}$$

Damit wird

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i I_{A_i \cap B_j}(\omega) = \sum_{i=1}^{n} \alpha_i I_{A_i}(\omega) = e(\omega) = \sum_{j=1}^{m} \beta_j I_{B_j}(\omega) = \sum_{j=1}^{m}\sum_{i=1}^{n} \beta_j I_{A_i \cap B_j}(\omega),$$

woraus wegen der Vertauschbarkeit der endlichen Summen $\alpha_i = \beta_j$ für alle $\omega \in A_i \cap B_j \neq \varnothing$ folgt. Andererseits ist

$$\int e\ d\mu \ = \ \int \sum_{i=1}^{n} \alpha_i I_{\bigcup_{j=1}^{m}(A_i \cap B_j)} \ = \ \sum_{i=1}^{n} \alpha_i \mu\left(\bigcup_{j=1}^{m} A_i \cap B_j\right) \ = \ \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i \mu(A_i \cap B_j)$$

und

$$\int e\ d\mu \ = \ \int \sum_{j=1}^{m} \beta_j I_{\bigcup_{i=1}^{n}(A_i \cap B_j)} \ = \ \sum_{j=1}^{m} \beta_j \mu\left(\bigcup_{i=1}^{n} A_i \cap B_j\right) \ = \ \sum_{i=1}^{n}\sum_{j=1}^{m} \beta_j \mu(A_i \cap B_j).$$

Da aber für $\omega \in A_i \cap B_j \neq \varnothing$, $\alpha_i = \beta_j$, gilt, müssen die beiden Summen identisch sein. ■

9.7 Satz (Eigenschaften des Integrals):
Für das Integral nichtnegativer Elementarfunktionen e, e_1 und e_2 gilt:

a) $\int I_A \, d\mu = \mu(A)$.

b) *Für* $\lambda \in \mathbb{R}^+$ *ist* $\int \lambda e \, d\mu = \lambda \int e \, d\mu$ *(Linearität).*

c) $\int (e_1 + e_2) \, d\mu = \int e_1 \, d\mu + \int e_2 \, d\mu$ *(Linearität).*

d) *Aus* $e_1 \leq e_2$ *folgt* $\int e_1 \, d\mu \leq \int e_2 \, d\mu$ *(Monotonie).*

e) $\int e_1 \, d(\mu + \nu) = \int e_1 \, d\mu + \int e_1 \, d\nu$ *(Linearität).*

Beweis:

a) Ergibt sich unmittelbar aus der Definition der Indikatorfunktion.

b) Aus $e(\omega) = \sum_{i=1}^{n} \alpha_i I_{A_i}(\omega)$ ergibt sich $\sum_{i=1}^{n} (\lambda \alpha_i) \mu(A_i) = \lambda \sum_{i=1}^{n} \alpha_i \mu(A_i)$.

c) Aus

$$e_1 = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i I_{A_i \cap B_j} \qquad \text{und} \qquad e_2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \beta_j I_{A_i \cap B_j}$$

folgt

$$e_1 + e_2 = \sum_{i=1}^{n} \sum_{j=1}^{m} (\alpha_i + \beta_j) I_{A_i \cap B_j} = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i I_{A_i \cap B_j} + \sum_{i=1}^{n} \sum_{j=1}^{m} \beta_j I_{A_i \cap B_j}.$$

d) Aus $e_1 \leq e_2$ folgt $e_2 - e_1 \geq 0$ und damit $\int (e_2 - e_1) \, d\mu \geq 0$. Eine Anwendung von c) ergibt weiter

$$\int e_2 \, d\mu \;=\; \int (e_2 - e_1) \, d\mu + \int e_1 \, d\mu \;\geq\; \int e_1 \, d\mu,$$

was zu zeigen war.

e) Es gilt:

$$\begin{aligned} \int e_1 \, d(\mu + \nu) &= \sum_{i=1}^{n} \alpha_i (\mu(A_i) + \nu(A_i)) = \sum_{i=1}^{n} \alpha_i \mu(A_i) + \sum_{i=1}^{n} \alpha_i \nu(A_i) \\ &= \int e_1 \, d\mu + \int e_1 \, d\nu. \end{aligned}$$

■

9.3 Das µ–Integral nichtnegativer messbarer numerischer Funktionen

9.8 Satz:
Jede nichtnegative messbare numerische Funktion $f\colon \Omega \to \overline{\mathbb{R}^+}$ ist punktweiser Grenzwert einer monoton nichtfallenden Folge nichtnegativer Elementarfunktionen.

Beweis:
Definiere für alle $\omega \in \Omega$:

$$f_n(\omega) := \begin{cases} \dfrac{k-1}{2^n} & , \dfrac{k-1}{2^n} \leq f(\omega) < \dfrac{k}{2^n} \\ n & , f(\omega) \geq n \end{cases} \qquad (k = 1, \ldots, n \cdot 2^n).$$

Zu zeigen ist $\lim\limits_{n\to\infty} f_n(\omega) = f(\omega)$.
Die Funktionen f_n lassen sich wie folgt durch Indikatorfunktionen ausdrücken:

$$f_n = \sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n} \cdot I_{A_{n,k}} + n \cdot I_{B_n}$$

mit $A_{n,k} := f^{-1}\left(\left[\dfrac{k-1}{2^n}, \dfrac{k}{2^n}\right)\right)$ und $B_n := f^{-1}([n, \infty))$.
Nun kann man wie folgt schließen:

$$\begin{aligned} & f \text{ ist messbar} \\ \Rightarrow & A_{n,k} \in \mathfrak{F} \text{ und } B_n \in \mathfrak{F} \text{ für alle } n \text{ und } k \\ \Rightarrow & \text{Die Indikatorfunktionen } I_{A_{n,k}} \text{ und } I_{B_n} \text{ sind messbar} \\ \Rightarrow & f_n \text{ ist messbar.} \end{aligned}$$

Für alle $\omega \in \Omega$ mit $f(\omega) < \infty$ und für alle $n \in \mathbb{N}$ mit $f(\omega) < n$ folgt

$$0 \leq f(\omega) - f_n(\omega) \leq \frac{1}{2^n} \quad \to 0 \quad (\text{ für } \; n \to \infty).$$

Für alle $\omega \in \Omega$ mit $f(\omega) = \infty$ gilt $f_n(\omega) = n \to \infty$ für $n \to \infty$. ∎

9.9 Beispiel:
Es sei $f\colon \mathbb{R} \to \mathbb{R}$, $f(\omega) := \omega^2$. In diesem Fall ist

$$f_1(\omega) = \begin{cases} 0 \;, & 0 \leq f(\omega) < \dfrac{1}{2} \\ \dfrac{1}{2} \;, & \dfrac{1}{2} \leq f(\omega) < 1 \\ 1 \;, & f(\omega) \geq 1 \end{cases}$$

bzw.

$$f_1(\omega) \;=\; \sum_{k=1}^{2} \frac{k-1}{2} I_{A_{1,k}}(\omega) + 1 \cdot I_{B_1}(\omega)$$

mit

$$\begin{aligned} A_{1,1} &= \left(-\frac{1}{\sqrt{2}}, 0\right] \cup \left[0, \frac{1}{\sqrt{2}}\right) = \left(-\sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}}\right), \\ A_{1,2} &= \left(-1, -\frac{1}{\sqrt{2}}\right] \cup \left[\frac{1}{\sqrt{2}}, 1\right), \\ B_1 &= (-\infty, -1] \cup [1, \infty). \end{aligned}$$

Für $f_2(\omega)$ erhalten wir

$$f_2(\omega) = \begin{cases} 0 \ , & 0 \leq f(\omega) < \frac{1}{4} \\ \frac{1}{4} \ , & \frac{1}{4} \leq f(\omega) < \frac{2}{4} \\ \frac{2}{4} \ , & \frac{2}{4} \leq f(\omega) < \frac{3}{4} \\ \frac{3}{4} \ , & \frac{3}{4} \leq f(\omega) < \frac{4}{4} \\ \frac{4}{4} \ , & \frac{4}{4} \leq f(\omega) < \frac{5}{4} \\ \frac{5}{4} \ , & \frac{5}{4} \leq f(\omega) < \frac{6}{4} \\ \frac{6}{4} \ , & \frac{6}{4} \leq f(\omega) < \frac{7}{4} \\ \frac{7}{4} \ , & \frac{7}{4} \leq f(\omega) < \frac{8}{4} \\ 2 \ , & f(\omega) \geq 2 \end{cases}$$

bzw.

$$f_2(\omega) = \sum_{k=1}^{8} \frac{k-1}{4} I_{A_{2,k}}(\omega) + 2 \cdot I_{B_2}(\omega)$$

mit

$$\begin{aligned} A_{2,1} &= \left(-\frac{1}{2}, \frac{1}{2}\right) \\ A_{2,2} &= \left(-\sqrt{\frac{1}{2}}, \frac{1}{2}\right] \cup \left[\frac{1}{2}, \sqrt{\frac{1}{2}}\right) \\ A_{2,3} &= \left(-\frac{\sqrt{3}}{2}, \sqrt{\frac{1}{2}}\right] \cup \left[\sqrt{\frac{1}{2}}, \frac{\sqrt{3}}{2}\right) \\ A_{2,4} &= \left(-1, \frac{\sqrt{3}}{2}\right] \cup \left[\frac{\sqrt{3}}{2}, 1\right) \\ &\vdots \\ B_2 &= \left(-\infty, \sqrt{2}\right] \cup \left[\sqrt{2}, \infty\right). \end{aligned}$$

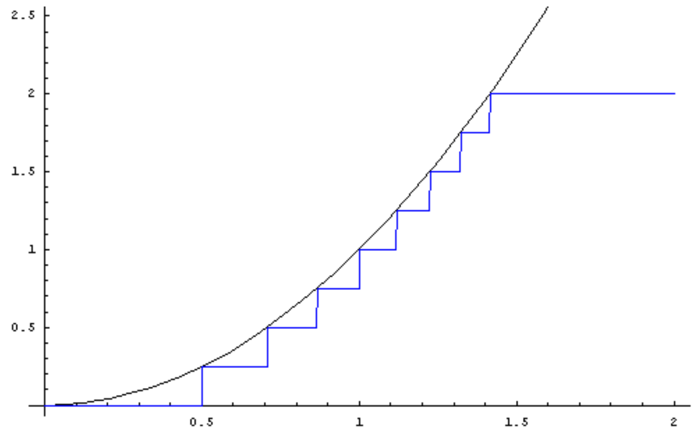


Abbildung 9.1: Näherung einer nichtnegativen messbaren Funktion durch Elementarfunktionen (dargestellt sind $f(\omega) := \omega^2$ und $f_2(\omega)$)

(Siehe auch Mathematica-Notebook zu diesem Beispiel.)

Als nächstes soll das μ–Integral einer nichtnegativen messbaren numerischen Funktion definiert werden. Zur Vorbereitung dienen die Sätze 9.10 und 9.11.

9.10 Satz:
Es sei $(f_n)_{n\in\mathbb{N}}$ eine monoton nichtfallende Folge von Elementarfunktionen von Ω nach $\mathbb{R}^+$. Dann gilt für jede nichtnegative Elementarfunktion h mit $h \leq \sup_n f_n = f$:

$$\int h \; d\mu \leq \sup_{n\in\mathbb{N}} \int f_n \; d\mu \, .$$

Beweis:
Der Fall $h = 0$ ist trivial. Es sei daher $h \neq 0$ mit der Darstellung $h = \sum_{i=1}^m \alpha_i I_{A_i}$ mit $A_i \in \mathfrak{F}$ und den Koeffizienten $\alpha_i \in \mathbb{R}^+$. Für $n \in \mathbb{N}$ und $0 < \beta < 1$ sei

$$B_n := \{\omega \in \Omega \mid \beta h(\omega) \leq f_n(\omega)\}.$$

Da h und f_n messbar sind, gilt aufgrund von Satz 8.12 $B_n \in \mathfrak{F}$ und aufgrund der Definition der Menge B_n, dass $\beta h I_{B_n} \leq f_n$ ist. Wegen Satz 9.7 b) und d) folgt

$$\int \beta h I_{B_n} \; d\mu = \beta \int h I_{B_n} \; d\mu \leq \int f_n \; d\mu \leq \sup_n \int f_n \; d\mu \, . \tag{9.2}$$

Da die Folge $(f_n)_{n\in\mathbb{N}}$ monoton nichtfallend ist und $h \leq \sup_n f_n$ gilt, folgt $B_n \nearrow \Omega$ für $n \to \infty$ und damit $(A_i \cap B_n) \nearrow A_i$ für $n \to \infty$. Aufgrund der Stetigkeit von μ (vgl. Satz 7.5) folgt schließlich $\mu(A_i \cap B_n) \to \mu(A_i)$. Damit wird

$$\int h \; d\mu = \sum_{i=1}^m \alpha_i \mu(A_i) = \lim_{n\to\infty} \sum_{i=1}^m \alpha_i \mu(A_i \cap B_n) = \lim_{n\to\infty} \int h I_{B_n} \; d\mu.$$

Somit ergibt sich

$$\beta \int h \; d\mu = \lim_{n\to\infty} \beta \int h I_{B_n} \; d\mu \overset{(9.2)}{\leq} \sup_n \int f_n \; d\mu.$$

Da $\beta \in (0, 1)$ beliebig gewählt war, folgt die Behauptung. ∎

9.11 Satz:
Es seien $(f_n)_{n\in\mathbb{N}}$ und $(g_n)_{n\in\mathbb{N}}$ zwei monoton nichtfallende Folgen nichtnegativer Elementarfunktionen. Gilt

$$\sup_n f_n(\omega) = \sup_n g_n(\omega) \qquad \forall\ \omega \in \Omega,$$

dann ist

$$\sup_n \int f_n\, d\mu = \sup_n \int g_n\, d\mu\,.$$

Beweis:
Gilt $\sup_n f_n = \sup_n g_n$, dann ist für alle $m \in \mathbb{N}$

$$f_m \le \sup_n g_n \qquad \text{und} \qquad g_m \le \sup_n f_n.$$

Mit Hilfe von Satz 9.10 kann weiter gefolgert werden, dass gilt:

$$\int f_m\ d\mu \le \sup_n \int g_n\ d\mu \qquad \text{und} \qquad \int g_m\ d\mu \le \sup_n \int f_n\ d\mu$$

bzw.

$$\sup_m \int f_m\ d\mu \le \sup_n \int g_n\ d\mu \qquad \text{und} \qquad \sup_m \int g_m\ d\mu \le \sup_n \int f_n\ d\mu,$$

woraus sich die Behauptung unmittelbar ergibt. ■

9.12 Definition (bestimmtes Integral):
Es sei f eine nichtnegative messbare numerische Funktion und $(f_n)_{n\in\mathbb{N}}$ eine monoton nichtfallende Folge nichtnegativer Elementarfunktionen mit $f = \sup_n f_n$. Dann heißt die Zahl

$$\int_\Omega f\ d\mu := \lim_{n\to\infty} \int_\Omega f_n\ d\mu = \sup_n \int_\Omega f_n\ d\mu$$

das <u>bestimmte Integral</u> von f über Ω. Man schreibt auch

$$\int_\Omega f\ d\mu = \int f\ d\mu = \int_\Omega f(\omega)\ d\mu(\omega) = \int_\Omega f(\omega)\mu(d\omega).$$

Die Definition 9.12 nutzt Satz 9.11 aus, wonach das Integral nicht von der speziellen Wahl der Folge $(f_n)_{n\in\mathbb{N}}$ abhängt.

9.13 Satz (Eigenschaften des Integrals):
Es seien f und g nichtnegative messbare Funktionen. Dann gilt:

a) $\displaystyle\int \alpha f\ d\mu = \alpha \int f\ d\mu \quad \forall\ \alpha \in \mathbb{R}^+.$

b) $\displaystyle\int (f+g) d\mu = \int f\ d\mu + \int g\ d\mu.$

c) $f \leq g \implies \int f \, d\mu \leq \int g \, d\mu.$

d) $\int f \, d(\mu + \nu) = \int f \, d\mu + \int f \, d\nu.$

Beweis:
Es seien $(f_n)_{n\in\mathbb{N}}$ und $(g_n)_{n\in\mathbb{N}}$ nichtnegative, monoton nichtfallende Elementarfunktionen mit $\sup\limits_n f_n = f$ und $\sup\limits_n g_n = g$.

a) Es sei $\alpha \in \mathbb{R}^+$. Dann gilt:

$$\begin{aligned} \int \alpha f \, d\mu &= \sup_n \int \alpha f_n \, d\mu && \text{nach Definition 9.12} \\ &= \alpha \sup_n \int f_n \, d\mu && \text{nach Satz 9.7 b)} \\ &= \alpha \int f \, d\mu && \text{nach Definition 9.12.} \end{aligned}$$

b)

$$\begin{aligned} \int (f+g) \, d\mu &= \sup_n \int (f_n + g_n) \, d\mu && \text{nach Definition 9.12} \\ &= \sup_n \left(\int f_n \, d\mu + \int g_n \, d\mu \right) && \text{nach Satz 9.7 c)} \\ &= \lim_{n\to\infty} \left(\int f_n \, d\mu + \int g_n \, d\mu \right) \\ &= \lim_{n\to\infty} \int f_n \, d\mu + \lim_{n\to\infty} \int g_n \, d\mu \\ &= \int f \, d\mu + \int g \, d\mu && \text{nach Definition 9.12.} \end{aligned}$$

c)

$$\begin{aligned} f \leq g &\implies f_m \leq \sup_n g_n && \forall\, m \in \mathbb{N} \\ &\implies \int f_m \, d\mu \leq \sup_n \int g_n \, d\mu && \forall\, m \in \mathbb{N} \quad \text{nach Satz 9.10} \\ &\implies \int f_m \, d\mu \leq \int g \, d\mu && \text{nach Definition 9.12} \\ &\implies \sup_m \int f_m \, d\mu \leq \int g \, d\mu \\ &\implies \int f \, d\mu \leq \int g \, d\mu. \end{aligned}$$

d)

$$\begin{aligned} \int f \, d(\mu+\nu) &= \sup_n \int f_n \, d(\mu+\nu) && \text{nach Definition 9.12} \\ &= \lim_{n\to\infty} \int f_n \, d(\mu+\nu) \\ &= \lim_{n\to\infty} \int f_n \, d\mu + \lim_{n\to\infty} \int f_n \, d\nu && \text{nach Satz 9.7 c)} \\ &= \int f \, d\mu + \int f \, d\nu && \text{nach Definition 9.12.} \end{aligned}$$

■

9.14 Definition (integrierbar, integrabel, μ-integrabel):
Es sei $(\Omega, \mathfrak{F}, \mu)$ ein Maßraum und $f\colon \Omega \longrightarrow \mathbb{R}^+ \cup \{\infty\}$ messbar. Die Funktion f heißt integrierbar, μ–integrabel oder kurz integrabel, wenn

$$\int_\Omega f \; d\mu < \infty$$

gilt.

9.4 Das µ–Integral allgemeiner messbarer numerischer Funktionen

9.15 Definition (Positiv-Teil, Negativ-Teil):
Es sei $f\colon \Omega \to \overline{\mathbb{R}}$ eine numerische Funktion. Der Positiv–Teil f^+ bzw. Negativ–Teil f^- von f ist definiert durch:

$$f^+(\omega) := max\{f(\omega), 0\} \qquad \text{und} \qquad f^-(\omega) := -\, min\{f(\omega), 0\}$$

für alle $\omega \in \Omega$.

Folgerung:

a) Es ist $f^+(\omega) = f(\omega)$, wenn $f(\omega) \geq 0$ ist und $f^+(\omega) = 0$, wenn $f(\omega) \leq 0$ ist.

b) Es ist $f^-(\omega) = -f(\omega)$, wenn $f(\omega) \leq 0$ ist und $f^-(\omega) = 0$, wenn $f(\omega) \geq 0$ ist.

c) Es gilt $f = f^+ - f^-$.

d) Es gilt $|f| = f^+ + f^-$.

9.16 Satz:
Eine numerische Funktion f ist genau dann messbar, wenn ihr Positiv– und ihr Negativ–Teil messbar sind.

Beweis:
Nach Satz 8.13 f) sind mit f auch $\max(f, 0)$ und $\min(f, 0)$ messbar. Umgekehrt ist mit f^+ und f^- auch $f = f^+ - f^-$ messbar. ■

9.17 Definition (quasiintegrabel, μ-integrierbar):
Es sei $(\Omega, \mathfrak{F}, \mu)$ ein Maßraum und $f\colon \Omega \to \overline{\mathbb{R}}$ eine $\mathfrak{F}$–$\overline{\mathfrak{B}}$–messbare Funktion.

a) Ist $\displaystyle\int f^+ \; d\mu < \infty$ oder $\displaystyle\int f^- \; d\mu < \infty$, so nennt man f (μ–)quasiintegrabel.

b) Ist f quasiintegrabel, so ist das μ–Integral von f definiert durch

$$\int f \, d\mu := \int f^+ \, d\mu - \int f^- \, d\mu.$$

c) f heißt μ–integrierbar, wenn sowohl $\int f^+ \, d\mu < \infty$ als auch $\int f^- \, d\mu < \infty$ sind.

9.18 Satz:
Es sei $f\colon \Omega \to \overline{\mathbb{R}}$ eine messbare numerische Funktion, dann sind folgende Aussagen äquivalent:

a) f ist μ–integrierbar.

b) f^+ und f^- sind μ–integrierbar.

c) Es gibt eine μ–integrierbare Funktion g mit $|f| \leq g$.

d) $|f|$ ist μ–integrierbar.

Beweis:
Die Äquivalenz zwischen a) und b) folgt unmittelbar aus der Definition 9.17.

b)$\Longrightarrow$ c)
Sind f^+ und f^- μ–integrierbar, dann sind: f^+, f^- und $g := |f| = f^+ + f^-$ messbar und es ist nach Satz 9.13

$$\int g \, d\mu = \int (f^+ + f^-) \, d\mu = \int f^+ \, d\mu + \int f^- \, d\mu < \infty.$$

Somit ist auch g μ–integrierbar und wegen $|f| = g = f^+ + f^-$ folgt die Behauptung.

c)$\Longrightarrow$ d)
Existiert eine μ–integrierbare Funktion g mit $|f| \leq g$, so gilt aufgrund der Monotonie des Integrals (Satz 9.13):

$$\int |f| \, d\mu \leq \int g \, d\mu.$$

Die Behauptung folgt nun aus der Annahme, dass $\int g \, d\mu < \infty$.

d)$\Longrightarrow$ b)
Es sei nun $|f|$ μ–integrierbar. Es gilt: $f^+ \leq |f|$ und $f^- \leq |f|$. Außerdem sind f^+, f^- und $|f|$ messbar und nichtnegativ. Damit wird

$$\int f^+ \, d\mu \leq \int |f| \, d\mu < \infty \qquad \text{und} \qquad \int f^- \, d\mu \leq \int |f| \, d\mu < \infty.$$

■

9.19 Satz (Eigenschaften des Integrals):
Es sei $(\Omega, \mathfrak{F}, \mu)$ ein Maßraum und $f, g\colon \Omega \to \overline{\mathbb{R}}$ seien messbare numerische und μ– und ν–integrierbare Funktionen. Dann gilt:

a) Für alle $\alpha \in \mathbb{R}$ ist αf μ–integrierbar, und es gilt

$$\int \alpha f \; d\mu = \alpha \int f \; d\mu.$$

b) Falls $f + g$ definiert ist, so ist $f + g$ μ–integrierbar und es gilt:

$$\int (f+g) \; d\mu = \int f \; d\mu + \int g \; d\mu.$$

c) max(f, g) und min(f, g) sind μ–integrierbar.

d) Aus $f \leq g$ folgt $\displaystyle\int f \; d\mu \leq \int g \; d\mu$.

e) Es gilt $\displaystyle\left|\int f \; d\mu\right| \leq \int |f| \; d\mu$.

f) Es gilt $\displaystyle\int f \; d(\mu + \nu) = \int f \; d\mu + \int f \; d\nu$.

Beweis:

a) Es sei $\alpha \in \mathbb{R}$. Für $\alpha \geq 0$ ist $(\alpha f)^+ = \alpha f^+$ und $(\alpha f)^- = \alpha f^-$. Für $\alpha \leq 0$ ist $(\alpha f)^+ = |\alpha| \, f^-$ und $(\alpha f)^- = |\alpha| \, f^+$. Alle diese Funktionen sind nichtnegativ und messbar. Die μ–Integrierbarkeit von f ist nach Satz 9.18 gleichwertig mit der μ–Integrierbarkeit von f^+ und f^-, woraus auch die μ–Integrierbarkeit von $(\alpha f)^+$ und $(\alpha f)^-$ folgt. Damit ist aber auch αf μ–integrierbar. Weiter ist für $\alpha \geq 0$

$$\begin{aligned}\int \alpha f \; d\mu \;&=\; \int (\alpha f)^+ \; d\mu - \int (\alpha f)^- \; d\mu \;=\; \int \alpha f^+ \; d\mu - \int \alpha f^- \; d\mu \\ &=\; \alpha \int f^+ \; d\mu - \alpha \int f^- \; d\mu \;=\; \alpha \left(\int f^+ \; d\mu - \int f^- \; d\mu \right) \;=\; \alpha \int f \; d\mu.\end{aligned}$$

Im Fall $\alpha < 0$ gilt $(\alpha f)^+ = |\alpha| \cdot f^-$ und $(\alpha f)^- = |\alpha| \cdot f^+$. Aufgrund von Satz 9.13 ist deshalb

$$\int_\Omega \alpha f^+ \, d\mu = |\alpha| \int_\Omega f^- \, d\mu \qquad \text{bzw.} \qquad \int_\Omega \alpha f^- \, d\mu = |\alpha| \int_\Omega f^+ \, d\mu \, .$$

Damit wird

$$\int_\Omega \alpha f \, d\mu = |\alpha| \left(\int_\Omega f^- \, d\mu - \int_\Omega f^+ \, d\mu \right) = \alpha \left(\int_\Omega f^+ \, d\mu - \int_\Omega f^- \, d\mu \right) = \alpha \int_\Omega f \, d\mu \, .$$

b) Nach Satz 9.18 sind $|f|$ und $|g|$ integrierbar. Wegen $|f+g| \leq |f| + |g|$ ist deshalb auch $f+g$ integrierbar. Außerdem ist $f+g = (f^+ + g^+) - (f^- + g^-)$, so dass gilt

$$\begin{aligned}
\int (f+g)\, d\mu &= \int (f^+ + g^+)\, d\mu - \int (f^- + g^-)\, d\mu \\
&= \int f^+\, d\mu + \int g^+\, d\mu - \int f^-\, d\mu - \int g^-\, d\mu \\
&= \int f^+\, d\mu - \int f^-\, d\mu + \int g^+\, d\mu - \int g^-\, d\mu \\
&= \int f\, d\mu + \int g\, d\mu.
\end{aligned}$$

c) Mit f und g sind auch $|f|$ und $|g|$ integrierbar (siehe Satz 9.18). Also ist auch $|f| + |g|$ μ–integrierbar. Da aber

$$|\max(f,g)| \leq |f| + |g| \qquad \text{und} \qquad |\min(f,g)| \leq |f| + |g|$$

gilt, folgt die Behauptung jetzt unmittelbar aus Satz 9.18 c).

d) Aus $f \leq g$ folgen die beiden Ungleichungen $f^+ \leq g^+$ und $g^- \leq f^-$. Damit wird:

$$\int f^+\, d\mu \leq \int g^+\, d\mu \qquad \text{und} \qquad \int g^-\, d\mu \leq \int f^-\, d\mu$$

und weiter

$$\int f\, d\mu = \int f^+\, d\mu - \int f^-\, d\mu \leq \int g^+\, d\mu - \int g^-\, d\mu = \int g\, d\mu.$$

e) Es gilt:

$$\begin{aligned}
\left|\int f\, d\mu\right| &= \left|\int f^+\, d\mu - \int f^-\, d\mu\right| \leq \left|\int f^+\, d\mu\right| + \left|\int f^-\, d\mu\right| \\
&= \int f^+\, d\mu + \int f^-\, d\mu = \int (f^+ + f^-)\, d\mu = \int |f|\, d\mu.
\end{aligned}$$

f)

$$\begin{aligned}
\int f\, d(\mu+\nu) &= \int f^+\, d(\mu+\nu) + \int f^-\, d(\mu+\nu) \\
&= \int f^+\, d\mu + \int f^+\, d\nu + \int f^-\, d\mu + \int f^-\, d\nu \\
&= \int f\, d\mu + \int f\, d\nu.
\end{aligned}$$

■

9.5 Erwartungswert und Varianz einer reellwertigen Zufallsvariable

9.20 Definition (Erwartungswert):
Es seien $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $X \colon \Omega \to \mathbb{R}$ eine Zufallsvariable. Ist X P–integrierbar, dann heißt

$$\boldsymbol{E}_P[X] \ := \ \int\limits_\Omega X(\omega)\ dP(\omega) \ = \ \int X \ dP$$

der Erwartungswert der Zufallsvariable X.

9.21 Beispiel:
Im Folgenden soll der Erwartungswert einiger Verteilungen berechnet werden:

- **Poisson–Verteilung:**
  Es seien $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $X \colon \Omega \to \mathbb{R}$ eine Zufallsvariable mit Verteilungsfunktion

$$F(x) := \begin{cases} \sum\limits_{j=0}^{\lfloor x \rfloor} \dfrac{\lambda^j}{j!} e^{-\lambda} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0. \end{cases}$$

  Es folgt zunächst

$$\begin{aligned} F\left(\frac{k}{2^n} - 0\right) - F\left(\frac{k-1}{2^n} - 0\right) &= \sum_{j=0}^{\lfloor \frac{k}{2^n} - 0 \rfloor} \frac{\lambda^j}{j!} e^{-\lambda} - \sum_{j=0}^{\lfloor \frac{k-1}{2^n} - 0 \rfloor} \frac{\lambda^j}{j!} e^{-\lambda} \\ &\overset{(*)}{=} \sum_{j=0}^{\lfloor \frac{k-1}{2^n} \rfloor} \frac{\lambda^j}{j!} e^{-\lambda} - \sum_{j=0}^{\lfloor \frac{k-2}{2^n} \rfloor} \frac{\lambda^j}{j!} e^{-\lambda} \\ &= \begin{cases} 0 & \text{für } \lfloor \frac{k-1}{2^n} \rfloor = \lfloor \frac{k-2}{2^n} \rfloor \\ \dfrac{\lambda^{\lfloor \frac{k-1}{2^n} \rfloor}}{(\lfloor \frac{k-1}{2^n} \rfloor)!} e^{-\lambda} & \text{für } \lfloor \frac{k-1}{2^n} \rfloor > \lfloor \frac{k-2}{2^n} \rfloor. \end{cases} \end{aligned}$$

  ($(*)$: $\left\lfloor \dfrac{k}{2^n} - 0 \right\rfloor = \lim\limits_{x \nearrow \frac{k}{2^n}} \lfloor x \rfloor = \left\lfloor \dfrac{k-1}{2^n} \right\rfloor$ für $n \in \mathbb{N}$ und $k \in \mathbb{N}_0$.)

  Es ist

$$\left\lfloor \frac{k-1}{2^n} \right\rfloor > \left\lfloor \frac{k-2}{2^n} \right\rfloor \iff 2^n | (k-1) \iff k = j \cdot 2^n + 1, \ j \in \mathbb{N}_0.$$

  Es soll $\int X(\omega)\, dP_F(\omega) = \lim\limits_{n \to \infty} \int X_n(\omega)\, dP_F(\omega)$ berechnet werden, wobei P_F das durch F

induzierte Maß auf $(\mathbb{R}, \mathfrak{B})$ und X_n die in Satz 9.8 verwendeten Funktionen bezeichnen.

$$\begin{aligned}
\int X_n(\omega) dP_F(\omega) &= \sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n} P_F\left(\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)\right) + n \cdot P_F([n,\infty)) \\
&= \frac{1}{2^n} \sum_{k=1}^{n\cdot 2^n} (k-1)\left[F\left(\frac{k}{2^n} - 0\right) - F\left(\frac{k-1}{2^n} - 0\right)\right] \\
&\quad + n \cdot (1 - F(n-0)) \\
&= \frac{1}{2^n} \sum_{\substack{k=2 \\ 2^n | (k-1)}}^{n\cdot 2^n} (k-1) \frac{\lambda^{\frac{k-1}{2^n}}}{(\frac{k-1}{2^n})!} e^{-\lambda} + n \cdot (1 - F(n-0)) \\
&= \frac{1}{2^n} \sum_{j=1}^{n-1} j \cdot 2^n \frac{\lambda^j}{j!} e^{-\lambda} + n \cdot (1 - F(n-0)) \\
&= \lambda \sum_{j=1}^{n-1} \frac{\lambda^{j-1}}{(j-1)!} e^{-\lambda} + n \cdot \left(1 - \sum_{j=0}^{\lfloor n-0 \rfloor} \frac{\lambda^j}{j!} e^{-\lambda}\right) \\
&= \underbrace{\lambda \sum_{j=1}^{n-1} \frac{\lambda^{j-1}}{(j-1)!}}_{\xrightarrow[n\to\infty]{} e^\lambda} e^{-\lambda} + \underbrace{n \cdot \left(1 - \sum_{j=0}^{n-1} \frac{\lambda^j}{j!} e^{-\lambda}\right)}_{\xrightarrow[n\to\infty]{} 0} \xrightarrow[n\to\infty]{} \lambda.
\end{aligned}$$

- **Exponentialverteilung:**
  Es seien $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $X \colon \Omega \to \mathbb{R}$ eine Zufallsvariable mit Verteilungsfunktion

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

  Es soll $\int X(\omega)\, dP_F(\omega) = \lim\limits_{n\to\infty} \int X_n(\omega)\, dP_F(\omega)$ berechnet werden, wobei P_F das durch F induzierte Maß auf $(\mathbb{R}, \mathfrak{B})$ und X_n die in Satz 9.8 verwendeten Funktionen bezeichnen.

$$\begin{aligned}
\int X_n(\omega) dP_F(\omega) &= \sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n} P_F\left(\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)\right) + n \cdot P_F([n,\infty)) \\
&= \frac{1}{2^n} \sum_{k=1}^{n\cdot 2^n} (k-1)\left[F\left(\frac{k}{2^n} - 0\right) - F\left(\frac{k-1}{2^n} - 0\right)\right] \\
&\quad + n \cdot (1 - F(n-0)) \\
&= \frac{1}{2^n}\Bigg[1 \cdot \left(F\left(\frac{2}{2^n}\right) - F\left(\frac{1}{2^n}\right)\right) + 2 \cdot \left(F\left(\frac{3}{2^n}\right) - F\left(\frac{2}{2^n}\right)\right) + \\
&\quad + 3 \cdot \left(F\left(\frac{4}{2^n}\right) - F\left(\frac{3}{2^n}\right)\right) + \ldots \\
&\quad + (n2^n - 2) \cdot \left(F\left(\frac{n2^n - 1}{2^n}\right) - F\left(\frac{n2^n - 2}{2^n}\right)\right) \\
&\quad + (n2^n - 1) \cdot \left(F\left(\frac{n2^n}{2^n}\right) - F\left(\frac{n2^n - 1}{2^n}\right)\right)\Bigg] + n \cdot (1 - F(n))
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2^n}\Big[-F\left(\frac{1}{2^n}\right) - F\left(\frac{2}{2^n}\right) - \cdots - F\left(\frac{n2^n-1}{2^n}\right) + \\
&\quad + (n2^n-1)\cdot F\left(\frac{n2^n}{2^n}\right)\Big] + n\cdot(1-F(n)) \\
&= -\frac{1}{2^n}\sum_{k=1}^{n\cdot 2^n-1} F\left(\frac{k}{2^n}\right) + \left(n-\frac{1}{2^n}\right)\cdot F\left(\frac{n2^n}{2^n}\right) + n\cdot(1-F(n)) \\
&= -\frac{1}{2^n}\sum_{k=1}^{n\cdot 2^n-1}\left(1-e^{-\lambda k/2^n}\right) + \left(n-\frac{1}{2^n}\right)\cdot\left(1-e^{-\lambda n}\right) + n\cdot e^{-\lambda n} \\
&= -\frac{n2^n-1}{2^n} + \frac{1}{2^n}\sum_{k=1}^{n\cdot 2^n-1}\left(e^{-\lambda/2^n}\right)^k + \left(n-\frac{1}{2^n}\right) \\
&\quad -\left(n-\frac{1}{2^n}\right)e^{-\lambda n} + n\cdot e^{-\lambda n} \\
&= \frac{1}{2^n}\cdot\sum_{k=1}^{n2^n-1}\left(e^{-\lambda/2^n}\right)^k + \frac{1}{2^n}\cdot e^{-\lambda n} \\
&= \frac{1}{2^n}\cdot\frac{1-(e^{-\lambda/2^n})^{n2^n}}{1-e^{-\lambda/2^n}} - \frac{1}{2^n} + \frac{1}{2^n}\cdot e^{-\lambda n}.
\end{aligned}
$$

Folglich gilt:

$$
\begin{aligned}
\int X(\omega)\,dP_F(\omega) &= \lim_{n\to\infty}\int X_n(\omega)dP_F(\omega) = \lim_{n\to\infty}\frac{1}{2^n}\cdot\frac{1-e^{-\lambda n}}{1-e^{-\lambda/2^n}} - \frac{1}{2^n} + \frac{1}{2^n}\cdot e^{-\lambda n} \\
&= \lim_{n\to\infty}\frac{1}{2^n}\cdot\frac{1-e^{-\lambda n}}{1-e^{-\lambda/2^n}}.
\end{aligned}
$$

Aus

$$\lim_{x\to 0}\frac{e^x-1}{x} = 1$$

folgt

$$\lim_{n\to\infty}\frac{1-e^{-\lambda/2^n}}{\lambda/2^n} = 1 \quad \text{bzw.} \quad \lim_{n\to\infty} 2^n\cdot(1-e^{-\lambda/2^n}) = \lambda\,.$$

Damit wird

$$\mathbf{E}_P[X] = \int X\,dP_F = \lim_{n\to\infty}\int X_n dP_F = \lim_{n\to\infty}\frac{1}{2^n\cdot(1-e^{-\lambda/2^n})}\cdot(1-e^{-\lambda n}) = \frac{1}{\lambda}\,.$$

- **Rechteckverteilung auf** $[-1,1]$**:**
  Es seien $(\Omega,\mathfrak{F},P)$ ein Wahrscheinlichkeitsraum und $X\colon\Omega\to\mathbb{R}$ eine Zufallsvariable mit Verteilungsfunktion

  $$F(x) := \begin{cases} 0, & x\le -1 \\ \frac{x+1}{2}, & -1\le x\le 1 \\ 1, & x>1. \end{cases}$$

  Die Darstellung

  $$\int X(\omega)\ dP_F(\omega) = \int X^+(\omega)\ dP_F(\omega) - \int X^-(\omega)\ dP_F(\omega)$$

legt nahe, Positiv– und Negativteil separat zu betrachten. X^+ und X^- werden dabei durch die Funktionen X_n^+ und X_n^- wie in Satz 9.8 approximiert:

$$X_n^{\pm} := \sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n} I_{A_{n,k}} + n \cdot I_{B_n}$$

mit

$$A_{n,k} := \left(X^{\pm}\right)^{-1}\left(\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)\right) \qquad \text{und} \qquad B_n := \left(X^{\pm}\right)^{-1}([n,\infty)).$$

1.) Positivteil: Für X^+ ergeben sich:

$$A_{n,k} := \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right) \qquad \text{und} \qquad B_n := [n,\infty).$$

Damit gilt:

$$\begin{aligned}
\int X_n^+(\omega)\, dP_F(\omega) &= \sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n} P_F\left(\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)\right) + n \underbrace{P_F([n,\infty))}_{=0 \text{ für } n\geq 1} \\
&= \sum_{\substack{k=1\\k\leq 2^n}}^{n\cdot 2^n} \frac{k-1}{2^n} P_F\left(\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)\right) \\
&\quad + \sum_{\substack{k=1\\k>2^n}}^{n\cdot 2^n} \frac{k-1}{2^n} \underbrace{P_F\left(\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)\right)}_{=F\left(\frac{k}{2^n}-0\right)-F\left(\frac{k-1}{2^n}-0\right)=1-1=0} \\
&= \frac{1}{2^n}\sum_{k=1}^{2^n}(k-1)\left(F\left(\frac{k}{2^n}-0\right) - F\left(\frac{k-1}{2^n}-0\right)\right) \\
&= \frac{1}{2^n}\sum_{k=1}^{2^n}(k-1)\underbrace{\left(\frac{1}{2}\left(\frac{k}{2^n}+1\right) - \frac{1}{2}\left(\frac{k-1}{2^n}+1\right)\right)}_{=\frac{1}{2^{n+1}}} \\
&= \frac{1}{2^{2n+1}}\sum_{k=1}^{2^n}(k-1) \;=\; \frac{1}{2^{2n+1}} \cdot \frac{2^n(2^n-1)}{2} \\
&= \frac{1}{4} - \frac{1}{2^{n+2}} \xrightarrow[n\to\infty]{} \frac{1}{4}.
\end{aligned}$$

2.) Negativteil: Für X^- ergeben sich:

$$A_{n,k} := \left(-\frac{k}{2^n}, -\frac{k-1}{2^n}\right] \qquad \text{und} \qquad B_n := (-\infty, -n].$$

Damit gilt:

$$\begin{aligned}
\int X_n^-(\omega)\ dP_F(\omega) &= \sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n} P_F\left(\left[-\frac{k}{2^n}, -\frac{k-1}{2^n}\right)\right) + n\,\underbrace{P_F((-\infty,-n])}_{=0 \text{ für } n<-1} \\
&= \sum_{\substack{k=1\\ k\leq 2^n}}^{n\cdot 2^n} \frac{k-1}{2^n} P_F\left(\left[-\frac{k}{2^n}, -\frac{k-1}{2^n}\right)\right) \\
&\quad + \sum_{\substack{k=1\\ k>2^n}}^{n\cdot 2^n} \frac{k-1}{2^n} \underbrace{P_F\left(\left[-\frac{k}{2^n}, -\frac{k-1}{2^n}\right)\right)}_{=F\left(-\frac{k-1}{2^n}-0\right)-F\left(-\frac{k}{2^n}-0\right)=0-0=0} \\
&= \frac{1}{2^n}\sum_{k=1}^{2^n}(k-1)\left(F\left(-\frac{k-1}{2^n}-0\right) - F\left(-\frac{k}{2^n}-0\right)\right) \\
&= \frac{1}{2^n}\sum_{k=1}^{2^n}(k-1)\underbrace{\left(\frac{1}{2}\left(-\frac{k-1}{2^n}+1\right) - \frac{1}{2}\left(-\frac{k}{2^n}+1\right)\right)}_{=\frac{1}{2^{n+1}}} \\
&= \frac{1}{2^{2n+1}}\sum_{k=1}^{2^n}(k-1) \;=\; \frac{1}{2^{2n+1}}\cdot\frac{2^n(2^n-1)}{2} \\
&= \frac{1}{4} - \frac{1}{2^{n+2}} \xrightarrow[n\to\infty]{} \frac{1}{4}.
\end{aligned}$$

Damit ergibt sich insgesamt:

$$\mathbf{E}_P[X] = \int X(\omega)\ dP_F(\omega) = \int X^+(\omega)\ dP_F(\omega) - \int X^-(\omega)\ dP_F(\omega) = \frac{1}{4} - \frac{1}{4} = 0.$$

Für die Praxis ist es wichtig, einfacher zu handhabende Verfahren zur Berechnung des Erwartungswertes einer reellwertigen Zufallsvariable zu haben. Der nachfolgende Satz zeigt, dass sich der Erwartungswert einer reellwertigen Zufallsvariable als uneigentliches Riemann–Integral aus der zugehörigen Verteilungsfunktion berechnen lässt. Aus dieser Darstellung folgt insbesondere, dass der Erwartungswert nicht von der speziellen Gestalt der Zufallsvariable, sondern nur von der speziellen Gestalt der Verteilungsfunktion abhängt.

9.22 Satz:
Es bezeichne F die Verteilungsfunktion von X bezüglich P. Ist X bezüglich P integrierbar, so gilt:

$$\boldsymbol{E}_P[X] \;=\; \int_0^\infty (1-F(x))\ dx - \int_{-\infty}^0 F(x)\ dx.$$

Existiert umgekehrt jedes der beiden letzteren Integrale, so ist X P–integrierbar, wobei der Erwartungswert wiederum durch diese Beziehung gegeben ist.

Beweis:
Idee: Betrachte den Positiv– und Negativteil des Erwartungswertes

$$\int X\ dP \ = \ \int X^+\ dP - \int X^-\ dP$$

und zeige

1. $\int X^- dP = \int\limits_{-\infty}^{0} F(x)dx$ und

2. $\int X^+\ dP \ = \ \int\limits_{0}^{\infty} (1 - F(x))\ dx.$

Zu 1.: Wir berechnen zunächst die im Beweis von Satz 9.8 definierte Folge $\{X_n^-\}$ von nichtnegativen Elementarfunktionen mit $X_n^- \nearrow X^-$. Aufgrund von Satz 9.8 und der Beziehung

$$X^-(\omega) := -X(\omega) \qquad \text{für} \qquad X(\omega) \leq 0$$

kann

$$X_n^-(\omega) := \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \cdot I_{A_{n,k}}(\omega) + n \cdot I_{B_n}(\omega) \qquad (n = 1, 2, \dots)$$

gewählt werden, wobei

$$\begin{aligned} A_{n,k} \ &:= \ \left\{\omega \mid \frac{k-1}{2^n} \leq X^-(\omega) < \frac{k}{2^n}\right\} = \left\{\omega \mid \frac{k-1}{2^n} \leq -X(\omega) < \frac{k}{2^n}\right\} \\ &= \ \left\{\omega \mid -\frac{k}{2^n} < X(\omega) \leq -\frac{k-1}{2^n}\right\} \end{aligned}$$

und

$$\begin{aligned} B_n \ &:= \ \{\omega \mid n \leq X^-(\omega) < \infty\} = \{\omega \mid n \leq -X(\omega) < \infty\} \\ &= \ \{\omega \mid -\infty < X(\omega) \leq -n\} \end{aligned}$$

ist. Damit wird

$$\begin{aligned} \int X_n^-\ dP \ &= \ \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \cdot P\left(-\frac{k}{2^n} < X \leq -\frac{k-1}{2^n}\right) + n \cdot P(-\infty < X \leq -n) \\ &= \ \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \cdot P_F\left(\left(-\frac{k}{2^n}, -\frac{k-1}{2^n}\right]\right) + n \cdot P_F((-\infty, -n]) \\ &= \ \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \cdot \left(F\left(-\frac{k-1}{2^n}\right) - F\left(-\frac{k}{2^n}\right)\right) + n \cdot F(-n) \\ &= \ \frac{1}{2^n} \cdot \left(F\left(-\frac{1}{2^n}\right) - F\left(-\frac{2}{2^n}\right)\right) + \frac{2}{2^n} \cdot \left(F\left(-\frac{2}{2^n}\right) - F\left(-\frac{3}{2^n}\right)\right) \\ &\quad + \dots \end{aligned}$$

$$
\begin{aligned}
&+\frac{n\cdot 2^n-1}{2^n}\cdot\left(F\left(-\frac{n\cdot 2^n-1}{2^n}\right)-F\left(-\frac{n\cdot 2^n}{2^n}\right)\right)+n\cdot F(-n)\\
=\;&\frac{1}{2^n}\Bigg(F\left(-\frac{1}{2^n}\right)+F\left(-\frac{2}{2^n}\right)+\dots\\
&+F\left(-\frac{n\cdot 2^n-1}{2^n}\right)+F\left(-\frac{n\cdot 2^n}{2^n}\right)\Bigg)\\
=\;&\frac{1}{2^n}\sum_{k=1}^{n2^n}F\left(-\frac{k}{2^n}\right).
\end{aligned}
$$

Da F monoton nichtfallend ist, gilt außerdem

$$\int_{-n}^{0} F(x)\;dx \geq \frac{1}{2^n}\sum_{k=1}^{n\cdot 2^n}F\left(-\frac{k}{2^n}\right)\geq \int_{-n-\frac{1}{2^n}}^{-\frac{1}{2^n}} F(x)\;dx.$$

Durch Grenzübergang $n\to\infty$, erhält man

$$\int X^-\;dP=\lim_{n\to\infty}\int X_n^-\;dP=\lim_{n\to\infty}\frac{1}{2^n}\sum_{k=1}^{n2^n}F\left(-\frac{k}{2^n}\right)=\int_{-\infty}^{0}F(x)\;dx. \tag{9.3}$$

Zu 2.: Wir betrachten eine Folge $\{X_n^+\}$ von nichtnegativen Elementarfunktionen mit $X_n^+\nearrow X^+$ und wie im Beweis von Satz 9.8 setzen wir:

$$X_n^+(\omega):=\sum_{k=1}^{n2^n}\frac{k-1}{2^n}\cdot I_{A_{n,k}}(\omega)+n\cdot I_{B_n}(\omega)\qquad(n=1,2,\dots),$$

wobei

$$A_{n,k}:=\left\{\omega\mid\frac{k-1}{2^n}\leq X^+(\omega)<\frac{k}{2^n}\right\}\qquad\text{und}\qquad B_n:=\{\omega\mid n\leq X^+(\omega)<\infty\}$$

sind. Damit wird

$$
\begin{aligned}
\int X_n^+\;dP\;&=\;\sum_{k=1}^{n2^n}\frac{k-1}{2^n}\cdot P_F\left(\left[\frac{k-1}{2^n},\frac{k}{2^n}\right)\right)+n\cdot P_F([n,\infty))\\
&=\;\sum_{k=1}^{n2^n}\frac{k-1}{2^n}\left[F\left(\frac{k}{2^n}-0\right)-F\left(\frac{k-1}{2^n}-0\right)\right]+n\cdot(1-F(n-0))\\
&=\;-\sum_{k=1}^{n2^n}\frac{k-1}{2^n}\left(1-F\left(\frac{k}{2^n}-0\right)\right)\\
&\quad+\sum_{k=1}^{n2^n}\frac{k-1}{2^n}\left(1-F\left(\frac{k-1}{2^n}-0\right)\right)+n\cdot(1-F(n-0))\\
&=\;\frac{1}{2^n}\sum_{k=1}^{n2^n}\left(1-F\left(\frac{k}{2^n}-0\right)\right).
\end{aligned}
$$

Da $1 - F(x)$ monoton nichtwachsend ist, gilt weiter

$$\int\limits_0^n (1 - F(x))\, dx \geq \frac{1}{2^n} \sum_{k=1}^{n2^n} \left(1 - F\left(\frac{k}{2^n} - 0\right)\right) = \int X_n^+ \, dP \geq \int\limits_{1/2^n}^{n+1/2^n} (1 - F(x))\, dx.$$

Durch den Grenzübergang $n \to \infty$ erhält man

$$\int X^+ \, dP = \lim_{n\to\infty} \int X_n^+ \, dP = \lim_{n\to\infty} \frac{1}{2^n} \sum_{k=1}^{n2^n} \left(1 - F\left(\frac{k}{2^n} - 0\right)\right) = \int\limits_0^\infty (1 - F(x))\, dx. \tag{9.4}$$

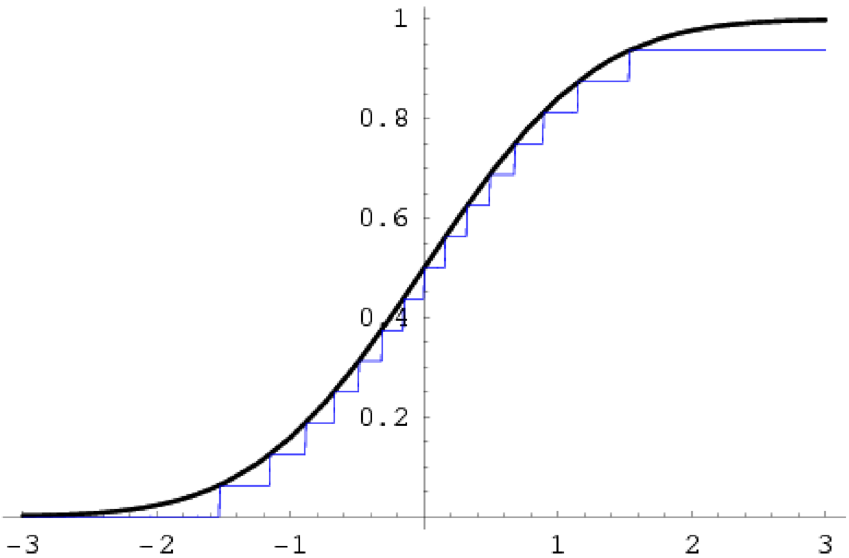


Abbildung 9.2: Approximation der Verteilungsfunktion der Standardnormalverteilung (dicke Linie) durch Treppenfunktionen; hier ist X_n für $n = 4$ dargestellt.

(Siehe auch Mathematica-Notebook zu dieser Approximation.)

Die Umkehrung ist eine unmittelbare Konsequenz der Beziehungen (9.3) und (9.4). ■

9.23 Beispiel:
Im Folgenden soll der Erwartungswert einiger Verteilungen nach dem Verfahren aus 9.22 berechnet werden:

- **Poisson–Verteilung:**
  Es seien $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $X\colon \Omega \to \mathbb{R}$ eine Zufallsvariable mit Verteilungsfunktion

$$F(x) := \begin{cases} \displaystyle\sum_{j=0}^{\lfloor x \rfloor} \frac{\lambda^j}{j!} e^{-\lambda} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0. \end{cases}$$

Dann gilt:

$$\begin{aligned}\mathbf{E}_P[X] &= \int_0^\infty (1-F(x))\,dx - \underbrace{\int_{-\infty}^0 F(x)\,dx}_{=0} = \int_0^\infty 1-\sum_{j=0}^{\lfloor x\rfloor}\frac{\lambda^j}{j!}e^{-\lambda}\,dx \\ &= \sum_{n=0}^\infty \left(1-\sum_{j=0}^n \frac{\lambda^j}{j!}e^{-\lambda}\right)\cdot \underbrace{P_F([n,n+1))}_{=1}.\end{aligned}$$

Betrachte zunächst die endliche Summe:

$$\begin{aligned}\sum_{n=0}^N\left(1-\sum_{j=0}^n\frac{\lambda^j}{j!}e^{-\lambda}\right) &= 1-\sum_{j=0}^0\frac{\lambda^j}{j!}e^{-\lambda}+1-\sum_{j=0}^1\frac{\lambda^j}{j!}e^{-\lambda}+\ldots+1-\sum_{j=0}^N\frac{\lambda^j}{j!}e^{-\lambda} \\ &= N+1-\sum_{j=0}^N (N+1-j)\frac{\lambda^j}{j!}e^{-\lambda} \\ &= (N+1)\left(1-\sum_{j=0}^N\frac{\lambda^j}{j!}e^{-\lambda}\right)+\sum_{j=0}^N\frac{\lambda^j}{(j-1)!}e^{-\lambda} \\ &= \frac{N+1}{e^\lambda}\left(e^\lambda-\sum_{j=0}^N\frac{\lambda^j}{j!}\right)+\lambda\sum_{j=0}^{N-1}\frac{\lambda^j}{j!}e^{-\lambda}\xrightarrow[N\to\infty]{} 0+\lambda\cdot 1=\lambda.\end{aligned}$$

- **Exponentialverteilung:**
  Es seien $(\Omega,\mathfrak{F},P)$ ein Wahrscheinlichkeitsraum und $X\colon\Omega\to\mathbb{R}$ eine Zufallsvariable mit Verteilungsfunktion
  $$F(x) := \begin{cases}1-e^{-\lambda x} & \text{für } x\ge 0\\ 0 & \text{für } x<0.\end{cases}$$
  Dann gilt:
  $$\begin{aligned}\mathbf{E}_P[X] &= \int_0^\infty (1-F(x))\,dx - \underbrace{\int_{-\infty}^0 F(x)\,dx}_{=0} \\ &= \int_0^\infty e^{-\lambda x}\,dx = \left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^\infty = \frac{1}{\lambda}.\end{aligned}$$

- **Rechteckverteilung auf $[-1,1]$:**
  Es seien $(\Omega,\mathfrak{F},P)$ ein Wahrscheinlichkeitsraum und $X\colon\Omega\to\mathbb{R}$ eine Zufallsvariable mit Verteilungsfunktion
  $$F(x) := \begin{cases}0, & x\le -1\\ \frac{x+1}{2}, & -1\le x\le 1\\ 1, & x>1.\end{cases}$$

Dann folgt:

$$\begin{aligned}
\mathbf{E}_P[X] &= \int_0^\infty (1 - F(x))\ dx - \int_{-\infty}^0 F(x)\ dx \\
&= \int_0^1 (1 - F(x))\ dx + \int_1^\infty (1 - \underbrace{F(x)}_{=1})\ dx - \int_{-1}^0 F(x)\ dx \\
&= \int_0^1 \left(1 - \frac{1}{2}(x+1)\right)\ dx - \int_{-1}^0 \frac{1}{2}(x+1)\ dx \\
&= \left[\frac{1}{2}x - \frac{1}{4}x^2\right]_0^1 - \frac{1}{2}\left[\frac{1}{2}x^2 + x\right]_{-1}^0 \\
&= \frac{1}{2} - \frac{1}{4} - \frac{1}{2}\left(-\frac{1}{2} + 1\right) = 0.
\end{aligned}$$

In allen Fällen erkennt man die Übereinstimmung mit den Ergebnissen aus Beispiel 9.21 und die Effizienz der alternativen Vorgehensweise.
Bereits in der Vorbemerkung zu Kapitel 9 wurde nahegelegt, den Erwartungswert einer Zufallsvariable X mit einer Dichte $f(x)$ als

$$\mathrm{E}[X] = \int x f(x)\ dx$$

festzulegen. Für den speziellen Fall der Exponentialverteilung wurde die Kompatibilität dieses Ansatzes mit der allgemeinen Definition des Erwartungswertes aus Satz 9.20 bereits überprüft. Offensichtlich handelt es sich hier um ein allgemeingültiges Resultat, wie der nachfolgende Satz bestätigt.

9.24 Satz:
Wenn in der Situation von Satz 9.22 die Verteilungsfunktion F stetig ist und eine Dichte f hat, so ist X genau dann P–integrierbar, wenn

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

gilt. In diesem Fall ist

$$\mathbf{E}_P\,[X] = \int_{-\infty}^{+\infty} x f(x)\ dx.$$

Beweis:
Für den Beweis wird wieder die Beziehung

$$\mathbf{E}_P\,[X] = \int X^+\, dP - \int X^-\, dP$$

ausgenutzt. Zunächst wird der Positivteil bearbeitet, d.h. es wird $\int X^+ \, dP = \int\limits_0^{+\infty} xf(x)\ dx$ gezeigt.

Für $X_n^+,\ n = 1, 2, \ldots$ wie im Beweis von Satz 9.22 werden dazu

1. $\int X_n^+ \, dP \leq \int\limits_0^{+\infty} xf(x)\ dx$ und

2. $\int X_n^+ \, dP \geq \int\limits_0^{+\infty} xf(x)\ dx - \frac{1}{2^n}$

gezeigt. Damit wird $\int X_n^+ \, dP$ von oben und unten eingeschlossen und es ergibt sich

$$\int X^+ \, dP = \lim_{n\to\infty} \int X_n^+ \, dP = \int\limits_0^{+\infty} xf(x)\ dx.$$

Zu 1.: Es gilt für $n = 1, 2, \ldots$:

$$\begin{aligned}
\int X_n^+\ dP &= \sum_{k=1}^{n2^n} \frac{k-1}{2^n} P_F\left(\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)\right) + n \cdot P_F([n, \infty)) \\
&= \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \int\limits_{(k-1)/2^n}^{k/2^n} f(x)\ dx + n \int\limits_n^{\infty} f(x)\ dx \\
&= \sum_{k=1}^{n2^n} \int\limits_{(k-1)/2^n}^{k/2^n} \frac{k-1}{2^n} f(x)\ dx + \int\limits_n^{\infty} nf(x)\ dx \\
&\leq \sum_{k=1}^{n2^n} \int\limits_{(k-1)/2^n}^{k/2^n} xf(x)\ dx + \int\limits_n^{\infty} xf(x)\ dx \\
&= \int\limits_0^{\infty} xf(x)\ dx.
\end{aligned}$$

Zu 2.: Andererseits ist für $n = 1, 2, \ldots$:

$$\begin{aligned}
\int X_n^+\ dP &= \sum_{k=1}^{n2^n} \int\limits_{(k-1)/2^n}^{k/2^n} \frac{k-1}{2^n} f(x)\ dx + n \int_n^{\infty} f(x)\ dx \\
&= \sum_{k=1}^{n2^n} \int\limits_{(k-1)/2^n}^{k/2^n} xf(x)\ dx - \sum_{k=1}^{n2^n} \int\limits_{(k-1)/2^n}^{k/2^n} \left(x - \frac{k-1}{2^n}\right) f(x)\ dx
\end{aligned}$$

$$
\begin{aligned}
& \underbrace{+n \int_{n}^{\infty} f(x)\, dx}_{\geq 0} \\
\geq \quad & \int_0^n x f(x)\, dx - \sum_{k=1}^{n2^n} \int_{(k-1)/2^n}^{k/2^n} \left(x - \frac{k-1}{2^n} \right) f(x)\, dx.
\end{aligned}
$$

Da aber für $x \in \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right]$ und für alle n und alle k gilt: $x - (k-1)/2^n \leq 1/2^n$, kann gefolgert werden

$$
\begin{aligned}
\int X_n^+ \, dP \quad & \geq \quad \int_0^n x f(x)\, dx - \frac{1}{2^n} \sum_{k=1}^{n2^n} \int_{(k-1)/2^n}^{k/2^n} f(x)\, dx \\
& = \quad \int_0^n x f(x)\, dx - \frac{1}{2^n} \underbrace{\int_0^n f(x)\, dx}_{\leq 1} \\
& \geq \quad \int_0^n x f(x)\, dx - \frac{1}{2^n} \cdot 1 \qquad (n = 1, 2, \dots).
\end{aligned}
$$

Entsprechend wird nun der Negativteil bearbeitet, d.h. es wird $\int X^- \, dP = \int_{-\infty}^{0} x f(x)\, dx$ gezeigt. Es ist

$$
\begin{aligned}
\int X_n^- \, dP \quad & = \quad \sum_{k=1}^{n2^n} \frac{k-1}{2^n} P_F \left(\left(\frac{k-1}{2^n}, \frac{k}{2^n} \right] \right) + n \cdot P_F((-\infty, -n]) \\
& = \quad \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \int_{-k/2^n}^{-(k-1)/2^n} f(x)\, dx + n \int_{-\infty}^{-n} f(x)\, dx \\
& = \quad \sum_{k=1}^{n2^n} \int_{-k/2^n}^{-(k-1)/2^n} \frac{k-1}{2^n} f(x)\, dx + \int_{-\infty}^{-n} n f(x)\, dx \\
& \leq \quad \sum_{k=1}^{n2^n} \int_{-k/2^n}^{-(k-1)/2^n} |x|\, f(x)\, dx + \int_{-\infty}^{-n} |x|\, f(x)\, dx \\
& = \quad \int_{-\infty}^{0} |x|\, f(x)\, dx = - \int_{-\infty}^{0} x f(x)\, dx \qquad (n = 1, 2, \dots)
\end{aligned}
$$

und andererseits

$$\begin{aligned}
\int X_n^- \, dP &= \sum_{k=1}^{n2^n} \int\limits_{-k/2^n}^{-(k-1)/2^n} \frac{k-1}{2^n} f(x)\, dx - \int_{-\infty}^{-n} n f(x)\, dx \\
&= \sum_{k=1}^{n2^n} \int\limits_{-k/2^n}^{-(k-1)/2^n} |x|\, f(x)\, dx - \sum_{k=1}^{n2^n} \int\limits_{-k/2^n}^{-(k-1)/2^n} \left(|x| - \frac{k-1}{2^n}\right) f(x)\, dx \\
&\quad \underbrace{+\, n \int\limits_{-\infty}^{-n} f(x)\, dx}_{\geq 0} \\
&\geq \sum_{k=1}^{n2^n} \int\limits_{-k/2^n}^{-(k-1)/2^n} |x|\, f(x)\, dx - \frac{1}{2^n} \sum_{k=1}^{n2^n} \int\limits_{-k/2^n}^{-(k-1)/2^n} f(x)\, dx \\
&\geq \int\limits_{-n}^{0} |x|\, f(x)\, dx - \frac{1}{2^n} \cdot 1 = - \int\limits_{-n}^{0} x f(x)\, dx - \frac{1}{2^n} \qquad (n = 1, 2, \ldots),
\end{aligned}$$

also folgt

$$\int X^- \, dP = \lim_{n\to\infty} \int X_n^- \, dP = - \int\limits_{-\infty}^{0} x f(x)\, dx.$$

■

9.25 Satz (Transformationssatz für Erwartungswerte):
Es sei X eine reelle Zufallsvariable über $(\Omega, \mathfrak{F}, P)$ mit Werten in $(\mathfrak{X}, \mathfrak{B}) = (\mathbb{R}, \mathfrak{B})$ und $g\colon (\mathfrak{X}, \mathfrak{B}) \to (\mathbb{R}, \mathfrak{B})$ eine messbare Abbildung, derart dass $g \geq 0$ oder $g(X)$ als Funktion von ω P–integrierbar ist. Dann gilt:

$$E[g(X)] = \int_\Omega g(X(\omega)) dP(\omega) = \int_\mathbb{R} g(x) dP_X(x) \quad \textit{kurz} \quad \mathrm{E}[g(X)] = \int g(X)\, dP = \int g\, dP_X,$$

wobei P_X das Bildmaß von P bzgl. X sei.

Beweis:
Der Beweis folgt dem Prinzip der „Algebraischen Induktion”, indem die Aussage sukzessive für Indikatorfunktionen (1), Elementarfunktionen (2), nichtnegative messbare Funktionen (3) und schließlich für allgemeine messbare Funktionen (4) bewiesen wird, so wie wir auch bei der Einführung des μ-Integrals vorgegangen sind.

(1): Für die Aussage von Satz 9.25 heißt dies konkret:
Betrachte Abbildungen g der Form $g = I_A$ mit $A \in \mathfrak{B}$. Offensichtlich gilt:

$$g(X(\omega)) = I_A(X(\omega)) = \begin{cases} 1 \ , \ X(\omega) \in A \\ 0 \ , \ X(\omega) \notin A \end{cases} = \begin{cases} 1 \ , \ \omega \in X^{-1}(A) \\ 0 \ , \ \omega \notin X^{-1}(A) \end{cases} = I_{X^{-1}(A)}(\omega).$$

Mit Hilfe von Satz 9.7 folgt deshalb:

$$\mathbf{E}_P[g(X)] = \mathbf{E}_P[I_{X^{-1}(A)}] = P(X^{-1}(A)) = P_X(A) = \mathbf{E}_{P_X}[I_A] = \mathbf{E}_{P_X}[g].$$

(2): Es sei nun g eine Elementarfunktion mit den endlich vielen Werten $\alpha_1, \ldots, \alpha_n$, $n \in \mathbb{N}$. Aufgrund von Satz 9.2 besitzt g eine Normaldarstellung der Form

$$g \;=\; \sum_{i=1}^{n} \alpha_i I_{A_i},$$

mit $A_i \in \mathfrak{B}$ für $i = 1, \ldots, n$. Deshalb lässt sich schreiben

$$\begin{aligned}\mathbf{E}_P[g(X)] \;&=\; \mathbf{E}_P\left[\sum_{i=1}^{n} \alpha_i I_{X^{-1}(A_i)}\right] \;=\; \sum_{i=1}^{n} \alpha_i \mathbf{E}_P[I_{X^{-1}(A_i)}] \;=\; \sum_{i=1}^{n} \alpha_i \mathbf{E}_{P_X}[I_{A_i}] \\ &=\; \mathbf{E}_{P_X}\left[\sum_{i=1}^{n} \alpha_i I_{A_i}\right] \;=\; \mathbf{E}_{P_X}[g].\end{aligned}$$

(3): Für $g \geq 0$ existiert aufgrund von Satz 9.8 eine monoton nichtfallende Folge g_n von nichtnegativen Elementarfunktionen mit $g = \sup\limits_n g_n$.

$$\begin{aligned}\mathbf{E}_P[g(X)] \;&=\; \mathbf{E}_P[\sup_n g_n(X)] \;\overset{\text{Satz 9.10}}{=}\; \sup_n \mathbf{E}_P[g_n(X)] \;=\; \sup_n \mathbf{E}_{P_X}[g_n] \\ &=\; \mathbf{E}_{P_X}[\sup_n g_n] \;=\; \mathbf{E}_{P_X}[g].\end{aligned}$$

(4): Allgemeine integrierbare messbare Funktionen behandelt man, indem man wieder ihren Positiv– und Negativteil betrachtet:

$$g = g^+ - g^- \;\Rightarrow\; (g \circ X)^+ = g^+ \circ X \;\text{ und }\; (g \circ X)^- = g^- \circ X.$$

Folglich gilt:

$$\begin{aligned}\mathbf{E}_P[g(X)] &= \int_\Omega (g(X))^+ \, dP - \int_\Omega (g(X))^- \, dP = \int_\Omega g^+(X) \, dP - \int_\Omega g^-(X) \, dP \\ &= \int_\mathbb{R} g^+ \, dP_X - \int_\mathbb{R} g^- \, dP_X = \int_\mathbb{R} g \, dP_X \,.\end{aligned}$$

■

9.26 Bemerkung:
Indem man für g die identische Abbildung $g\colon x \to x$ betrachtet, lässt sich allgemein schreiben:

$$\mathbf{E}_P[X] = \int\limits_\Omega X(\omega) \, dP(\omega) = \int\limits_\mathbb{R} x \, dP_X(x) =: \int\limits_\mathbb{R} x \, dF_X(x) \,.$$

9.27 Satz (Erwartungswert und Riemann–Integral):
Es sei X eine stetige Zufallsvariable über $(\Omega, \mathfrak{F}, P)$ mit der Dichte f und es sei $g\colon (\mathbb{R}, \mathfrak{B}) \to (\mathbb{R}, \mathfrak{B})$ eine messbare Abbildung mit der Eigenschaft, dass $|g| \cdot f$ Lebesgue–integrierbar ist. Dann ist $g(X)$ P–integrierbar mit

$$\mathrm{E}[g(X)] \;=\; \int\limits_{-\infty}^{+\infty} g(x) f(x)\ dx.$$

(Siehe auch Lebensdaten von Lebesgue und Riemann im Anhang D.)

Beweis:
Übung. ■

Aufgrund der Sätze 9.22, 9.24 und 9.25 stehen unterschiedliche Methoden zur Berechnung des Erwartungswertes einer Zufallsgröße zur Auswahl. Diese Möglichkeiten sollen an einem Beispiel veranschaulicht werden.

9.28 Beispiel:
Es sei $X \stackrel{d}{=} \mathcal{R}(0,2)$ und $g(x) := x^2$. Gesucht ist $\mathbf{E}_P[X^2] = \mathbf{E}_P[g(X)]$. Die Rechteckverteilung besitzt die Dichte

$$f(x) \;=\; \begin{cases} 1/2 & , \ x \in (0,2) \\ 0 & , \ x \notin (0,2) \end{cases}$$

und die Verteilungsfunktion

$$F(x) \;=\; \begin{cases} 0 & , \ x \le 0 \\ x/2 & , \ 0 < x < 2 \\ 1 & , \ x \ge 2 \end{cases} .$$

1. Möglichkeit: Aufgrund von Satz 9.27 gilt:

$$\mathbf{E}_P\left[X^2\right] = \int\limits_{-\infty}^{+\infty} x^2 f(x)\ dx \;=\; \int\limits_{0}^{2} x^2 \cdot \frac{1}{2}\ dx = \frac{1}{2} \cdot \frac{1}{3} x^3 \Big|_0^2 = \frac{1}{2} \cdot \frac{8}{3} = \frac{4}{3}.$$

2. Möglichkeit: Wir setzen $Y = X^2$. Es gilt

$$\widetilde{F}(y) = P(Y \le y) = P(X^2 \le y) = P(X \le \sqrt{y}) = \frac{1}{2}\sqrt{y}, \qquad 0 \le y \le 4.$$

Die Anwendung von Satz 9.22 liefert dann

$$\begin{aligned} \mathbf{E}_P\left[X^2\right] \;&=\; \int\limits_0^\infty (1 - \widetilde{F}(y))\ dy \;=\; \int\limits_0^4 \left(1 - \frac{1}{2}\sqrt{y}\right)\ dy \;=\; 4 - \frac{1}{2} \cdot \frac{y^{3/2}}{3/2}\Big|_0^4 \\ &=\; 4 - \frac{1}{2} \cdot \frac{64^{1/2}}{3/2} = 4 - \frac{1}{3} \cdot 8 = \frac{4}{3}. \end{aligned}$$

3. Möglichkeit: Die Dichte von Y lautet:

$$\widetilde{f}(y) = \begin{cases} \frac{1}{4\sqrt{y}} & , \ 0 < y < 4 \\ 0 & , \ \text{sonst} \end{cases} .$$

Damit wird

$$\mathbf{E}_P[X^2] = \int_0^\infty y\widetilde{f}(y)\, dy = \int_0^4 y\frac{1}{4\sqrt{y}}\, dy = \int_0^4 \frac{1}{4}\sqrt{y}\, dy = \frac{1}{4}\cdot\frac{2}{3}\sqrt{y^3}\Big|_0^4 = \frac{4}{3}.$$

9.29 Definition (Varianz):
Es sei $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \longrightarrow \mathbb{R}$ eine Zufallsgröße, für die sowohl $\mathbf{E}_P\,[X]$ als auch $\mathbf{E}_P\left[(X - \mathrm{E}[X])^2\right]$ existiert. Dann heißt

$$\mathrm{Var}[X] \ := \ \mathbf{E}_P\left[(X - \mathrm{E}[X])^2\right]$$

die Varianz von X.

Wie im diskreten Fall beweist man

9.30 Satz:
Es sei $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \longrightarrow \mathbb{R}$ eine Zufallsgröße, für die $\mathbf{E}_P\left[X^2\right]$ existiert. Dann existiert auch $\mathbf{E}_P[X]$ und es gilt:

a) $\mathrm{Var}[X] = \mathbf{E}_P\left[X^2\right] - (\mathbf{E}_P[X])^2$ *(Verschiebungssatz).*

b) $\mathrm{Var}[aX + b] = a^2 \cdot \mathrm{Var}[X]$.

Beweis:
Die Existenz von $\mathbf{E}_P\,[X]$ folgt aus der Abschätzung $|X| \leq \max(1, X^2)$ und Satz 9.18c. Die weiteren Schritte sind identisch mit dem Beweis von Satz 2.25. ■

9.31 Beispiel:
Es sei $X \stackrel{d}{=} \mathrm{Exp}(\lambda)$, dann gilt:

$$\begin{aligned} \mathbf{E}_P\,[X] &= \int_0^\infty x f(x)\, dx = \int_0^\infty x\lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}. \\ \mathbf{E}_P\left[X^2\right] &= \int_0^\infty P(X^2 > x)\, dx = \int_0^\infty P(X > \sqrt{x})\, dx = \int_0^\infty (1 - F(\sqrt{x}))\, dx \\ &= \int_0^\infty e^{-\lambda\sqrt{x}}\, dx = 2\int_0^\infty e^{-\lambda t} t\, dt = \frac{2}{\lambda^2}. \end{aligned}$$

Damit ergibt sich:

$$\mathrm{Var}[X] = \mathbf{E}\left[X^2\right] - (\mathbf{E}[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

9.32 Bemerkung:
Neben dem Erwartungswert und der Varianz einer Zufallsvariablen werden manchmal noch weitere Kennzahlen zur Charakterisierung einer Verteilungsfunktion herangezogen:
Die Größe

$$\sigma := \sqrt{\mathrm{Var}[X]},$$

heißt Standardabweichung oder Streuung von X. Im Falle $\mathbf{E}_P[X] \neq 0$ nennt man den Quotienten

$$\frac{\sqrt{\mathrm{Var}[X]}}{\mathbf{E}_P[X]}$$

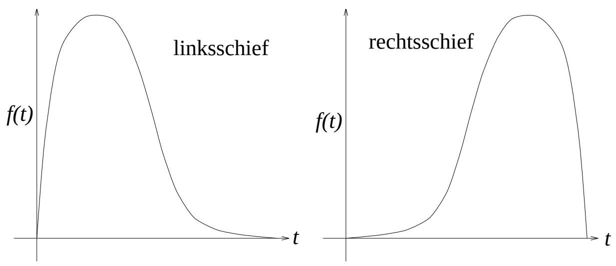
den Variationskoeffizienten von X. Beide Kennzahlen charakterisieren — ebenso wie die Varianz — die mittlere Abweichung der Zufallsgröße von ihrem Erwartungswert.
Andere Eigenschaften des Verteilungsgesetzes kann man mit Hilfe der sogenannten höheren Momente beschreiben:

$$\mathbf{E}_P\left[X^k\right], \ \mathbf{E}_P\left[|X|^k\right], \ \mathbf{E}_P\left[(X - \mathrm{E}[X])^k\right], \ \mathbf{E}_P\left[|X - \mathrm{E}[X]|^k\right] \ (k \in \mathbb{N}).$$

Sie heißen der obigen Reihenfolge nach das k–te Moment, das k–te absolute Moment, das k–te zentrale Moment und das k–te zentrale absolute Moment von X. Das zweite zentrale Moment entspricht der Varianz. Den Quotienten

$$\frac{\mathbf{E}_P\left[(X - \mathrm{E}[X])^3\right]}{\mathrm{Var}[X]^{\frac{3}{2}}}$$

bezeichnet man als die Schiefe von X.



Diese Kennzahl charakterisiert Abweichungen von der Symmetrie des Verteilungsgesetzes von X. Ist X symmetrisch verteilt zu einem Punkt $x \in \mathbb{R}$, wie zum Beispiel jede normalverteilte Zufallsvariable, so ist die Schiefe gleich Null, während Zufallsvariablen X mit Dichtefunktionen der Form wie in der obenstehenden Abbildung eine positive bzw. eine negative Schiefe besitzen.

9.33 Bemerkung:
Sofern keine Verwechslungen möglich sind, wird auch anstelle von $\mathbf{E}_P[X]$ kurz $\mathrm{E}[X]$ geschrieben.

9.6 Tabelle mit Kenngrößen verschiedener Verteilungen

Zum Abschluss dieses Kapitels soll noch eine tabellarische Übrsicht über die Kenngrößen der bisher behandelten kontinuerilichen Verteilungen gegeben werden.

Rechteckverteilung ($R(a,b)$, $a<b$)	
Dichte	$f(x) := \begin{cases} \frac{1}{b-a} & , \quad x \in [a,b], \\ 0 & , \quad x \notin [a,b] \end{cases}$
Verteilung	$F(x) := \begin{cases} 0 & , \quad x < a, \\ \frac{x-a}{b-a} & , \quad a \le x \le b, \\ 1 & , \quad x > b. \end{cases}$
Erwartungswert	$\mathbf{E}[X] = \frac{a+b}{2}$
Varianz	$\mathrm{Var}[X] = \frac{(b-a)^2}{12}$

Exponentialverteilung ($\mathrm{Exp}(\lambda)$, $\lambda > 0$)	
Dichte	$f(x) := \begin{cases} \lambda e^{-\lambda x} & , \quad x \ge 0, \\ 0 & , \quad x < 0. \end{cases}$
Verteilung	$F(x) := \begin{cases} 1 - e^{-\lambda x} & , \quad x \ge 0, \\ 0 & , \quad x < 0 \end{cases}$
Erwartungswert	$\mathbf{E}[X] = \frac{1}{\lambda}$
Varianz	$\mathrm{Var}[X] = \frac{1}{\lambda^2}$

Weibull–Verteilung ($\lambda, \beta > 0$)	
Dichte	$f(x) := \begin{cases} \lambda \cdot \beta \cdot x^{\beta-1} \cdot e^{-\lambda x^\beta} & , \quad x \ge 0 \\ 0 & , \quad x < 0 \end{cases}$
Verteilung	$F(x) := \begin{cases} 1 - e^{-\lambda x^\beta} & , \quad x \ge 0 \\ 0 & , \quad x < 0 \end{cases}$
Erwartungswert	$\mathbf{E}[X] = \frac{\lambda^{-1/\beta}}{\beta} \Gamma\left(\frac{1}{\beta}\right)$
Varianz	$\mathrm{Var}[X] = \frac{\lambda^{-2/\beta}}{\beta} \left(2\Gamma\left(\frac{2}{\beta}\right) - \frac{1}{\beta}\left(\Gamma\left(\frac{1}{\beta}\right)\right)^2 \right)$

Standardnormalverteilung ($\mathcal{N}(0,1)$)	
Dichte	$f(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
Verteilung	$F(x) := \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \, dt$
Erwartungswert	$\mathbf{E}[X] = 0$
Varianz	$\mathrm{Var}[X] = 1$

Normalverteilung ($\mathcal{N}(\mu,\sigma)$, $\mu \in \mathbb{R}$, $\sigma > 0$)	
Dichte	$f(x) := \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
Verteilung	$F(x) := \int\limits_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = \Phi\left(\frac{x-\mu}{\sigma}\right)$
Erwartungswert	$\mathbf{E}[X] = \mu$
Varianz	$\mathrm{Var}[X] = \sigma^2$

Logarithmische Normalverteilung ($\mu \in \mathbb{R}$, $\sigma > 0$)	
Dichte	$f(x) := \begin{cases} \frac{1}{\sigma x \sqrt{2\pi}} e^{-(\log(x)-\mu)^2/2\sigma^2} & , \quad x > 0 \\ 0 & , \quad x \le 0 \end{cases}$
Verteilung	$F(x) := \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)$
Erwartungswert	$\mathbf{E}[X] = e^{\mu + \sigma^2/2}$
Varianz	$\mathrm{Var}[X] = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2}$

Cauchy–Verteilung ($\lambda > 0$, $\mu \in \mathbb{R}$)	
Dichte	$f(x) := \frac{1}{\pi} \cdot \frac{\lambda}{\lambda^2+(x-\mu)^2}$
Verteilung	$F(x) := \int_{-\infty}^{x} f(t)\ dt$
Erwartungswert	$\mathbf{E}[X] = \mu$
Varianz	existiert nicht
Gammaverteilung ($Gamma(b,p)$, $b, p \in \mathbb{R}^+$)	
Dichte	$f(x) := \begin{cases} \frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx} & , \quad x \geq 0 \\ 0 & , \quad x < 0 \end{cases}$
Verteilung	$F(x) := \int_{-\infty}^{x} f(t)\ dt$
Erwartungswert	$\mathbf{E}[X] = \frac{p}{b}$
Varianz	$\mathrm{Var}[X] = \frac{p}{b^2}$
Die χ^2–Verteilung ($Gamma(\frac{1}{2}, \frac{n}{2})$, $n \in \mathbb{N}$)	
Dichte	$f(x) := \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & , \quad x \geq 0 \\ 0 & , \quad x < 0 \end{cases}$
Verteilung	$F(x) := \int_{-\infty}^{x} f(t)\ dt$
Erwartungswert	$\mathbf{E}[X] = n$
Varianz	$\mathrm{Var}[X] = 2n$
Erlang–Verteilung ($Gamma(b,n)$, $n \in \mathbb{N}$, $b \in \mathbb{R}^+$)	
Dichte	$f(x) := \begin{cases} \frac{b^n}{(n-1)!} x^{n-1} e^{-bx} & , \quad x \geq 0, \\ 0 & , \quad x < 0 \end{cases}$
Verteilung	$F(x) := \begin{cases} 1 - e^{-bx} \cdot \sum\limits_{k=0}^{n-1} \dfrac{(bx)^k}{k!} & , \quad x \geq 0, \\ 0 & , \quad x < 0. \end{cases}$
Erwartungswert	$\mathbf{E}[X] = \frac{n}{b}$
Varianz	$\mathrm{Var}[X] = \frac{n}{b^2}$
Betaverteilung ($\mathrm{Beta}(p,q)$)	
Dichte	$f(x) := \begin{cases} \frac{(b-a)^{1-p-q}}{B(p,q)} (x-a)^{p-1}(b-x)^{q-1} & , \quad x \in (a,b) \\ 0 & , \quad x \notin (a,b) \end{cases}$
Verteilung	$F(x) := \int_{-\infty}^{x} f(t)\ dt$
Erwartungswert	$\mathbf{E}[X] = \frac{bp+aq}{p+q}$
Varianz	$\mathrm{Var}[X] = \frac{(a-b)^2 pq}{(p+q)^2(1+p+q)}$

(Siehe auch Mathematica-Notebook zu den Verteilungen.)

Eine umfangreiche Sammlung von Informationen zu den Verteilungen aus diesem Abschnitt und vielen weiteren Verteilungen findet sich unter `www.xycoon.com`.

9.7 Weitere Hilfsmittel aus der Maß– und Integrationstheorie

In diesem Abschnitt geht es zunächst um fast überall bestehende Eigenschaften.

9.34 Definition (μ–Nullmenge):
Es sei $(\Omega, \mathfrak{F}, \mu)$ ein Maßraum. Eine Teilmenge $N \subseteq \Omega$ heißt $\underline{\mu\text{–Nullmenge}}$, wenn $N \in \mathfrak{F}$ und $\mu(N) = 0$ ist.

9.35 Definition (μ–fast–überall):
Man sagt, die Aussage P über Elemente von Ω ist <u>*μ–fast–überall*</u> *(auf Ω) wahr, wenn es eine μ–Nullmenge $N \subseteq \Omega$ gibt, so dass P für alle $\omega \in \overline{N} = \Omega \backslash N$ wahr ist.*

9.36 Satz:
Es sei $(\Omega, \mathfrak{F}, \mu)$ ein Maßraum und $f\colon \Omega \to \mathbb{R}^+$ sei integrierbar. Es ist genau dann $\int_\Omega f\, d\mu = 0$, wenn $f = 0$ f.ü. .

Beweis:

1. Es sei $\int f d\mu = 0$: Wegen der Messbarkeit von f ist $T(f) = \{\omega \mid f(\omega) > 0\} \in \mathfrak{F}$ und für $n = 1, 2, 3, \ldots$ sei $A_n := \left\{\omega \mid f(\omega) \geq \frac{1}{n}\right\} \in \mathfrak{F}$. Es gilt $A_n \nearrow T(f)$. Nun gilt $f \geq \frac{1}{n} \cdot I_{A_n}$ und nach Voraussetzung $0 = \int f d\mu \geq \int \frac{1}{n} I_{A_n} d\mu = \frac{1}{n}\mu(A_n)$ für alle n. Damit ist auch $\mu\left(\bigcup\limits_{n=1}^{\infty} A_n\right) = \mu(T(f)) = 0$, d.h. es ist $f = 0$ f.ü..
2. Es sei jetzt $f = 0$ f.ü.: Für die Folge von Elementarfunktionen $f_n = nI_{T(f)}(\omega)$ gilt $\int f_n d\mu = 0$ für alle $n \in \mathbb{N}$. Der Grenzwert $\hat{f} := \lim\limits_{n\to\infty} f_n$ ist eine nichtnegative messbare Funktion, für die $f \leq \hat{f}$ gilt. Dann ist nach der Definition und wegen der Monotonie des Integrals $\int f d\mu \leq \int \hat{f} d\mu = 0$.

■

9.37 Satz:
Es sei $(\Omega, \mathfrak{F}, \mu)$ ein Maßraum und $f, g\colon \Omega \to \mathbb{R}$ seien messbar. Ist f μ–integrierbar und $f = g$ f.ü., dann ist auch g μ–integrierbar und es gilt

$$\int\limits_\Omega f\, d\mu = \int\limits_\Omega g\, d\mu.$$

Beweis:
Wir setzen $h := f - g$, dann ist h messbar und $h = 0$ f.ü. Damit ist auch $h^+ = h^- = 0$ f.ü. Somit ist dann $\int h^+ d\mu = \int h^- d\mu = 0$, also $\int f^+ d\mu = \int g^+ d\mu$ und $\int f^- d\mu = \int g^- d\mu$. Daher ist g integrabel und $\int f d\mu = \int g d\mu$. ■

9.38 Definition (Integral):
Es sei $(\Omega, \mathfrak{F}, \mu)$ ein Maßraum und $f\colon \Omega \to \mathbb{R}$ quasiintegrabel. Für $A \in \mathfrak{F}$ nennt man

$$\int\limits_A f\, d\mu := \int\limits_\Omega f \cdot I_A\, d\mu$$

das <u>*Integral*</u> *von f über A.*

Im Folgenden sollen einige Eigenschaften des Integrals genauer betrachtet werden.

9.39 Satz:
Es sei $(\Omega, \mathfrak{F}, \mu)$ ein Maßraum und $f\colon \Omega \to \mathbb{R}$ integrabel. Ist $f > 0$ f.ü. auf $A \in \mathfrak{F}$, dann gilt:

$$\int\limits_A f d\mu = 0 \;\Leftrightarrow\; \mu(A) = 0.$$

Beweis:
Es ist f.ü. $I_A \cdot f \geq 0$ und es sei $\int_\Omega I_A f \, d\mu = 0$ und aufgrund von Satz 9.36 ist daher $I_A f = 0$ f.ü.. Folglich ist $I_A = 0$ f.ü. und mithin ist A eine μ–Nullmenge. Umgekehrt folgt aus $\mu(A) = 0$, dass $f \cdot I_A = 0$ f.ü.. Für diesen Fall besagt Satz 9.36, dass $\int_\Omega f I_A \, d\mu = \int_A f \, d\mu = 0$. ■

Selbstverständlich gilt 9.39 auch für $f < 0$ f.ü.. Der Beweis verläuft analog.

9.40 Satz:
Es sei $(\Omega, \mathfrak{F}, \mu)$ ein Maßraum und $f \colon \Omega \to \mathbb{R}$ eine integrierbare Funktion mit $\int_A f \, d\mu = 0$ für alle $A \in \mathfrak{F}$, dann ist $f = 0$ f.ü..

Beweis:
Es seien $P := \{\omega \mid f(\omega) > 0\}$ und $N := \{\omega \mid f(\omega) < 0\}$. Wegen $\int f \, d\mu = 0$ und $f(\omega) > 0$ f.ü. auf P folgt aus 9.39, dass $\mu(P) = 0$ gilt. Analog schließt man auf $\mu(N) = 0$. Daher ist der Träger von f : $\{\omega \mid f(\omega) \neq 0\} = P \cup N$ eine Nullmenge, also $f = 0$ f.ü., was zu zeigen war. ■

9.41 Satz:
Es sei $(\Omega, \mathfrak{F}, \mu)$ ein Maßraum und $f, g \colon \Omega \to \mathbb{R}$ seien integrierbare Funktionen. Gilt für alle $A \in \mathfrak{F}$

$$\int\limits_A f d\mu = \int\limits_A g d\mu,$$

so ist $f = g$ f.ü.

Beweis:
Nach Satz 9.40 folgt aus $\int\limits_A (f-g) d\mu = 0 \;\forall\, A \in \mathfrak{F} : f - g = 0$ f.ü., womit alles gezeigt ist. ■

9.42 Satz (Satz von der monotonen Konvergenz):
Es sei $(\Omega, \mathfrak{F}, \mu)$ ein Maßraum und $(f_n)_{n \in \mathbb{N}}$ eine Folge messbarer Funktionen $f_n \colon \Omega \to \overline{\mathbb{R}^+}$ mit $f_n \leq f_{n+1}$ für alle $n \in \mathbb{N}$. Dann gilt:

(i) $\displaystyle\int_\Omega \lim_{n \to \infty} f_n d\mu = \int_\Omega \sup_{n \in \mathbb{N}} f_n d\mu = \sup_{n \in \mathbb{N}} \int_\Omega f_n d\mu = \lim_{n \to \infty} \int_\Omega f_n d\mu$

(ii) $\displaystyle f := \sup_{n \in \mathbb{N}} f_n$ *ist integrierbar oder* $\displaystyle\sup_{n \in \mathbb{N}} \int_\Omega f_n \, d\mu < \infty$.

Beweis:
Wegen der Monotonie von $(f_n)_{n \in \mathbb{N}}$ und $(\int f_n)_{n \in \mathbb{N}}$ sind die Suprema gleich den Limiten. Es sei $f := \sup_{n \in \mathbb{N}} f_n$. Dann zeigt $f_n \leq f$ sofort $\int f_n \, d\mu \leq \int f \, d\mu$ und somit

$$\sup_{n \in \mathbb{N}} \int f_n \, d\mu \leq \int f \, d\mu.$$

Ist umgekehrt $(h_{n,m})_{m \in \mathbb{N}}$ eine monoton nichtfallende Folge von Elementarfunktionen mit $h_{n,m} \to f_n$, so gilt aufgrund von Satz 9.10

$$\int h_{n,m} \, d\mu \leq \sup_{n \in \mathbb{N}} \int f_n \, d\mu \,.$$

Alles weitere folgt mit Satz 9.11. ■

Aus Satz 9.42 ergibt sich

9.43 Satz:
Es sei $(\Omega, \mathfrak{F}, \mu)$ *ein Wahrscheinlichkeitsraum und* $(g_n)_{n\in\mathbb{N}}$ *eine Folge messbarer Funktionen* $g_n\colon \Omega \to \overline{\mathbb{R}^+}$. *Dann gilt:*

$$\int_\Omega \sum_{n=1}^{\infty} g_n \, d\mu = \sum_{n=1}^{\infty} \int_\Omega g_n \, d\mu \,.$$

Beweis:
Man wende Satz 9.42 an mit $f_n := \sum_{k=1}^n g_k$ für alle $n \in \mathbb{N}$ und $f := \sup_k g_k$. ■

9.44 Satz (Lemma von Fatou):
Es sei $(\Omega, \mathfrak{F}, \mu)$ *ein Maßraum und* $(f_n)_{n\in\mathbb{N}}$ *eine Folge messbarer Funktionen* $f_n\colon \Omega \to \overline{\mathbb{R}^+}$. *Dann gilt*

$$\int_\Omega \liminf_{n\to\infty} f_n \, d\mu \le \liminf_{n\to\infty} \int_\Omega f_n \, d\mu \,.$$

Beweis:
Für jedes $\omega \in \Omega$ sei $g_n(\omega) := \inf\limits_{k\ge n} f_k(\omega)$. Dann ist $g_n \ge 0$, messbar und die Folge $(g_n)_{n\in\mathbb{N}}$ ist monoton nichtfallend mit

$$\lim_{n\to\infty} g_n(\omega) = \sup_{n\in\mathbb{N}} g_n(\omega) = \liminf_{n\to\infty} f_n(\omega) \,.$$

Da für alle n gilt: $f_n \ge g_n$, lässt sich mit Hilfe von Satz 9.42 folgern, dass

$$\liminf_{n\to\infty} \int f_n \, d\mu \ge \lim_{n\to\infty} \int g_n \, d\mu = \int \lim_{n\to\infty} g_n \, d\mu = \int \liminf_{n\to\infty} f_n \, d\mu \,.$$

■

9.45 Satz (Satz von der majorisierten Konvergenz):
Es sei $(\Omega, \mathfrak{F}, \mu)$ *ein Maßraum und* $(f_n)_{n\in\mathbb{N}}$ *eine Folge integrierbarer Funktionen* $f_n\colon \Omega \to \mathbb{R}$, *die f.ü. gegen eine messbare Funktion* $f\colon \Omega \to \mathbb{R}$ *konvergiert. Existiert eine integrierbare Funktion* $g\colon \Omega \to \mathbb{R}$ *mit* $|f_n| \le g$ *f.ü. für alle* $n \in \mathbb{N}$, *dann ist auch* f *integrierbar und es gilt:*

$$\int f \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu \,.$$

Beweis:
O.B.d.A. kann angenommen werden, dass $f_n \ge 0$, $\forall\, n \in \mathbb{N}$ und $f \ge 0$. Es sei jetzt $N \in \mathfrak{F}$ eine Nullmenge mit

$$\begin{aligned} \lim_{n\to\infty} f_n(\omega) &= f(\omega) \qquad \forall\, \omega \in \Omega\backslash N \,, \\ |f_n(\omega)| &\le g(\omega) \qquad \forall\, \omega \in \Omega\backslash N \,. \end{aligned}$$

Ersetzt man nun f_n durch $f_n \cdot (1 - I_N)$ und f durch $f \cdot (1 - I_N)$, lässt sich deshalb $N = \varnothing$ annehmen. Mit dem Lemma von Fatou (Satz 9.44) folgt dann, dass

$$\int f \, d\mu \le \liminf_{n\to\infty} \int f_n \, d\mu \le \int g \, d\mu$$

und

$$\int g\,d\mu - \int f\,d\mu = \int (g-f)\,d\mu \le \liminf_{n\to\infty} \int (g-f_n)\,d\mu = \int g\,d\mu - \limsup_{n\to\infty} \int f_n\,d\mu\,.$$

Folglich gilt

$$\limsup_{n\to\infty} \int f_n\,d\mu \le \int f\,d\mu \le \liminf_{n\to\infty} \int f_n\,d\mu \le \limsup_{n\to\infty} \int f_n\,d\mu\,,$$

was zu zeigen war. ■

Literatur zu Kapitel 9

Folgende Bücher werden als begleitende Literatur zu diesem Kapitel des Skriptes empfohlen:

- H. Bauer:
  Wahrscheinlichkeitstheorie,
  5. Auflage, Walter de Gruyter, Berlin, 2002.
  ISBN: 3110172364

- W. Behnen, G. Neuhaus:
  Grundkurs Stochastik,
  3. Auflage, Teubner-Verlag, Stuttgart, 1995.
  ISBN: 3930737698

- P. Billingsley:
  Probability and Measure,
  2nd edition, John Wiley and Sons, New York, 1986.
  ISBN: 0471007102

- P. Gänssler und W. Stute:
  Wahrscheinlichkeitstheorie,
  Springer–Verlag, Berlin, 1977.
  ISBN: 3540084185

- E. Henze:
  Einführung in die Maßtheorie,
  Bibl. Institut, Mannheim, 1971.
  ISBN: 341100505X

- R. Mathar/D. Pfeiffer:
  Stochastik für Informatiker,
  Teubner, 1990.
  ISBN: 3519022400

- P. P. Spies:
  Grundlagen stochastischer Modelle,
  Hanser, München, 1982.
  ISBN: 3446137114

Kapitel 10

Zuverlässigkeit

Die Zuverlässigkeitstheorie stellt eine Anwendung der Wahrscheinlichkeitstheorie dar. Gegenstand der Zuverlässigkeitstheorie ist die Zuverlässigkeit technischer Systeme, die gemessen, analysiert und optimiert werden soll. Analyse und Optimierung werden anhand bestimmter Kenngrößen wie Lebensdauerverteilung und Ausfallrate vorgenommen.

Schlüsselwörter: Lebensdauerverteilung, Zuverlässigkeitsfunktion, Ausfallrate, Serienschaltung, Parallelschaltung, Brückenschaltung, Operationspfade, dynamische Optimierung, Zustand, Entscheidungsfolge, Bellmansches Optimalitätsprinzip, Bellmansche Funktionalgleichung, Aufteilungsproblem

10.1 Einführung und Grundbegriffe

In der Zuverlässigkeitstheorie werden folgende Aufgabenstellungen betrachtet:

- Modellierung des Ausfallverhaltens und der Abnutzung von Systemen mit Hilfe stochastischer Modelle. (Stichwörter: Lebensdauerverteilungen, Zuverlässigkeit)
- Untersuchung des Zusammenhangs zwischen Zuverlässigkeitskenngrößen eines Systems und seiner Subsysteme. (Stichwörter: Systemfunktionen, monotone Systeme)
- Untersuchung und Optimierung von Maßnahmen zur Erhaltung bzw. Wiederherstellung der Arbeitsfähigkeit von Systemen, einschließlich der Interdependenzen zwischen Mensch und Maschine. (Stichwörter: Instandsetzungs– bzw. Instandhaltungsstrategien, repair–men–problem)
- Schätzung und Prüfung der Zuverlässigkeitskenngrößen durch Anwendung mathematisch–statistischer Verfahren sowie Entwicklung spezifischer Schätz– und Prüfverfahren. (Stichwort: Lebensdaueranalyse)

In dieser Vorlesung soll lediglich ein Einblick in die Zuverlässigkeitstheorie gegeben werden. Daher beschränkt sich dieses Kapitel auf den ersten der oben genannten Bereiche.

10.1 Definition (Lebensdauer-, Überlebens- und Zuverlässigkeitsfunktion):
Als Lebensdauer einer Maschine oder Komponente wird die Zeit zwischen Inbetriebnahme und Ausfall bezeichnet. Sie wird durch eine nichtnegative Zufallsgröße T beschrieben.
Die Lebensdauerfunktion sei die Verteilungsfunktion der Lebensdauer T, d.h.

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

Mit Hilfe der Lebensdauerfunktion lässt sich die Überlebens- bzw. Zuverlässigkeitsfunktion $\overline{F}(t)$ der Komponente als

$$\overline{F}(t) := 1 - F(t) = P(T > t), \quad t \in \mathbb{R}^+,$$

definieren.

Im Weiteren sei F eine stetige Verteilungsfunktion und f die dazugehörige Dichte.

10.2 Bemerkung (mittlere Lebensdauer, Varianz der Lebensdauer):
Typische Kenngrößen der Lebensdauerfunktion sind die mittlere Lebensdauer $\mathrm{E}[T]$ und die Varianz der Lebensdauer $\mathrm{Var}[T]$.

10.3 Definition (Ausfallrate):
Es sei F eine stetige Verteilungsfunktion. Dann heißt

$$a(t) := \lim_{h \to 0+} \frac{1}{h} P(t < T \leq t + h \mid T > t)$$

die Ausfallrate eines Bauteils mit der Lebensdauer T.
$\Delta h \cdot a(t)$ ist die Wahrscheinlichkeit dafür, dass eine Komponente nach Erreichen des Lebensalters t innerhalb der Zeitspanne Δh ausfällt.

10.4 Satz:
Es sei $a(t)$ die Ausfallrate eines Bauteils mit stetiger Lebensdauerverteilung. Dann gilt für die Zuverlässigkeitsfunktion:

$$\overline{F}(t) = \exp\left(-\int\limits_0^t a(u)\ du\right), \quad t \in \mathbb{R}^+.$$

Beweis:
Es gilt:

$$P(t < T \leq t+h \mid T > t) = \frac{P(t < T \leq t+h)}{P(T > t)} = \frac{F(t+h) - F(t)}{1 - F(t)}.$$

Somit folgt:

$$\begin{aligned} a(t) &= \lim_{h \to 0} \frac{1}{h} \cdot \frac{F(t+h) - F(t)}{1 - F(t)} = \lim_{h \to 0} \frac{F(t+h) - F(t)}{h} \cdot \frac{1}{1 - F(t)} \\ &= \frac{d}{dt} F(t) \cdot \frac{1}{1 - F(t)} = \frac{f(t)}{1 - F(t)}, \quad t \in \mathbb{R}^+, \end{aligned}$$

bzw.

$$a(t) = \frac{f(t)}{1 - F(t)} = -\frac{d}{dt} \ln(1 - F(t)), \quad t \in \mathbb{R}^+.$$

Diese Differentialgleichung lässt sich durch Integration lösen:

$$\ln(1 - F(u))\Big|_0^t = -\int_0^t a(u)\ du + c,$$

was gleichbedeutend ist mit

$$1 - F(t) = c' \cdot \exp\left(-\int_0^t a(u)\ du\right), \quad t \in \mathbb{R}^+.$$

Die Bestimmung der Konstanten c' erfolgt mit Hilfe der Anfangsbedingung $F(0) := 0$, die $c' = 1$ und damit insgesamt

$$\overline{F}(t) = 1 - F(t) = \exp\left(-\int\limits_0^t a(u)\ du\right), \quad t \in \mathbb{R}^+,$$

ergibt. ■

Neben der Ausfallrate ist die bedingte Restlebensdauer, d.h. die Lebensdauerverteilung unter der Bedingung, dass das Bauteil bereits bis zum Zeitpunkt t_0 überlebt hat, von Interesse.

10.5 Definition (Bedingte Überlebenswahrscheinlichkeit, bedingte Restlebensdauer):
Es sei

$$\overline{F}_{t_0}(t) := P(T > t + t_0 | T > t_0) = \frac{\overline{F}(t + t_0)}{\overline{F}(t_0)} = \frac{1 - F(t + t_0)}{1 - F(t_0)}, \quad t \in \mathbb{R}^+,$$

die bedingte Überlebenswahrscheinlichkeit ab dem Zeitpunkt $t_0 \in \mathbb{R}^+$. Entsprechend wird die bedingte Restlebensdauer ab dem Zeitpunkt $t_0 \in \mathbb{R}^+$, d.h. die Verteilung der Restlebensdauer *unter der Annahme, dass das Bauteil bereits bis zum Zeitpunkt $t_0 \in \mathbb{R}^+$ überlebt hat, definiert als*

$$F_{t_0}(t) := 1 - \overline{F}_{t_0}(t) = P(T \leq t + t_0 | T > t_0) = \frac{F(t+t_0)}{F(t_0)}, \quad t \in \mathbb{R}^+.$$

Die bedingte Überlebenswahrscheinlichkeit $\overline{F}_{t_0}(t)$ lässt sich ebenfalls durch die Ausfallrate $a(t)$ ausdrücken:

$$\overline{F}_{t_0}(t) = exp\left(-\int_0^t a(u+t_0)du\right), \quad t \in \mathbb{R}^+.$$

Damit lassen sich folgende Eigenschaften formulieren:

- Die Ausfallrate $a(t)$ ist genau dann eine monoton wachsende Funktion, wenn für jedes feste t die bedingte Überlebenswahrscheinlichkeit $\overline{F}_{t_0}(t)$ eine fallende Funktion des erreichten Lebensalters ist.

- Die Ausfallrate $a(t)$ ist genau dann eine monoton fallende Funktion, wenn für jedes feste t die bedingte Überlebenswahrscheinlichkeit $\overline{F}_{t_0}(t)$ eine wachsende Funktion des erreichten Lebensalters ist.

Klassen von Lebensdauerverteilungen

Parametrische Klassen

Im günstigsten Fall kann die Lebensdauer mit einer bekannten Verteilung beschrieben werden. In Kapitel 7.3 wurden bereits die Exponentialverteilung, ihre Verallgemeinerung, die Weibullverteilung, sowie die logarithmische Normalverteilung als Kandidaten zur Modellierung von Lebensdauern genannt. Nachstehende Tabelle stellt die im vorigen Abschnitt definierten Begriffe für diese Verteilungen zusammen:

Es gilt jeweils für alle $t \in \mathbb{R}^+$:

Exponentialverteilung	**Weibull–Verteilung**	**Log-Normalverteilung**
	Verteilungsfunktion $F(t) =$	
$1 - e^{-\alpha t},\ \alpha > 0$	$1 - e^{-\alpha t^\beta},\ \alpha, \beta > 0$	$\Phi\left(\frac{\log(t)-\mu}{\sigma}\right)$
	Dichte $f(t) =$	
$\alpha e^{-\alpha t}$	$\alpha\beta t^{\beta-1} e^{-\alpha t^\beta}$	$\frac{1}{\sigma t\sqrt{2\pi}} e^{-(\log(t)-\mu)^2/2\sigma^2}$
	Zuverlässigkeitsfunktion $\overline{F}(t) = 1 - F(t) =$	
$e^{-\alpha t}$	$e^{-\alpha t^\beta}$	$1 - \Phi\left(\frac{\log(t)-\mu}{\sigma}\right)$
	Mittlere Lebensdauer $\mathrm{E}[T] =$	
$\frac{1}{\alpha}$	$\frac{\Gamma(1/\beta+1)}{\alpha^{1/\beta}}$	$e^{\mu+\sigma^2/2}$

Exponentialverteilung	**Weibull–Verteilung**	**Log-Normalverteilung**
Varianz der Lebensdauer $\mathrm{Var}[T]=$		
$\frac{1}{\alpha^2}$	$\frac{\Gamma(2/\beta+1)-(\Gamma(1/\beta+1))^2}{(\alpha^{1/\beta})^2}$	$e^{2\mu+2\sigma^2}-e^{2\mu+\sigma^2}$
Ausfallrate $a(t)=\frac{f(t)}{1-F(t)}=$		
$\frac{\alpha\cdot e^{-\alpha t}}{e^{-\alpha t}}=\alpha$	$\alpha\beta t^{\beta-1}$	$\frac{\frac{1}{\sigma t\sqrt{2\pi}}e^{-(\log(t)-\mu)^2/2\sigma^2}}{1-\Phi\left(\frac{\log(t)-\mu}{\sigma}\right)}$
Bedingte Überlebenswahrscheinlichkeit $\overline{F}_{t_0}(t)=$		
$\frac{1-(1-e^{-\alpha(t+t_0)})}{1-(1-e^{-\alpha t_0})}=e^{-\alpha t}$	$e^{-\alpha(t_0+t)^\beta+\alpha t_0^\beta}$	$\frac{1-\Phi\left(\frac{\log(t+t_0)-\mu}{\sigma}\right)}{1-\Phi\left(\frac{\log(t_0)-\mu}{\sigma}\right)}$

Besonders hervorzuheben ist hier die bedingte Restwahrscheinlichkeit der Exponentialverteilung, die, unabhängig vom Beobachtungszeitpunkt, wieder exponentiell verteilt ist.

Nichtparametrische Klassen

Die vorgestellten Lebensdauerverteilungen der parametrischen Klasse sind durch ihre Verteilungsfunktionen bzw. deren Parameter charakterisiert. Durch die Wahl dieser Parameter bedingt, können sowohl fallende, als auch steigende Ausfallraten auftreten. Unter Umständen ist es jedoch schwierig, einer gegebenen Ausfallrate eine entsprechende Verteilungsfunktion anzupassen. In diesen Fällen wird die Verteilungsfunktion oft anhand der Eigenschaften der Ausfallrate klassifiziert.

10.6 Definition (Increasing Failure Rate, Decreasing Failure Rate):
Eine Verteilungsfunktion $F(t)$ ist eine Increasing- (IFR) oder eine Decreasing Failure Rate (DFR) Verteilung, wenn die bedingte Überlebenswahrscheinlichkeit $\overline{F}_{t_0}(t)$ bei beliebigem, aber festem $t>0$ monoton fällt bzw. wächst in t_0.

10.7 Beispiel:
Die Weibull–Verteilung für $\beta>1$ und die Erlangverteilung sind IFR–Verteilungen. Für $\beta<1$ ist die Weibull–Verteilung eine DFR–Verteilung. Die Exponentialverteilung ist sowohl eine IFR-, als auch eine DFR-Verteilung. Die logarithmische Normalverteilung hingegen stellt keinen der beiden Typen dar.

10.8 Satz:
Eine Verteilungsfunktion $F(t)$ ist genau dann vom Typ IFR bzw. DFR, wenn die Funktion $\ln\overline{F}(t)$ konkav bzw. konvex ist.

Beweis:
Eine Funktion heißt **konvex**, wenn die Sekante durch zwei Punkte des Graphen stets oberhalb des Graphen liegt. Liegt die Sekante stets unterhalb des graphen, so heißt die Funktion **konkav**. Es gelten folgende Kriterien:

- Eine Funktion f ist konvex (konkav) genau dann, wenn f' monoton wachsend (fallend) ist.
- Eine Funktion f ist konvex (konkav) genau dann, wenn $f''\geq 0$ ($f''\leq 0$).

Zu zeigen ist somit:

- $F(t)$ ist vom Typ DFR $\iff \frac{d}{dt}\ln\overline{F}(t)$ ist monoton wachsend.
- $F(t)$ ist vom Typ IFR $\iff \frac{d}{dt}\ln\overline{F}(t)$ ist monoton fallend.

Es wird hier nur der Typ DFR betrachtet. Für den Typ IFR gilt die entsprechende Argumentation.
Es sei $F(t)$ vom Typ DFR, d.h. für alle beliebig fest gewählten $t \in \mathbb{R}^+$ ist $\overline{F}_{t_0} := \frac{\overline{F}(t+t_0)}{\overline{F}(t_0)}$ monoton wachsend in t_0. Dies ist äquivalent zu

$$\begin{aligned}
& \ln\frac{\overline{F}(t+t_0)}{\overline{F}(t_0)} = \ln\overline{F}(t+t_0) - \ln\overline{F}(t_0) \text{ ist monoton wachsend in } t_0 \ \forall\ t \in \mathbb{R}^+ \\
\iff\ & \frac{\ln\overline{F}(t+t_0) - \ln\overline{F}(t_0)}{t} \text{ ist monoton wachsend in } t_0 \ \forall\ t \in \mathbb{R}^+ \\
\iff\ & \frac{d}{dt_0}\ln\overline{F}(t_0) \text{ ist monoton wachsend in } t_0
\end{aligned}$$

Die Richtung „$\Longrightarrow$" in der letzten Äquivalenz folgt durch Grenzübergang und die Richung „$\Longleftarrow$" ergibt sich, da $\frac{d}{dt_0}\ln\overline{F}(t_0)$ für alle $t_0 \in \mathbb{R}^+$ monoton wachsend ist. ∎

Zusammengesetzte Lebensdauerverteilungen

Die typische Form einer Ausfallrate ist die sogenannte Badewannenkurve. Die Kurve besteht aus drei Bereichen: Am Anfang steht ein hoher, schnell abklingender Bereich. Dies ist die Zeit der Kinderkrankheiten. Der mittlere Bereich hat eine fast konstante Ausfallrate, während zum Schluss die Ausfallrate durch Alterserscheinungen wieder ansteigt.

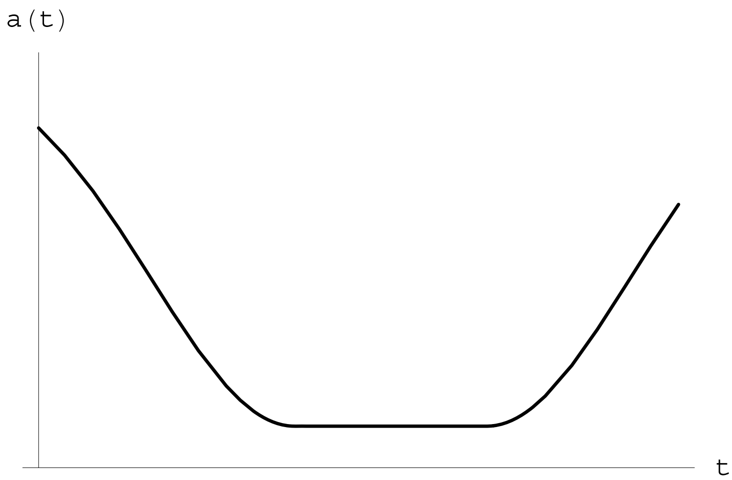


Abbildung 10.1: typischer Ausfallratenverlauf; sogenannte Badenwannenkurve

Bei der Betrachtung der Ausfallraten der bisherigen Verteilungsfunktionen lässt sich feststellen, dass keine dieser Funktionen einer Badewannenkurve ähnlich ist. Um ein solches Ausfallverhalten nachbilden zu können, müssen mehrere Lebensdauerverteilungen kombiniert werden.

Eine zusammengesetzte Verteilungsfunktion lässt sich wie folgt definieren:

Es seien $f_1, \dots, f_k$ Dichten von Verteilungsfunktionen, $p_i \geq 0$ für $i = 1, \dots, k$. Ist die Summe aller p_i gleich eins, so definiert

$$f(t) := \sum_{i=1}^{k} p_i \cdot f_i(t)$$

die zusammengesetzte Dichte. Die zugehörige Verteilungsfunktion lautet

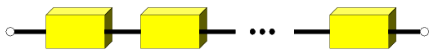
$$F(t) := \sum_{i=1}^{k} p_i \cdot F_i(t).$$

10.2 Zuverlässigkeit von zusammengesetzten Systemen

Dieser Abschnitt behandelt die Zuverlässigkeit zusammengesetzter Systeme. Als Grundmuster fungieren Serien– und die Parallelschaltungen. Mit Hilfe dieser Grundmuster lassen sich anschließend komplexere Systeme bilden und analysieren.

Zuverlässigkeit von Seriensystemen

Das einfachste zusammengesetzte System ist die Serienschaltung.



Die angegebene Serienschaltung habe die Länge n. Das System ist intakt, wenn alle Komponenten intakt sind. Es sei X_i für $i = 1, \dots, n$ die Lebensdauer der Komponente i. Y bezeichne die Lebensdauer des Gesamtsystems. Außerdem wird angenommen, dass die X_i stochastisch unabhängig seien. Es gilt dann:

$$Y = \min(X_1, \dots, X_n).$$

Die Lebensdauerverteilung des Seriensystems berechnet sich damit wie folgt:

$$\begin{aligned} F_Y(t) &= P(Y \leq t) \;=\; 1 - P(Y > t) \;=\; 1 - P(\min(X_1, \dots, X_n) > t) \\ &= 1 - P(X_1 > t, \dots, X_n > t) \;=\; 1 - [P(X_1 > t) \cdot \dots \cdot P(X_n > t)] \\ &= 1 - [(1 - F_{X_1}(t)) \cdot \dots \cdot (1 - F_{X_n}(t))] \;=\; 1 - \prod_{i=1}^{n} \bar{F}_i(t), \quad t \in \mathbb{R}^+. \end{aligned}$$

Die Zuverlässigkeitsfunktion des Gesamtsystems lautet:

$$\overline{F}(t) = 1 - F_Y(t) = \prod_{i=1}^{n} \bar{F}_i(t), \quad t \in \mathbb{R}^+.$$

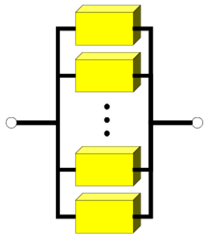
10.9 Bemerkung:
Aus $\overline{F}(t) = \bar{F}_1(t) \cdot \dots \cdot \bar{F}_n(t)$ folgt

$$\exp\left(-\int_0^t a_Y(u)du\right) = \exp\left(-\left[\int_0^t a_1(u)du + \dots + \int_0^t a_n(u)du\right]\right)$$

und somit $a_Y(t) = a_1(t) + \cdots + a_n(t)$, d.h. im Fall eines Seriensystems addieren sich die Ausfallraten.

Zuverlässigkeit von Parallelsystemen

Ein anderes einfaches zusammengesetztes System ist die Parallelschaltung.



Wiederum bestehe die Schaltung aus n Komponenten und die Lebensdauern seien gegeben durch die Zufallsgrößen X_i, $i = 1, \ldots, n$. Die Lebensdauer des Gesamtsystems sei Z und die X_i seien wieder stochastisch unabhängig. Die Parallelschaltung ist solange intakt, bis auch die letzte Komponente ausgefallen ist, d.h. es gilt

$$Z = \max(X_1, \ldots, X_n).$$

Die Lebensdauerverteilung ergibt sich damit zu

$$\begin{aligned} F_Z(t) &= P(Z \leq t) \;=\; P(\max(X_1, \ldots, X_n) \leq t) \;=\; P(X_1 \leq t, \ldots, X_n \leq t) \\ &= P(X_1 \leq t) \cdot \cdots \cdot P(X_n \leq t) \;=\; F_{X_1}(t) \cdot \cdots \cdot F_{X_n}(t), \quad t \in \mathbb{R}^+. \end{aligned}$$

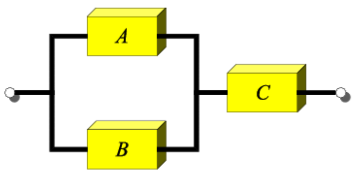
Die Zuverlässigkeitsfunktion des Gesamtsystems lässt sich also berechnen durch

$$\overline{F}(t) = 1 - F_Z(t) = 1 - \prod_{i=1}^{n} (1 - \bar{F}_i(t)), \quad t \in \mathbb{R}^+.$$

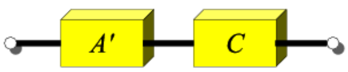
Allgemeine Schaltbilder

Teilsysteme

Allgemeine Schaltbilder, etwa von der nachstehenden Form, können durch eine Zerlegung in Teilsysteme gelöst werden.



In diesem Fall ist eine Dekomposition in die Teilsysteme A' und C naheliegend:



Die Zuverlässigkeitsfunktion berechnet sich nun primär nach dem Gesetz für eine Serienschaltung. In Kurzform ergibt sich

$$\overline{F}(t) = \bar{F}_{A'}(t) \cdot \bar{F}_C(t), \quad t \in \mathbb{R}^+,$$

wobei sich $\bar{F}_{A'}(t)$ aus der Parallelschaltung von A und B errechnet:

$$\bar{F}_{A'}(t) = 1 - [(1 - \bar{F}_A(t))(1 - \bar{F}_B(t))] = \bar{F}_A(t) + \bar{F}_B(t) - \bar{F}_A(t) \cdot \bar{F}_B(t), \quad t \in \mathbb{R}^+.$$

Daraus resultiert:

$$\begin{aligned}\overline{F}(t) &= (\bar{F}_A(t) + \bar{F}_B(t) - \bar{F}_A(t) \cdot \bar{F}_B(t)) \cdot \bar{F}_C(t) \\ &= \bar{F}_A(t) \cdot \bar{F}_C(t) + \bar{F}_B(t) \cdot \bar{F}_C(t) - \bar{F}_A(t) \cdot \bar{F}_B(t) \cdot \bar{F}_C(t), \quad t \in \mathbb{R}^+.\end{aligned}$$

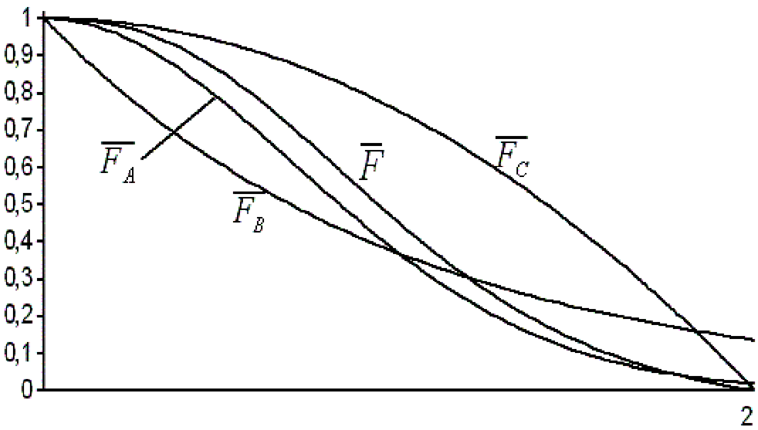


Abbildung 10.2: Überlebensfunktion $\overline{F}$ mit $\overline{F}_A(t) := \exp(-t^2)$, $\overline{F}_B(t) := \exp(-t)$ und $\overline{F}_C(t) := \left(1 - \frac{t^2}{4}\right) \cdot I_{[0,2]}(t)$, $t \in \mathbb{R}^+$.

10.10 Bemerkung:
Die Argumentation lässt sich auch umkehren. So kann man die Ausgangsschaltung auch als Erweiterung einer einfachen Serienschaltung verstehen, in der ein Element durch eine Parallelschaltung mehrerer Elemente ersetzt wird. Wie obige Rechnung zeigt, wird dadurch die Zuverlässigkeit der Serienschaltung vergrößert. Dies führt zu einem wesentlichen Ergebnis der Zuverlässigkeitstheorie:
Die Zuverlässigkeit eines Systems lässt sich durch die Erhöhung der Zuverlässigkeit seiner Teilsysteme vergrößern.

Operationspfade

Eine weitere Möglichkeit zur Berechnung der Zuverlässigkeitsfunktion allgemeiner Schaltbilder stellen die sogenannten Operationspfade dar. Ein Operationspfad ist eine Serienschaltung von Bauelementen, die, wenn sie alle intakt sind, auch das System intakt halten. In dem obigen Beispiel gibt es zwei mögliche Operationspfade: $X := AC$ und $Y := BC$. Damit funktioniert das Gesamtsystem noch, falls Operationspfad X intakt ist, die Bauelemente A und

C also funktionieren. Entsprechendes gilt für Y. Das System ist also intakt, wenn entweder X intakt oder Y intakt, oder sowohl X, als auch Y intakt sind.
Die Zuverlässigkeit des Gesamtsystems ist somit

$$P(\{\text{„}X\text{ intakt"}\} \cup \{\text{„}Y\text{ intakt"}\}).$$

Diese Wahrscheinlichkeit soll kurz mit $P(X \cup Y)$ bezeichnt werden. Es gilt:

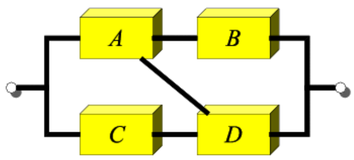
$$\begin{aligned} P(X \cup Y) &= P(X) + P(Y) - P(X \cap Y) \\ &= P(AC) + P(BC) - P(AC \cap BC) \\ &= P(A) \cdot P(C) + P(B) \cdot P(C) - P(A) \cdot P(B) \cdot P(C), \end{aligned}$$

Womit sich dasselbe Gesamtergebnis wie in Abschnitt „Teilsysteme" ergibt:

$$\overline{F}(t) = \bar{F}_A(t) \cdot \bar{F}_C(t) + \bar{F}_B(t) \cdot \bar{F}_C(t) - \bar{F}_A(t) \cdot \bar{F}_B(t) \cdot \bar{F}_C(t).$$

Brückenschaltung

In einer Brückenschaltung



gibt es drei Operationspfade: $X := AB$, $Y := AD$ und $Z := CD$. Die Intaktwahrscheinlichkeit bzw. die Zuverlässigkeit lässt sich nach dem oben beschriebenen Verfahren wie folgt berechnen:

$$\begin{aligned} P(X \cup Y \cup Z) &= P(X) + P(Y) + P(Z) - P(X \cap Y) - P(X \cap Z) - P(Y \cap Z) \\ &\quad + P(X \cap Y \cap Z) \\ &= P(AB) + P(AD) + P(CD) - P(AB \cap AD) - P(AB \cap CD) \\ &\quad - P(AD \cap CD) + P(AB \cap AD \cap CD) \\ &= P(A) \cdot P(B) + P(A) \cdot P(D) + P(C) \cdot P(D) - P(A) \cdot P(B) \cdot P(D) \\ &\quad - P(A) \cdot P(B) \cdot P(C) \cdot P(D) - P(A) \cdot P(C) \cdot P(D) \\ &\quad + P(A) \cdot P(B) \cdot P(C) \cdot P(D) \\ &= P(A) \cdot P(B) + P(A) \cdot P(D) + P(C) \cdot P(D) - P(A) \cdot P(B) \cdot P(D) \\ &\quad - P(A) \cdot P(C) \cdot P(D). \end{aligned}$$

Zusammenfassend ergibt sich:

$$\begin{aligned} \overline{F}(t) &= \bar{F}_A(t) \cdot \bar{F}_B(t) + \bar{F}_A(t) \cdot \bar{F}_D(t) + \bar{F}_C(t) \cdot \bar{F}_D(t) \\ &\quad -\bar{F}_A(t) \cdot \bar{F}_B(t) \cdot \bar{F}_D(t) \\ &\quad -\bar{F}_A(t) \cdot \bar{F}_C(t) \cdot \bar{F}_D(t), \quad t \in \mathbb{R}^+. \end{aligned}$$

(Siehe auch PowerPoint-Präsentation zur Brückenschaltung.)
Die Berechnung lässt sich mit Hilfe des folgenden Schemas vereinfachen:

(i) Erstelle eine Tabelle mit jeweils einer Spalte pro Operationspfad und Systemkomponente, sowie eine Spalte für das Vorzeichen.

(ii) Trage alle möglichen Kombinationen der Operationspfade mit den daran beteiligten Komponenten ein.

(iii) Bewerte eine ungerade Anzahl Operationen mit dem Vorzeichen Plus, eine gerade Anzahl mit Minus.

Pfade			Einheiten					Vorzeichen
X	Y	Z	A	B	C	D		
X			A	B				$+$
	Y		A			D		$+$
		Z			C	D		$+$
X	Y		A	B		D		$-$
X		Z	A	B	C	D	$\surd$	$-$
	Y	Z	A		C	D		$-$
X	Y	Z	A	B	C	D	$\surd$	$+$

Die mit einem Haken versehenen Zeilen bestehen aus denselben beteiligten Komponenten mit jeweils unterschiedlichen Vorzeichen. Diese werden im Folgenden nicht mehr berücksichtigt. Die verbleibenden Zeilen enthalten nach Aufsummieren mit dem angegebenen Vorzeichen die Formel für die Intaktwahrscheinlichkeit.

Literatur zu Kapitel 10

Folgende Bücher werden als begleitende Literatur zu diesem Kapitel des Skriptes empfohlen:

- F. Beichelt:
  Zuverlässigkeits– und Instandhaltungstheorie,
  Teubner, Stuttgart, 1993.
  ISBN: 3519029855
- K.–W. Gaede:
  Zuverlässigkeit — Mathematische Modelle,
  Hanser, München, 1977.
  ISBN: 3446123709

Kapitel 11

Produkträume und mehrdimensionale Zufallsvariable

Mehrdimensionale Zufallsvariablen dienen zur Erfassung von Zufallsexperimenten, bei denen mehrere Größen gleichzeitig beobachtet werden. In diesem Kapitel geht es im Wesentlichen darum, die für eindimensionale Zufallsvariablen eingeführten Begriffe auf mehrdimensionale Zufallsvariablen zu übertragen. Bei der gleichzeitigen Beobachtung verschiedener Zufallsvariablen spielt außerdem der Begriff der stochastischen Unabhängigkeit eine wichtige Rolle.

Schlüsselwörter: Produkt–σ–Algebra, mehrdimensionale Verteilungsfunktion, Randverteilungen, Produktmaß, Unabhängigkeit von Zufallsgrößen, Transformation von Zufallsvariablen, Transformationssatz für Dichten, Faltung, Satz von Fubini, Kovarianz.

11.1 Zufällige Vektoren

Zusammengesetzte Stichprobenräume, bzw. sogenannte Produkträume $\Omega := \Omega_1 \times \cdots \times \Omega_n$ ergeben sich, wenn aufeinanderfolgende oder parallele Beobachtungen eines Zufallsexperiments vorliegen. Einige solcher Situationen wurden bereits betrachtet:

1. n–maliges Werfen mit einem Würfel. In diesem Fall ist $\Omega_i := \{1,2,3,4,5,6\}$ für $i = 1,\ldots,n$ und $\Omega := \Omega_1 \times \cdots \times \Omega_n = \{1,2,\ldots,6\}^n$
2. Gruppierte Daten aus einer Vorsorgeuntersuchung
(Gewicht, Körpergröße, Blutdruck, Cholesterin–Spiegel, ...)
3. Belegungszustand eines Produktionssystems
(Pufferbelegung an Arbeitsstation 1, Pufferbelegung an Arbeitsstation 2, ...)
4. Räuber–Beute–Modell
(Anzahl der Räuber und Anzahl der Beutetiere jeweils zu einem bestimmten Zeitpunkt)

Um die genannten Fälle behandeln zu können, wird zunächst eine zum kartesischen Produkt $\Omega := \Omega_1 \times \cdots \times \Omega_n$ passende σ–Algebra benötigt. Dazu werden Messräume $(\Omega_i, \mathfrak{F}_i)$, $i = 1,\ldots,n$, betrachtet. Angesichts der Darstellung

$$\Omega := \bigtimes_{i=1}^{n} \Omega_i$$

könnte man versucht sein,

$$\mathfrak{F} := \bigtimes_{i=1}^{n} \mathfrak{F}_i$$

als σ–Algebra über Ω zu verwenden. Doch im Allgemeinen führt das kartesische Produkt von σ–Algebren nicht wieder zu einer σ–Algebra, wie das folgende Beispiel zeigt:

Es werden $n := 2$, $A := [-1,0] \times [-1,0]$ und $B := [0,1] \times [0,1]$ gewählt. Wie man sich anhand von Abbildung 11.1 schnell überzeugt, ist $A \cup B$ kein kartesisches Produkt zweier Teilmengen von $\mathbb{R}$.

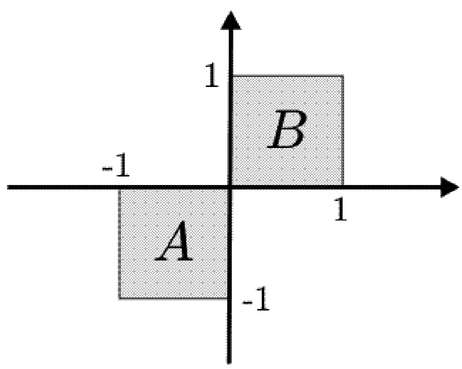


Abbildung 11.1: $A \cup B$ ist kein kartesisches Produkt.

Deswegen benutzt man die folgende Begriffsbildung:

11.1 Definition (Produkt–$\boldsymbol{\sigma}$–Algebra):
Es seien $(\Omega_i, \mathfrak{F}_i)$, $i = 1, \ldots, n$, Messräume und

$$\mathfrak{E}^n := \left\{ \bigtimes_{i=1}^{n} A_i \mid A_i \in \mathfrak{F}_i,\ i = 1, \ldots, n \right\}$$

das System aller kartesischen Produkte aus den Mengen der gegebenen σ–Algebren. Dann heißt die durch

$$\bigotimes_{i=1}^{n} \mathfrak{F}_i := \sigma(\mathfrak{E}^n)$$

definierte σ–Algebra über $\Omega := \bigtimes_{i=1}^{n} \Omega_i$ die Produkt–σ–Algebra der σ–Algebren $\mathfrak{F}_1, \ldots, \mathfrak{F}_n$.
Der Spezialfall $\Omega_i := \mathbb{R}^1$, $\mathfrak{F}_i := \mathfrak{B}^1$, $i = 1, \ldots, n$, führt auf die σ–Algebra der Borelschen Mengen des $\mathbb{R}^n$.

11.2 Satz (Rechenregeln für Rechteckmengen):
Es seien $A_1, A_{1i} \in \Omega_1$, $i \in I_1$ und $A_2, A_{2i} \in \Omega_2$, $i \in I_2$. Dann gilt:

a) $\bigcap\limits_{j=1}^{\infty} (A_{1j} \times A_{2j}) = \left(\bigcap\limits_{j=1}^{\infty} A_{1j} \right) \times \left(\bigcap\limits_{j=1}^{\infty} A_{2j} \right)$,

b) $\overline{(A_1 \times A_2)} = (\overline{A_1} \times \Omega_2) \cup (A_1 \times \overline{A_2}) = (\Omega_1 \times \overline{A_2}) \cup (\overline{A_1} \times A_2)$,

c) $\bigcup\limits_{j=1}^{\infty} (A_{1j} \times A_2) = \left(\bigcup\limits_{j=1}^{\infty} A_{1j} \right) \times A_2$,

$\bigcup\limits_{j=1}^{\infty} (A_1 \times A_{2j}) = A_1 \times \left(\bigcup\limits_{j=1}^{\infty} A_{2j} \right)$,

d) $\bigcup\limits_{j=1}^{\infty} (A_{1j} \times A_{2j}) \subset \left(\bigcup\limits_{j=1}^{\infty} A_{1j} \right) \times \left(\bigcup\limits_{j=1}^{\infty} A_{2j} \right)$,

e) $A_1 \times A_2 = \varnothing \Leftrightarrow$ *wenn* $A_1 = \varnothing$ *oder* $A_2 = \varnothing$,

f) $A_1 \times A_2 \subset B_1 \times B_2 \Leftrightarrow A_1 \subset B_1$ *und* $A_2 \subset B_2$ $(A_1, A_2 \neq \varnothing)$,

Beweis:

a)

$$\begin{aligned}
& (\omega_1, \omega_2) \in \bigcap_{j=1}^{\infty} (A_{1j} \times A_{2j}) \\
\Leftrightarrow\ & (\omega_1, \omega_2) \in (A_{1j} \times A_{2j})\ \forall\ j \in \mathbb{N} \\
\Leftrightarrow\ & \omega_1 \in A_{1j}, \omega_2 \in A_{2j}\ \forall\ j \in \mathbb{N} \\
\Leftrightarrow\ & \omega_1 \in \bigcap_{j=1}^{\infty} A_{1j}\ \wedge\ \omega_2 \in \bigcap_{j=1}^{\infty} A_{2j} \\
\Leftrightarrow\ & (\omega_1, \omega_2) \in \left(\bigcap_{j=1}^{\infty} A_{1j} \right) \times \left(\bigcap_{j=1}^{\infty} A_{2j} \right).
\end{aligned}$$

b) Es gilt

$$\begin{aligned}
&(\omega_1,\omega_2) \in \overline{A_1 \times A_2} \Leftrightarrow (\omega_1,\omega_2) \notin (A_1 \times A_2) \\
\Leftrightarrow\ & \omega_1 \in A_1 \ \wedge\ \omega_2 \notin A_2 \text{ oder } \omega_1 \notin A_1 \ \wedge\ \omega_2 \in \Omega_2 \\
\Leftrightarrow\ & (\omega_1,\omega_2) \in (A_1 \times \overline{A_2}) \text{ oder } (\omega_1,\omega_2) \in (\overline{A_1} \times \Omega_2) \\
\Leftrightarrow\ & (\omega_1,\omega_2) \in (A_1 \times \overline{A_2}) \cup (\overline{A_2} \times \Omega_2).
\end{aligned}$$

Entsprechend c)-f) (Übung!). ■

Konstruktion einer Produkt–σ–Algebra

Betrachte den zweidimensionalen Fall mit den beiden Messräumen $(\Omega_1, \mathfrak{F}_1)$ und $(\Omega_2, \mathfrak{F}_2)$. Konstruiere eine geeignete σ–Algebra $\mathfrak{F}_1 \otimes \mathfrak{F}_2$ zum Produktraum $\Omega_1 \times \Omega_2$:
Betrachte die sogenannten **Zylindermengen**

$$\mathfrak{F}_{(1)} := \{A_1 \times \Omega_2 \mid A_1 \in \mathfrak{F}_1\} \ \text{ und } \ \mathfrak{F}_{(2)} := \{\Omega_1 \times A_2 \mid A_2 \in \mathfrak{F}_2\}.$$

Mit Hilfe von Satz 11.2 lässt sich schnell nachprüfen, dass $\mathfrak{F}_{(1)}$ und $\mathfrak{F}_{(2)}$ σ-Algebren über $\Omega_1 \times \Omega_2$ darstellen. Zu zeigen ist: $\mathfrak{F}_1 \otimes \mathfrak{F}_2 = \sigma(\{\mathfrak{F}_{(1)} \cup \mathfrak{F}_{(2)}\})$.

- Weil die Zylindermengen spezielle Mengen von $\mathfrak{F}_1 \otimes \mathfrak{F}_2$ sind, gilt

$$\sigma(\{\mathfrak{F}_{(1)} \cup \mathfrak{F}_{(2)}\}) \subseteq \mathfrak{F}_1 \otimes \mathfrak{F}_2.$$

- Da sich andererseits aber jede Rechteckmenge $A_1 \times A_2$ als Durchschnitt zweier Zylindermengen darstellen lässt, nämlich in der Form

$$(A_1 \times A_2) = (A_1 \times \Omega_2) \cap (\Omega_1 \times A_2),$$

  gilt

$$\{A_1 \times A_2 \mid A_i \in \mathfrak{F}_i, i = 1,2\} \subseteq \sigma(\{\mathfrak{F}_{(1)} \cup \mathfrak{F}_{(2)}\})$$

  und damit

$$\sigma(\{A_1 \times A_2 \mid A_i \in \mathfrak{F}_i, i = 1,2\}) = \mathfrak{F}_1 \otimes \mathfrak{F}_2 \subseteq \sigma(\{\mathfrak{F}_{(1)} \cup \mathfrak{F}_{(2)}\}).$$

Wie im eindimensionalen Fall werden Wahrscheinlichkeitsmaße auf der Borelschen σ–Algebra $\mathfrak{B}^n$ durch sogenannte n–dimensionale Verteilungsfunktionen erzeugt.

11.3 Definition (n–dimensionale Verteilungsfunktion):
Es sei P ein Wahrscheinlichkeitsmaß auf der σ–Algebra $\mathfrak{B}^n$. Die durch

$$F_P(x_1, \ldots, x_n) := P\left(\bigtimes_{i=1}^{n} (-\infty, x_i]\right), \qquad (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

definierte Funktion heißt die zu P gehörende n–dimensionale Verteilungsfunktion.

Die Funktion F_P ist durch eine Reihe von Eigenschaften, die im Folgenden betrachtet werden sollen, ausgezeichnet.

11.4 Satz:
Es sei $F := F_P$ eine Verteilungsfunktion über $(\mathbb{R}^n, \mathfrak{B}^n)$. Dann gilt:

a) Δ-Monotonie:
Für alle $a := (a_1, \ldots, a_n)$ und $b := (b_1, \ldots, b_n) \in \mathbb{R}^n$ mit $a \leq b$ (d.h. $a_i \leq b_i, 1 \leq i \leq n$) gilt:

$$\Delta_n^{a,b} F := \sum_{(\theta_1,\ldots,\theta_n)\in\{0,1\}^n} (-1)^{\sum_{i=1}^n \theta_i} F(\theta_1 a_1 + (1-\theta_1)b_1, \ldots, \theta_n a_n + (1-\theta_n)b_n) \geq 0.$$

b) F ist in jeder Variable rechtsseitig stetig, d.h. es gilt:

$$\lim_{x_1\to y_1+0,\ldots,x_n\to y_n+0} F(x_1, \ldots, x_n) = F(y_1, \ldots, y_n) \quad \forall\ (y_1, \ldots, y_n) \in \mathbb{R}^n.$$

c) F ist normiert, d.h. es gilt:

$$\lim_{x_1\to+\infty,\ldots,x_n\to+\infty} F(x_1, \ldots x_n) = 1$$

und für $(x_1, \ldots, x_n) \in \mathbb{R}^n$ und jedes $i \in \{1, \ldots, n\}$ gilt:

$$\lim_{x_i\to-\infty} F(x_1, \ldots, x_n) = 0\,.$$

Beweis:
Für $n := 1$ reduziert sich die Aussage a) auf den bereits bekannten Fall (vgl. Satz 7.2) und es ist

$$\Delta_1^{a,b} F = F(b) - F(a) \geq 0 \quad \text{für} \quad b \geq a.$$

Für $n := 2$ erhält man

$$\Delta_2^{a,b} F = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2). \tag{11.1}$$

Es wird gezeigt, dass $\Delta_2^{a,b} F$ mit $P((a,b]_{(2)})$ zusammenfällt. Dafür wird die Abbildung 11.2 benutzt. Offensichtlich gilt:

$$\begin{aligned} ((a_1, a_2), (b_1, b_2)] \quad &= \quad \{((-\infty, -\infty), (b_1, b_2)] \setminus ((-\infty, -\infty), (a_1, b_2)]\} \\ &\quad\ \setminus \{((-\infty, -\infty), (b_1, a_2)] \setminus ((-\infty, -\infty), (a_1, a_2)]\}\,. \end{aligned}$$

$$\begin{aligned} P((a,b]_{(2)}) &= P(((a_1, a_2), (b_1, b_2)]) \\ &= P(((a_1, -\infty), (b_1, b_2)]) - P(((a_1, -\infty), (b_1, a_2)]) \\ &= P(((-\infty, -\infty), (b_1, b_2)]) - P(((-\infty, -\infty), (a_1, b_2)]) \\ &\ \ - P(((-\infty, -\infty), (b_1, a_2)]) + P(((-\infty, -\infty), (a_1, a_2)]) \\ &= F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \\ &= \Delta_2^{a,b} F \end{aligned}$$

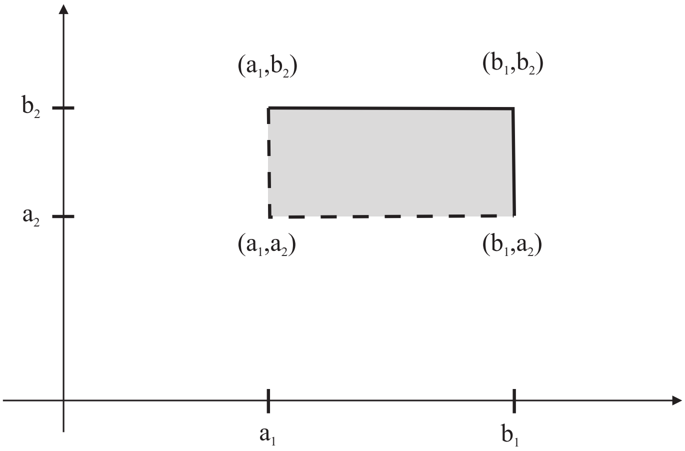


Abbildung 11.2: Zur Berechnung von $P((a, b]_{(2)})$.

in Übereinstimmung mit (11.1). Für $n \geq 3$ verlaufen die Beweise analog, sie werden jedoch mit wachsendem n formal aufwendiger.

b) folgt aus der Beziehung

$$\begin{aligned}
&\lim_{\substack{k_1\to\infty,\dots,\\ k_n\to\infty}} F\left(x_1 + \frac{1}{k_1}, \dots, x_n + \frac{1}{k_n}\right)\\
&= \lim_{\substack{k_1\to\infty,\dots,\\ k_n\to\infty}} P\left(\left((-\infty, \dots, -\infty), \left(x_1 + \frac{1}{k_1}, \dots, x_n + \frac{1}{k_n}\right)\right]\right)\\
&= P\left(\bigcap_{\substack{k_1=1,\dots,\\ k_n=1}}^{\infty} \left((-\infty, \dots, -\infty), \left(x_1 + \frac{1}{k_1}, \dots, x_n + \frac{1}{k_n}\right)\right]\right)\\
&= P(((-\infty, \dots, -\infty), (x_1, \dots, x_n)]) = F(x_1, \dots, x_n)\,.
\end{aligned}$$

Die unter c) beschriebenen Beziehungen ergeben

$$\begin{aligned}
\lim_{x_1\to\infty,\dots,x_n\to\infty} F(x_1, \dots, x_n) &= \lim_{x_1\to\infty,\dots,x_n\to\infty} P(((-\infty, \dots, -\infty), (x_1, \dots, x_n)])\\
&= P\left(\bigcup_{x_1=1,\dots,x_n=1}^{\infty} ((-\infty, \dots, -\infty), (x_1, \dots, x_n)]\right)\\
&= P(\mathbb{R}^n) = 1\,,\\
\lim_{x_i\to-\infty} F(x_1, \dots, x_i, \dots, x_n) &= \lim_{k\to\infty} F(x_1, \dots, -k, \dots, x_n)\\
&= \lim_{k\to\infty} P(((-\infty, \dots, -\infty), (x_1, \dots, -k, \dots, x_n)])\\
&= P\left(\bigcap_{k=1}^{\infty} ((-\infty, \dots, -\infty), (x_1, \dots, -k, \dots, x_n)]\right)\\
&= P(\varnothing) = 0\,.
\end{aligned}$$

■

11.5 Definition (maßdefinierende Funktion, Verteilungsfunktion):
Eine messbare Funktion $F\colon \mathbb{R}^n \to \mathbb{R}$ heißt maßdefinierende Funktion über $\mathbb{R}^n$, falls sie Δ-monoton und rechtsseitig stetig ist. Sie heißt Verteilungsfunktion über $\mathbb{R}^n$, falls sie außerdem normiert ist.

11.6 Satz:
Zu jeder maßdefinierenden Funktion F über $\mathbb{R}^n$ gibt es genau ein Maß μ_F auf $(\mathbb{R}^n, \mathfrak{B}^n)$ mit

$$\mu_F((a,b]_{(n)}) = \Delta_n^{a,b} F \quad \forall\, a,b \in \mathbb{R}^n \text{ mit } a \leq b.$$

Ist F eine Verteilungsfunktion über $\mathbb{R}^n$, so ist μ_F ein Wahrscheinlichkeitsmaß über $(\mathbb{R}^n, \mathfrak{B}^n)$, das mit P_F bezeichnet wird.

Beweis:
Zu zeigen sind:

1. Existenz von μ_F.
2. Ist F Verteilungsfunktion, so ist μ_F Wahrscheinlichkeitsmaß.

Der Beweis folgt dabei der Argumentation wie im eindimensionalen Fall (vgl. Satz 7.2).

Zu 1.: Die Δ-Monotonie von F bedingt zunächst, dass μ_F auf dem Semiring $\mathbb{I}^n$ der links offenen und rechts abgeschlossenen Intervalle des $\mathbb{R}^n$ einen Inhalt definiert.

(a) Ist μ_F σ–additiv und somit ein Prämaß, kann μ_F aufgrund des ersten Maßfortsetzungssatzes eindeutig zu einem Prämaß auf dem von $\mathbb{I}^n$ erzeugten Ring $\mathfrak{R}(\mathbb{I}^n)$ fortgesetzt werden.
Die σ–Additivität folgt aus der rechtsseitigen Stetigkeit von F.

(b) Ist μ_F σ–endlich, so existiert nach Aussage des zweiten Fortsetzungssatzes für Maße eine eindeutige Fortsetzung zu einem Maß auf der von $\mathbb{I}^n$ erzeugten σ–Algebra $\sigma(\mathbb{I}^n) = \mathfrak{B}^n$.
Die σ–Endlichkeit folgt wegen

$$\mathbb{R}^n = \bigcup_{k=1}^{\infty} \underbrace{(-k,k] \times \ldots \times (-k,k]}_{n\text{-mal}}$$

und

$$P(\underbrace{(-k,k] \times \ldots \times (-k,k]}_{n\text{-mal}}) = \Delta_n^{-k,k} F < \infty \qquad \forall\, k \in \mathbb{N}.$$

Zu 2.: Der Beweis verläuft analog zur entsprechenden Aussage von vgl. Satz 7.2. ■

Offensichtlich kann aus einer Verteilungsfunktion F über $\mathbb{R}^n$ durch Einschränkung auf eine Komponente $i \in \{1,\ldots,n\}$ vermöge

$$\begin{aligned} F_i(x) &:= P_F(\mathbb{R}^{i-1} \times (-\infty, x] \times \mathbb{R}^{n-i}) \\ &= \lim_{\substack{x_k \to \infty \\ k \neq i}} F(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n), \quad x \in \mathbb{R}, \end{aligned}$$

eine Verteilungsfunktion F_i über $\mathbb{R}$ erzeugt werden, denn wegen Satz 11.4 c) gelten

$$\lim_{x\to\infty} F_i(x) = 1 \qquad \text{und} \qquad \lim_{x\to-\infty} F_i(x) = 0.$$

Die übrigen Eigenschaften einer Verteilungsfunktion folgen aus der Δ-Monotonie und der rechtsseitigen Stetigkeit von F.

11.7 Definition (Randverteilungsfunktion, Randverteilung):
Gegeben sei ein Wahrscheinlichkeitsmaß P über $(\mathbb{R}^n, \mathfrak{B}^n)$. Die zugehörige n-dimensionale Verteilungsfunktion sei F. Für $i \in \{1, \ldots, n\}$ wird

$$\begin{aligned} F_i(x) &:= P_F(\mathbb{R}^{i-1} \times (-\infty, x] \times \mathbb{R}^{n-i}) \\ &= \lim_{\substack{x_k\to\infty \\ k\neq i}} F(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n), \quad x \in \mathbb{R}, \end{aligned}$$

die <u>i−te Randverteilungsfunktion</u> von F bzw. P genannt. Das zugehörige Wahrscheinlichkeitsmaß über $(\mathbb{R}, \mathfrak{B})$ heißt <u>i-te Randverteilung</u> von P.

Aus der gemeinsamen Verteilung kann eindeutig auf ihre Randverteilungen geschlossen werden. Die Umkehrung gilt jedoch nicht, wie das nachfolgende Beispiel zeigt.

11.8 Beispiel:
Es wird die zweidimensionale Verteilungsfunktion

$$F(x,y) := \begin{cases} (1-e^{-x})(1-e^{-y}+e^{-x}) & ,0 < x \le y \\ 1-e^{-y} & ,0 < y < x \\ 0 & ,x \le 0 \vee y \le 0 \end{cases}$$

betrachtet.

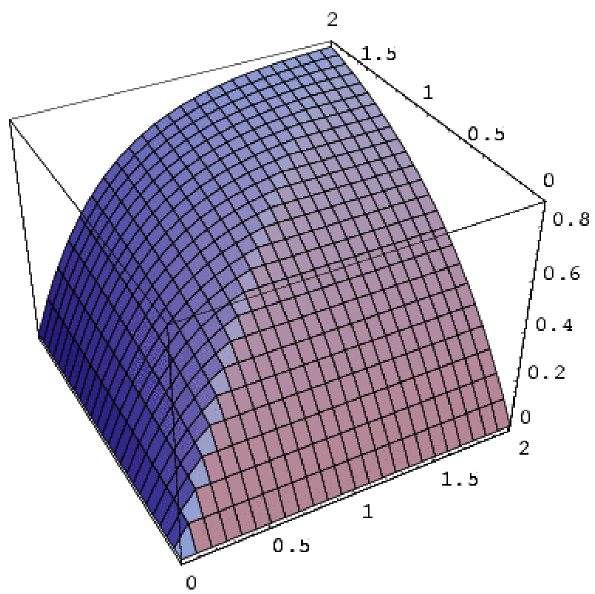


Abbildung 11.3: Verteilungfunktion $F(x,y)$ für $x = 0, \ldots, 5$ und $y = 0, \ldots, 5$.

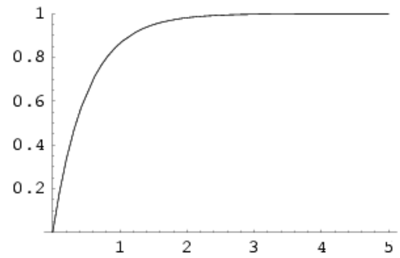


Abbildung 11.4: Randverteilung $F_1(x)$ für $x = 0, \ldots, 5$.

Die zugehörigen Randverteilungsfunktionen lauten:

$$F_1(x) := \lim_{y \to \infty} F(x, y) = \begin{cases} 1 - e^{-2x} & , x > 0 \\ 0 & , x \leq 0 \end{cases}$$

und

$$F_2(y) := \lim_{x \to \infty} F(x, y) = \begin{cases} 1 - e^{-y} & , y > 0 \\ 0 & , y \leq 0. \end{cases}$$

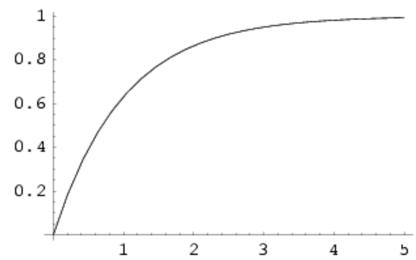


Abbildung 11.5: Randverteilung $F_2(y)$ für $y = 0, \ldots, 5$.

Wie sich sofort nachprüfen lässt, besitzt die Verteilungsfunktion

$$\tilde{F}(x, y) := \begin{cases} (1 - e^{-2x})(1 - e^{-y}) & , x > 0 \wedge y > 0 \\ 0 & , \text{ sonst} \end{cases}$$

dieselben Randverteilungsfunktionen. Doch es ist $\tilde{F} \neq F$.

Genauso wie im eindimensionalen Fall lassen sich auch im mehrdimensionalen Fall Verteilungsfunktionen durch sogenannte Dichten erzeugen.

11.9 Definition (Dichte):
Eine integrierbare Funktion $f : \mathbb{R}^n \to \mathbb{R}$ (bzgl. des Lebesgue-Maßes λ) heißt eine Dichte, wenn gilt:

a) $f(x) \geq 0 \ \forall \ x \in \mathbb{R}^n$,

b) $\int_{\mathbb{R}^n} f \, d\lambda = 1$.

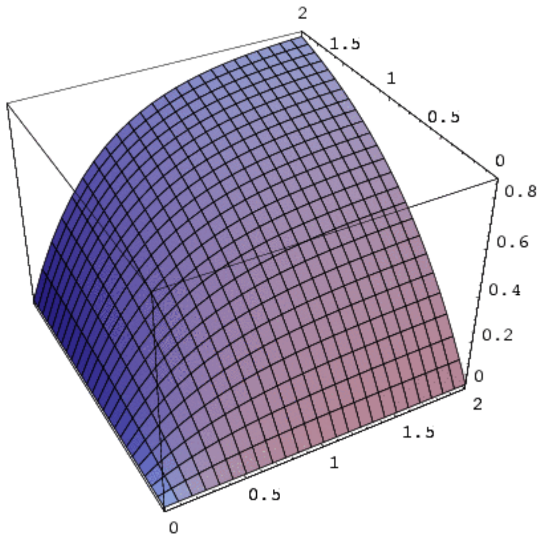


Abbildung 11.6: Verteilungfunktion $\tilde{F}(x, y)$ für $x = 0, \ldots, 5$ und $y = 0, \ldots, 5$.

11.10 Satz:
Es sei $f\colon \mathbb{R}^n \to \mathbb{R}$ *eine Dichte, dann definiert*

$$F(x) := \int\limits_{\times_{i=1}^{n}(-\infty, x_i]} f \, d\lambda \tag{11.2}$$

eine stetige Verteilungsfunktion $F\colon \mathbb{R}^n \to \mathbb{R}$.

Beweis:
Es wird $\nu_f(A) := \int_A f \, d\lambda$ gesetzt und behauptet, dass ν_f ein Maß auf $\mathbb{R}^n$ definiert.

(i) Wegen der Voraussetzung a) ist $\nu_f(A) \geq 0 \ \forall \ A \in \mathfrak{B}^n$.

(ii) ν_f ist σ–additiv, denn mit Hilfe des Satzes von der monotonen Konvergenz 9.42 folgt:

$$\begin{aligned}
\nu_f\left(\bigcup_{j \in \mathbb{N}} A_j\right) &= \int_{\bigcup\limits_{j \in \mathbb{N}} A_j} f \, d\lambda \\
&= \int_{\mathbb{R}^n} f \left(\sum_{j \in \mathbb{N}} I_{A_j}\right) d\lambda \\
&= \lim_{k \to \infty} \int_{\mathbb{R}^n} \sum_{j=1}^{k} f \; I_{A_j} \, d\lambda \\
&= \lim_{k \to \infty} \sum_{j=1}^{k} \int_{\mathbb{R}^n} f \; I_{A_j} \, d\lambda \\
&= \sum_{j=1}^{\infty} \nu_f(A_j).
\end{aligned}$$

(iii) ν_f ist normiert, denn wegen der Voraussetzung b) gilt: $\nu_f(\mathbb{R}^n) = \int_{\mathbb{R}^n} f\, d\lambda \overset{11.9\text{ b)}}{=} 1.$

Folglich definiert $F(x) = \nu_f \left(\bigtimes_{i=1}^{n} (-\infty, x_i] \right)$ eine Verteilungsfunktion.
Die Stetigkeit von F sieht man folgendermaßen: Seien $x, x_o \in \mathbb{R}^n$, $y := \max(x, x_0)$ und $z := \min(x, x_0)$. Dann gilt unter Zuhilfenahme von Satz 6.7:

$$\begin{aligned} |F(x) - F(x_0)| &\leq (F(x) - F(z)) + (F(x_0) - F(z)) \\ &\leq \int_{(z,x]} f\ d\lambda + \int_{(z,x_0]} f\ d\lambda \\ &\leq \nu_f((x_0, y]) + \nu_f((z, x_0]) \underset{x \to x_0}{\longrightarrow} 0, \end{aligned}$$

da $(z, x_0]$ und $(x_0, y]$ für $x \to x_0$ monoton fallend gegen die leere Menge gehen. ■

11.11 Beispiel (Zweidimensionale Normalverteilung):
Es sei

$$f(x,y) := \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left(-\frac{1}{2(1-\varrho^2)} \left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\varrho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right) \right),\ (x,y) \in \mathbb{R}^2,$$

für $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 \in \mathbb{R}^+$ und $\varrho \in (-1, 1)$. Wir zeigen, dass f eine Dichte ist. f heißt Dichte der zweidimensionalen Normalverteilung. In diesem Zusammenhang ist die Beziehung

$$\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2c}\right)\ dx = \sqrt{2\pi c}$$

von Bedeutung, die sich folgendermaßen beweisen lässt:

$$\begin{aligned} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2c}\right)\ dx &= \left(\left(\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2c}\right)\ dx \right)^2 \right)^{\frac{1}{2}} \\ &= \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{x_1^2 + x_2^2}{2c}\right) dx_1 dx_2 \right)^{\frac{1}{2}} \\ &= \left(\int_{0}^{2\pi} \int_{0}^{\infty} \exp\left(-\frac{r^2}{2c}\right) \cdot r\ dr d\varphi \right)^{\frac{1}{2}} \\ &= \Bigg(2\pi \underbrace{\left[-c \exp\left(-\frac{r^2}{2c}\right) \right]_0^{\infty}}_{=0-(-c)} \Bigg)^{\frac{1}{2}} = \sqrt{2\pi c}. \end{aligned}$$

Wir integrieren zuerst nach y und dann nach x:

$$
\begin{aligned}
&\int_{-\infty}^{\infty} f(x,y)dy \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2(1-\varrho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\varrho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right) dy \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2(1-\varrho^2)}\left(\frac{(x-\mu_1)^2(1-\varrho^2)+(x-\mu_1)^2\varrho^2}{\sigma_1^2} - 2\varrho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right) dy \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{1}{2(1-\varrho^2)}\left(\frac{(x-\mu_1)^2\varrho^2}{\sigma_1^2} - 2\varrho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right) dy \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}\right) \exp\left(-\frac{1}{2(1-\varrho^2)\sigma_2^2}\left(y-\mu_2-\frac{(x-\mu_1)\varrho\sigma_2}{\sigma_1}\right)^2\right) dy \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}\right) \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2(1-\varrho^2)\sigma_2^2}\left(y-\left(\mu_2+\frac{(x-\mu_1)\varrho\sigma_2}{\sigma_1}\right)\right)^2\right) dy
\end{aligned}
$$

Mit der Vereinbarung $c_1 := \mu_2 - \frac{(x-\mu_1)\varrho\sigma_2}{\sigma_1}$ und der Substitution $t = y - c_1$ wird hieraus

$$
\begin{aligned}
&\int_{-\infty}^{\infty} f(x,y)dy \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}\right) \underbrace{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2(1-\varrho^2)\sigma_2^2}t^2\right) dt}_{=\sqrt{2\pi(1-\varrho^2)\sigma_2^2}\text{ s.o.}} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{2\pi}\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}\right) \sqrt{2\pi(1-\varrho^2)\sigma_2^2} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}\right).
\end{aligned}
$$

Damit wird

$$
\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x-\mu_1)^2}{\sigma_1^2}\right) dx \\
&\overset{t:=x-\mu_1}{=} \frac{1}{\sqrt{2\pi}\sigma_1} \underbrace{\int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2\sigma_1^2}\right) dt}_{=\sqrt{2\pi\sigma_1^2}\text{ s.o.}} = 1.
\end{aligned}
$$

11.12 Satz (Erzeugung von Zufallsvektoren):
Es sei $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum. Sind $X_1, \ldots, X_n$ reellwertige Zufallsvariablen über $(\Omega, \mathfrak{F}, P)$, so ist $X := (X_1, \ldots, X_n)$ eine $\mathbb{R}^n$-wertige Zufallsvariable bzw. ein n-dimensionaler Zufallsvektor über $(\Omega, \mathfrak{F}, P)$ und umgekehrt.

Beweis:
Aufgrund von Satz 8.11 gilt für alle $i \in \{1, \ldots, n\}$:

$$X_i \text{ ist } \mathfrak{F} - \mathfrak{B}\text{-messbar } \Leftrightarrow \{\omega \mid X_i(\omega) \leq x_i\} \in \mathfrak{F} \;\; \forall\, x_i \in \mathbb{R}.$$

Da die Intervalle $(-\infty, x]_{(n)}, x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, ein Erzeugendensystem der σ-Algebra $\mathfrak{B}^n$ bilden, kann mit Hilfe von Satz 8.9 geschlossen werden:

$$\begin{aligned} X^{-1}(\{(-\infty, x]_{(n)}, x \in \mathbb{R}^n\}) &= \{\omega \mid X_1(\omega) \leq x_1, \ldots, X_n(\omega) \leq x_n\} \\ &= \bigcap_{i=1}^{n} \{\omega \mid X_i(\omega) \leq x_i\} \in \mathfrak{F}, \end{aligned}$$

so dass $X = (X_1, \ldots, X_n)$ $\mathfrak{F} - \mathfrak{B}^n$-messbar ist.
Umgekehrt folgt aus der $\mathfrak{F}$–$\mathfrak{B}^n$–Messbarkeit von X mit $B_i \in \mathfrak{B}^1$ und $X_i^{-1}(B_i) = X^{-1}(\mathbb{R} \times \ldots \times \mathbb{R} \times B_i \times \mathbb{R} \times \ldots \times \mathbb{R}) \subseteq \mathfrak{F}$, dass alle X_i messbar sind. ■

11.13 Definition (Verteilungsfunktion):
Es seien $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum, $X := (X_1, \ldots, X_n)$ ein n-dimensionaler reellwertiger Zufallsvektor über $(\Omega, \mathfrak{F}, P)$ und P_X das Bildmaß von P unter X. Dann heißt $F_X \colon \mathbb{R}^n \to \mathbb{R}$ mit

$$\begin{aligned} F_X(x) := F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) &= P_X((-\infty, x_1] \times \ldots \times (-\infty, x_n]) \\ &= P(\{\omega \in \Omega \mid X_1(\omega) \leq x_1, \ldots, X_n(\omega) \leq x_n\}) \\ &= P(X_1 \leq x_1, \ldots, X_n \leq x_n) \end{aligned}$$

die (n-dimensionale) Verteilungsfunktion von X. Das mit X korrespondierende Bildmaß P_X von P heißt Verteilung von X.

Angesichts Satz 11.12 lässt die i-te Randverteilung von F folgende Interpretation zu:

$$F_i(y) = P((X_1, \ldots, X_n) \in \mathbb{R}^{i-1} \times (-\infty, y] \times \mathbb{R}^{n-i}) = P(X_i \leq y),$$

d.h. F_i ist die Verteilungsfunktion der Zufallsvariablen X_i. Falls F_i eine Dichte f_i besitzt, nennt man f_i i-te Randverteilungsdichte.

11.14 Beispiel (Dreifacher Münzwurf):
Wir betrachten die Ereignisse „Zahl" (Z) und „Wappen" (W) und setzen

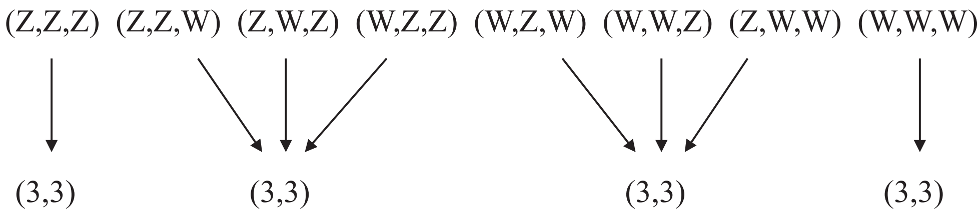
$$\Omega := \{(Z,Z,Z), (Z,Z,W), (Z,W,Z), (W,Z,Z), (W,Z,W), (W,W,Z), (Z,W,W), (W,W,W)\}.$$

Außerdem definieren wir:

$X_1 \colon \Omega \to \Omega_1$, $X_1 :=$ Anzahl von „Zahl" (Z) bei 3 Würfen, $\Omega_1' := \{0, 1, 2, 3\}$, $X_2 \colon \Omega \to \Omega_2$,

$X_2 :=$ Absolutbetrag der Differenz zwischen der Anzahl von „Zahl" (Z) und der Anzahl von „Wappen" (W) bei 3 Würfen, $\Omega_2' := \{1, 3\}$.

Wir wollen das gemeinsame Verhalten von X_1 und X_2 studieren, d.h. wir sind an der (diskreten) Zufallsvariablen (X_1, X_2) in $\Omega_1' \times \Omega_2'$ und ihrer Verteilung $P_{(X_1,X_2)}$ interessiert.



Damit erhalten wir für $p_{ij} := P((X_1, X_2) = (i, j))$, $i = 1, 2, 3$, $j = 1, 3$ die Werte:

	X_1				
X_2	0	1	2	3	P_{X_2}
1	0	$\frac{3}{8}$	$\frac{3}{8}$	0	$\frac{5}{8}$
3	$\frac{1}{8}$	0	0	$\frac{1}{8}$	$\frac{2}{8}$
P_{X_1}	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	

P_{X_1} und P_{X_2} bezeichnen die zugehörigen Randverteilungen.

11.2 Unabhängigkeit von Zufallsvariablen

Nach Definition 11.7 kann in eindeutiger Weise von der gemeinsamen Verteilung mehrerer Zufallsvariablen auf die Randverteilungen geschlossen werden. Wie in Beispiel 11.8 gezeigt wurde, gilt die Umkehrung jedoch nicht immer. In diesem Abschnitt wird u.a. gezeigt, dass die Umkehrung genau dann gilt, wenn die Zufallsvariablen unabhängig sind.

11.15 Definition (stochastisch unabhängig):
Es bezeichne $(\Omega, \mathfrak{F}, P)$ einen Wahrscheinlichkeitsraum und es sei I eine nichtleere Indexmenge. Die Familie der Mengensysteme $(\mathfrak{M}_i)_{i \in I}$ mit $\mathfrak{M}_i \subseteq \mathfrak{F}$ heißt (stochastisch) unabhängig bzgl. P genau dann, wenn für jede nichtleere, endliche Teilmenge $J \subseteq I$ gilt:

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j) \quad \forall\ A_j \in \mathfrak{M}_j \text{ und } j \in J.$$

11.16 Definition (erzeugte σ-Algebra):
Es seien $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum, $(\Omega', \mathfrak{F}')$ ein Messraum und $X : \Omega \to \Omega'$ eine $\mathfrak{F} - \mathfrak{F}'$-messbare Abbildung. Unter der von X erzeugten σ-Algebra versteht man die σ-Algebra $\sigma(X) := X^{-1}(\mathfrak{F}')$.

11.17 Definition (Unabhängigkeit von Zufallsgrößen):
Die Zufallsvariablen $X_1, \ldots, X_n$ heißen stochastisch unabhängig, wenn die von ihnen erzeugten σ–Algebren unabhängig sind.

11.18 Satz:
Es seien $X_1, \ldots, X_n$ Zufallsvariablen über $(\Omega, \mathfrak{F}, P)$ mit Werten in Messräumen $(\chi_i, \mathfrak{A}_i)$, $i = 1, \ldots, n$. Ferner seien $E_1, \ldots, E_n$ durchschnittsstabile Erzeugendensysteme der σ–Algebren $\mathfrak{A}_1, \ldots, \mathfrak{A}_n$ mit $\chi_i \in E_i$, $i = 1, \ldots, n$. Dann gilt:
Die Zufallsvariablen $X_1, \ldots, X_n$ sind genau dann stochastisch unabhängig, wenn die Mengensysteme $\widehat{E}_1, \ldots, \widehat{E}_n$ mit

$$\widehat{E}_i := \{X_i^{-1}(A_i) \mid A_i \in E_i\}, \qquad i = 1, \ldots, n,$$

unabhängig sind.

Beweis:
Da die Abbildung X_i messbar ist, liegt das Mengensystem $\widehat{E}_i$ in $\sigma(X_i)$ und bildet somit ein Erzeugendensystem von $\sigma(X_i)$. Es soll gezeigt werden:

Mengensysteme $\widehat{E}_1, \dots, \widehat{E}_n$ unabhängig $\Leftrightarrow$ Mengensysteme $\sigma(\widehat{E}_1), \dots, \sigma(\widehat{E}_n)$ unabhängig.

Dazu wird zunächst gezeigt:

Mengensysteme $\widehat{E}_1, \dots, \widehat{E}_n$ unabhängig $\Leftrightarrow$ Mengensysteme $\sigma(\widehat{E}_1), \widehat{E}_2 \dots, \widehat{E}_n)$ unabhängig.

Dafür wird

$$D_1 := \{A \in \sigma(\widehat{E}_1) \mid \{A\}, \widehat{E}_2, \dots, \widehat{E}_n \text{ unabhängig } \}$$

betrachtet und gezeigt, dass D_1 ein Dynkin–System ist. Damit gilt dann mit $\widehat{E}_1 \in D_1$ auch $\mathfrak{D}(\widehat{E}_1) \subseteq D_1$, wobei $\mathfrak{D}(\widehat{E}_1)$ das von $\widehat{E}_1$ erzeugte Dynkin–System ist.
Da $\widehat{E}_1$ als durchschnittsstabil vorausgesetzt ist, folgt mit Hilfe von Satz 5.12, dass $\mathfrak{D}(\widehat{E}_1) = \sigma(\widehat{E}_1)$ und damit $\sigma(\widehat{E}_1) \subseteq D_1$ ist. Damit ist gezeigt, dass die Mengensysteme $\sigma(\widehat{E}_1)$, $\widehat{E}_2$, …, $\widehat{E}_n$ unabhängig sind. Analog kann geschlossen werden, dass $\widehat{E}_1, \dots, \widehat{E}_{i-1}, \sigma(\widehat{E}_i), \widehat{E}_{i+1}, \dots, \widehat{E}_n$ für $i = 2, \dots, n$ und damit auch $\sigma(\widehat{E}_1), \sigma(\widehat{E}_2), \dots, \sigma(\widehat{E}_n)$ unabhängig sind.
Um zu zeigen, dass D_1 ein Dynkin–System ist, müssen folgende Punkte nachgewiesen werden:

(i) $\Omega \in D_1$,

(ii) für $A, B \in D_1$ mit $A \subseteq B$ gilt $B \setminus A \in D_1$,

(iii) für jede Folge $(A_n)_{n\in\mathbb{N}}$ paarweise disjunkter Mengen aus D_1 ist auch $\bigcup\limits_{n\in\mathbb{N}} A_n \in D_1$.

Zu (i): Da $\chi_i \in E_1$ ist, ist $\Omega \in \widehat{E}_1$. Für beliebige $A_i \in \widehat{E}_i$, $i = 2, \dots, n$, gilt dann:

$$P\left(\Omega \cap \bigcap_{i=2}^{n} A_i\right) = P(\Omega) \cdot P\left(\bigcap_{i=2}^{n} A_i\right) = P(\Omega) \cdot \prod_{i=2}^{n} P(A_i),$$

d.h. $\Omega \in D_1$.

Zu (ii): Für Mengen $A, B \in D_1$ mit $A \subset B$

$$P\left(A \cap \bigcap_{i=2}^{n} A_i\right) = P(A) \cdot \prod_{i=2}^{n} P(A_i)$$

und

$$P\left(B \cap \bigcap_{i=2}^{n} A_i\right) = P(B) \cdot \prod_{i=2}^{n} P(A_i).$$

Daraus ergibt sich

$$\begin{aligned}
P\left((B \setminus A) \cap \bigcap_{i=2}^{n} A_i\right) &= P\left(B \cap \bigcap_{i=2}^{n} A_i \setminus A \cap \bigcap_{i=2}^{n} A_i\right) \\
&= P\left(B \cap \bigcap_{i=2}^{n} A_i\right) - P\left(A \cap \bigcap_{i=2}^{n} A_i\right) \\
&= P(B) \cdot \prod_{i=2}^{n} P(A_i) - P(A) \cdot \prod_{i=2}^{n} P(A_i) \\
&= (P(B) - P(A)) \cdot \prod_{i=2}^{n} P(A_i) = P(B \setminus A) \cdot \prod_{i=2}^{n} (A_i).
\end{aligned}$$

Folglich ist auch $B \setminus A \in D_1$.

Zu (iii): Für eine Folge $(B_n)_{n\in\mathbb{N}}$ paarweise fremder Mengen aus D_1 gilt weiter

$$\begin{aligned} P\left(\left(\bigcup_{m\in\mathbb{N}} B_m\right) \cap A_2 \cap \ldots \cap A_n\right) &= P\left(\bigcup_{m\in\mathbb{N}} (B_m \cap A_2 \cap \ldots \cap A_n)\right) \\ &= \sum_{m\in\mathbb{N}} P(B_m \cap A_2 \cap \ldots \cap A_n) \\ &= \sum_{m\in\mathbb{N}} \left(P(B_m) \cdot \prod_{i=2}^{n} P(A_i)\right) \\ &= \left(\sum_{m\in\mathbb{N}} P(B_m)\right) \cdot \prod_{i=2}^{n} P(A_i) \\ &= P\left(\bigcup_{m\in\mathbb{N}} B_m\right) \cdot \prod_{i=2}^{n} P(A_i). \end{aligned}$$

Folglich gilt $\bigcup_{m\in\mathbb{N}} B_m \in D_1$. ■

11.19 Satz:
Es seien $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $(\Omega_i, \mathfrak{F}_i), i = 1, \ldots, n$, Messräume. Die Abbilundgen $X_i \colon \Omega \to \Omega_i$ sei $\mathfrak{F} - \mathfrak{F}_i$-messbar. $Y : \Omega \to \bigtimes_{i=1}^{n} \Omega_i$ sei definiert durch $Y(\omega) := (X_1(\omega), \ldots, X_n(\omega))\ \forall\ \omega \in \Omega$. Dann gilt: Die Zufallsvariablen $X_1, \ldots X_n$ sind genau dann stochastisch unabhängig, wenn

$$P_Y(A_1 \times \ldots \times A_n) = \prod_{i=1}^{n} P_{X_i}(A_i) \qquad \forall\ A_i \in \mathfrak{F}_i, i = 1, \ldots, n \tag{11.3}$$

gilt.

Beweis:
(i) Es wird die Gültigkeit von (11.3) angenommen. Dann gilt für alle $A_i \in \mathfrak{F}_i$:

$$\begin{aligned} P_Y(A_1 \times \ldots \times A_n) &= P(Y^{-1}(A_1 \times \ldots \times A_n)) \\ &= P\left(\bigcap_{i=1}^{n} X_i^{-1}(A_i)\right). \end{aligned}$$

Da aber $\{X_i^{-1}(A_i) \mid A_i \in \mathfrak{F}_i\}$ die von X_i erzeugte σ-Algebra darstellt, entspricht die Beziehung

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

der Unabhängigkeit der Zufallsvariablen $X_1, \ldots, X_n$.

(ii) Sind die Zufallsvariablen $X_1, \ldots, X_n$ stochastisch unabhängig, so gilt aufgrund der Definition 11.17 zunächst

$$P\left(\bigcap_{i=1}^{n} X_i^{-1}(A_i)\right) = \prod_{i=1}^{n} P(X_i^{-1}(A_i)) \;\; \forall\, A_i \in \mathfrak{F}_i, i = 1, \ldots, n.$$

Mit

$$P\left(\bigcap_{i=1}^{n} X_i^{-1}(A_i)\right) = P(Y^{-1}(A_1 \times \ldots \times A_n)) = P_Y(A_1 \times \ldots \times A_n)$$

und

$$\prod_{i=1}^{n} P(X_i^{-1}(A_i)) = \prod_{i=1}^{n} P_{X_i}(A_i)$$

folgt

$$P_Y(A_1 \times \ldots \times A_n) = \prod_{i=1}^{n} P_{X_i}(A_i) \;\; \forall\, A_i \in \mathfrak{F}_i, i = 1, \ldots, n$$

in Übereinstimmung mit 11.3.

■

11.20 Satz:
Setzt man in Satz 11.19 $(\Omega_i, \mathfrak{F}_i) := (\mathbb{R}, \mathfrak{B})$ *für* $i = 1, \ldots, n$*, dann ist die Aussage aus Definition 11.3 äquivalent zu*

$$F_X(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) \qquad \forall\, x = (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

wobei F_X *und* F_{X_i} *die Verteilungsfunktionen der Zufallsvariablen* X *und* X_i *bezeichnen.*

Beweis:
Übung.

■

11.21 Satz:

a) Es seien $X_1, \ldots, X_n$ *Zufallsvariablen über* $(\Omega, \mathfrak{F}, P)$ *mit Werten in* $(\mathbb{R}, \mathfrak{B})$*. Sind die Zufallsvariablen* $X_1, \ldots, X_n$ *stochastisch unabhängig und stetig verteilt mit den Dichten* $f_{X_i}, i = 1, \ldots, n$*, so besitzt auch* $X := (X_1, \ldots, X_n)$ *eine Dichte und es gilt*

$$f_X(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \;\forall\, x = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

b) Ist umgekehrt $X := (X_1, \ldots, X_n)$ *stetig verteilt mit einer Dichte* f_X *von der Form*

$$f_X(x_1, \ldots, x_n) = \prod_{i=1}^{n} g_i(x_i) \;\; \forall\, x := (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

wobei $g_i(x_i) \geq 0$ *und* $\int_{\mathbb{R}} g_i(x_i)\, dx_i = 1$ *für* $i = 1, \ldots, n$ *gilt, dann sind die Zufallsvariablen* $X_1, \ldots, X_n$ *stochastisch unabhängig und* g_i *ist eine Dichte von* X_i, $i = 1, \ldots, n$.

Beweis:

a) Satz 11.20 impliziert

$$F_X(x_1,\ldots,x_n) = \prod_{i=1}^n F_{X_i}(x_i) = \prod_{i=1}^n \int_{-\infty}^{x_i} f_{X_i}(y_i)dy_i$$
$$= \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{X_1}(y_1)\cdot \ldots \cdot f_{X_n}(y_n)dy_1\ldots dy_n$$
$$= \int_{\times_{i=1}^n(-\infty,x_i]} f_X(y)\,dy\,.$$

Da die unter dem Integral stehende Funktion das Produkt von Dichten ist, ist sie somit Dichte von $X := (X_1,\ldots,X_n)$.

b) Da die Funktionen g_i Dichten sind, sind die Funktionen $G_i(x_i) = \int_{-\infty}^{x_i} g_i(y_i)dy_i, i = 1,\ldots,n$, stetige Verteilungsfunktionen und es gilt:

$$F_X(x_1,\ldots,x_n) = \int_{\times_{i=1}^n(-\infty,x_i]} f_X(y)\,dy$$
$$= \int_{\times_{i=1}^n(-\infty,x_i]} \prod_{i=1}^n g_i(y_i)\,dy$$
$$= \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} g_1(y_1)\cdot \ldots \cdot g_n(y_n)\,dy_1\cdots dy_n$$
$$= \prod_{i=1}^n \int_{-\infty}^{x_i} g_i(y_i)\,dy_i$$
$$= \prod_{i=1}^n G_i(x_i)\,.$$

Andererseits ist

$$F_{X_i}(x_i) = \lim_{\substack{x_k\to\infty\\ k\neq i}} F_X(x_1,\ldots,x_n) = G_i(x_i)\,.$$

Also gilt

$$F_X(x_1,\ldots,x_n) = \prod_{i=1}^n F_{X_i}(x_i)\,,$$

weshalb die Zufallsvariablen stochastisch unabhängig sind (vergleiche Satz 11.20). ■

11.22 Beispiel (Fortführung Beispiel 11.11):
(X,Y) sei eine $\mathbb{R}^2$-wertige Zufallsvariable mit Dichte

$$f_{(X,Y)}(x,y) := \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}}\exp\left(-\frac{1}{2(1-\varrho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\varrho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right),\ (x,y)\in\mathbb{R}^2,$$

wobei $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 \in \mathbb{R}^+$ und $\varrho \in (-1, 1)$ (Dichte der zweidimensionalen Normalverteilung). Wir zeigen, dass X und Y stochastisch unabhängig sind, wenn $\varrho = 0$ gilt.
Für die Randverteilungsdichte gilt aufgrund von Beispiel 11.12

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}\right), \; x \in \mathbb{R},$$

und entsprechend

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\frac{(y-\mu_1)^2}{\sigma_2^2}\right), \; y \in \mathbb{R}.$$

Zu zeigen ist deshalb die Gleichung

$$\begin{aligned} f_{(X,Y)}(x,y) &= \frac{1}{2\pi\sigma_1\sigma_2 \underbrace{\sqrt{1-\varrho^2}}_{\stackrel{!}{=}1}} \cdot \\ &\quad \exp\Bigg(-\frac{1}{2\underbrace{\sqrt{1-\varrho^2}}_{\stackrel{!}{=}1}}\Bigg(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\underbrace{\varrho}_{\stackrel{!}{=}0}\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\Bigg)\Bigg) \\ &= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{1}{2}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right) \;=\; f_X(x)\cdot f_Y(y), \end{aligned}$$

die offensichtlich aber nur für $\varrho = 0$ zu erfüllen ist.

11.23 Satz:
Es seien $(\Omega_i, \mathfrak{F}_i, P_i), i = 1, \ldots, n,$ Wahrscheinlichkeitsräume. Dann existiert genau ein Wahrscheinlichkeitsmaß P auf $\otimes_{i=1}^n \mathfrak{F}_i$ mit

$$P(A_1 \times \ldots \times A_n) = \prod_{i=1}^n P_i(A_i) \quad \forall\ A_i \in \mathfrak{F}_i, i = 1, \ldots, n.$$

Beweis:
Zunächst wird der Fall $n := 2$ betrachtet. Die Menge der kartesischen Produkte

$$K := \{A_1 \times A_2 \mid A_1 \in \mathfrak{F}_1, A_2 \in \mathfrak{F}_2\}$$

bildet einen Semiring, wobei der Beweis analog zum Fall $\mathbb{I}^2$ (vgl. Kapitel 5.1) geführt werden kann (Übung!). Betrachte

$$\tilde{P}(A_1 \times A_2) := P_1(A_1) \cdot P_2(A_2) \;\; \forall\ A_1 \times A_2 \in K.$$

Wenn $\tilde{P}$ ein Prämaß und σ–endlich ist, so kann mithilfe der Maßfortsetzungssätze 6.9 und 6.18 die Existenz einer eindeutigen Erweiterung von $\tilde{P}$ zu einem Maß P auf der von K erzeugten σ–Algebra gefolgert werden.
Der allgemeine Fall ergibt sich durch vollständige Induktion nach n mit dem Schritt

$$P^{(n+1)}(A_1 \times \ldots \times A_{n+1}) = P^{(n)}(A_1 \times \ldots \times A_n) \cdot P_{n+1}(A_{n+1}).$$

Zu zeigen sind:

(i) $\tilde{P}$ ist nichtnegativ.

(ii) $\tilde{P}$ ist σ–additiv.

(iii) $\tilde{P}$ ist σ–endlich.

Zu (i): (Nichtnegativität)
Es ist $P(A_1 \times A_2) \geq 0 \ \ \forall \ A_1 \times A_2 \in K$ und außerdem $P(\varnothing) = 0$, da für $A_1 \times A_2 = \varnothing$ entweder $A_1 = \varnothing$ oder $A_2 = \varnothing$ ist (vgl. Satz 11.2).

Zu (ii): (σ–Additivität)
Betrachte eine Folge $(A_{1,m} \times A_{2,m})_{m\in\mathbb{N}}$ paarweise fremder Mengen $A_{1,m} \times A_{2,m} \in K, m \in \mathbb{N}$, mit $\bigcup\limits_{m=1}^{\infty} A_{1,m} \times A_{2,m} = A_1 \times A_2 \in K$. Es gilt:

$$\begin{aligned} I_{A_1}(\omega_1) \cdot I_{A_2}(\omega_2) = I_{A_1 \times A_2}(\omega_1, \omega_2) &= \sum_{m=1}^{\infty} I_{A_{1,m} \times A_{2,m}}(\omega_1, \omega_2) \\ &= \sum_{m=1}^{\infty} I_{A_{1,m}}(\omega_1) \cdot I_{A_{2,m}}(\omega_2) \ \ \forall \ \omega_1 \in \Omega_1, \omega_2 \in \Omega_2. \end{aligned}$$

Für jedes $\omega_1 \in \Omega_1$ werden durch $f\colon \Omega_2 \to \mathbb{R}$ mit $f(\omega_2) := I_{A_1}(\omega_1) \cdot I_{A_2}(\omega_2)$ und $f_k\colon \Omega_2 \to \mathbb{R}$ mit

$$f_k(\omega_2) := \sum_{m=1}^{k} I_{A_{1,m}}(\omega_1) \cdot I_{A_{2,m}}(\omega_2), k \in \mathbb{N},$$

nichtnegative $\mathfrak{F}_2$–$\mathfrak{B}$–messbare Elementarfunktionen definiert.
$(f_k)_{k\in\mathbb{N}}$ ist monoton nichtfallend mit $\lim\limits_{k\to\infty} f_k = f$. Mithilfe des Satzes von der monotonen Konvergenz (9.42) gilt:

$$\begin{aligned} g(\omega_1) &:= I_{A_1}(\omega_1) \cdot P_2(A_2) = \int f dP_2 = \lim_{k\to\infty} \int f_k dP_2 \\ &= \lim_{k\to\infty} g_k(\omega_1). \end{aligned}$$

Dabei ist $g_k(\omega_1) = \sum\limits_{m=1}^{k} I_{A_{1,m}}(\omega_1) \cdot P_2(A_{2,m}), k \in \mathbb{N}$.
Für $k \in \mathbb{N}$ sind g und g_k nichtnegative $\mathfrak{F}_1$–$\mathfrak{B}$–messbare Elementarfunktionen. $(g_k)_{k\in\mathbb{N}}$ ist monoton nichtfallend mit $\lim\limits_{k\to\infty} g_k = g$. Wie oben wird gefolgert:

$$\begin{aligned} \tilde{P}(A_1 \times A_2) &= P_1(A_1) \cdot P_2(A_2) = \int g dP_1 = \lim_{k\to\infty} \int g_k dP_1 \\ &= \lim_{k\to\infty} \sum_{m=1}^{k} P_1(A_{1,m}) \cdot P_2(A_{2,m}) = \sum_{m=1}^{\infty} P(A_{1,m} \times A_{2,m}). \end{aligned}$$

Zu (iii): (σ–Endlichkeit)
Seien $(B_{1,m})_{m\in\mathbb{N}}$ und $(B_{2,m})_{m\in\mathbb{N}}$ (disjunkte) Zerlegungen von Ω_1 bzw. Ω_2, so gilt für die Mengen $B_{1,m} \times B_{2,n}, m, n \in \mathbb{N}$

$$\bigcup_{m=1}^{\infty} \bigcup_{n=1}^{\infty} (B_{1,m} \times B_{2,n}) = \bigcup_{m=1}^{\infty} (B_{1,m} \times \Omega_2) = \Omega_1 \times \Omega_2$$

und

$$\tilde{P}(B_{1,m} \times B_{2,n}) = P_1(B_{1,m}) \cdot P_2(B_{2,n}) < \infty.$$

■

Satz 11.23 gilt in entsprechender Form auch für Maßräume und σ–endliche Maße (siehe Bauer, Kapitel 8).

11.24 Satz:
Es seien $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $(\Omega_i, \mathfrak{F}_i), i = 1, \ldots, n$, Messräume. Sind die Zufallsvariablen $X_i \colon \Omega \to \Omega_i$ stochastisch unabhängig, so sind auch die Zufallsvektoren $(X_1, \ldots, X_{r_1})$, $(X_{r_1+1}, \ldots, X_{r_2}), \ldots, (X_{r_{m-1}+1}, \ldots, X_{r_m})$, $1 \leq r_1 < r_2 < \ldots < r_m \leq n$, stochastisch unabhängig.

Beweis:
O.B.d.A. genügt es zu zeigen, dass $Y := (X_1, \ldots, X_r)$ und $Z := (X_{r+1}, \ldots, X_s), 1 \leq r < s \leq n$ stochastisch unabhängig sind. Aufgrund der Voraussetzung und aufgrund von Definition 11.17 sind die Zufallsvariablen $X_1, \ldots, X_r$ sowie $X_{r+1}, \ldots, X_s$ stochastisch unabhängig. Es wird Satz 11.19 angewandt mit dem sich für alle $A := C_1 \times \ldots \times C_r$ und $B := C_{r+1} \times \ldots \times C_s$ mit $C_i \in \mathfrak{F}_i, i = 1, 2, \ldots, s$ folgendes ergibt:

$$P_Y(A) = \prod_{i=1}^{r} P_{X_i}(C_i) \quad \text{und} \quad P_Z(B) = \prod_{i=r+1}^{s} P_{X_i}(C_i),$$

woraus

$$\begin{aligned} P_Y(A) \cdot P_Z(B) &= \prod_{i=1}^{r} P_{X_i}(C_i) \cdot \prod_{i=r+1}^{s} P_{X_i}(C_i) \\ &= P_{(X_1,\ldots,X_n)}(C_1 \times \ldots \times C_s \times \Omega_{s+1} \times \ldots \times \Omega_n) \\ &= P_{(X_1,\ldots,X_s)}(C_1 \times \ldots \times C_s) = P_{(Y,Z)}(A \times B) \end{aligned}$$

folgt, weshalb $Y^{-1}(E_1)$ und $Z^{-1}(E_2)$ mit $E_1 := \{A = C_1 \times \ldots \times C_r \mid C_i \in \mathfrak{F}_i, i = 1, \ldots, r\}$ und $E_2 := \{B = C_{r+1} \times \ldots \times C_s \mid C_i \in \mathfrak{F}, i = r+1, \ldots, s\}$ stochastisch unabhängig sind. Da E_1 und E_2 durchschnittsstabil sind, folgt die Behauptung mit Satz 11.18. ■

11.25 Satz:
Es sei $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $(\Omega_i, \mathfrak{F}_i)$ sowie $(\Omega_i', \mathfrak{F}_i'), i = 1, \ldots, n$, seien Messräume. Die Abbildung $X_i \colon \Omega \to \Omega_i$ sei $\mathfrak{F} - \mathfrak{F}_i$-messbar und $g_i : \Omega_i \to \Omega_i'$ sei $\mathfrak{F}_i - \mathfrak{F}'{}_i$-messbar. In diesem Fall folgt aus der Unabhängigkeit der Zufallsvariablen $X_1, \ldots, X_n$ die Unabhängigkeit der Zufallsvariablen $g(X_1), \ldots, g(X_n)$.

Beweis:
Aufgrund der Voraussetzungen sind die σ-Algebren $X_i^{-1}(\mathfrak{F}_i), i = 1, \ldots, n$, unabhängig. Wegen der Messbarkeit der g_i gilt $g_i^{-1}(\mathfrak{F}_i') \subseteq \mathfrak{F}_i$ und weiter

$$X_i^{-1}(g_i^{-1}(\mathfrak{F}_i')) = (g_i \circ X_i)^{-1}(\mathfrak{F}_i') \subseteq X_i^{-1}(\mathfrak{F}_i), \qquad i = 1, \ldots, n,$$

weswegen auch die σ-Algebren $(g_i \circ X_i)^{-1}(\mathfrak{F}_i')$ unabhängig sind. ■

Führt man die Sätze 11.23 und 11.24 zusammen, erhält man den folgenden Satz:

11.26 Satz:
Es sei $(\Omega, \mathfrak{F}, P)$ *ein Wahrscheinlichkeitsraum und* $(\Omega_i, \mathfrak{F}_i), 1 \leq i \leq n$, *sowie* $(\chi_j, \mathfrak{A}_j), 1 \leq j \leq m$, *seien Messräume.* $X_i \colon \Omega \to \Omega_i$ *sei* $\mathfrak{F} - \mathfrak{F}_i$*-messbar und* $g_j \colon \bigotimes_{i=r_{j-1}+1}^{r_j} \Omega_i \to \bigotimes_{i=r_{j-1}+1}^{r_j} \chi_i$ *sei* $\mathfrak{F}_i - \mathfrak{A}_i$*−messbar für* $1 \leq j \leq m$. *Dann gilt: Sind* $X_i, \ldots, X_n$ *stochastisch unabhängig, dann sind auch* $g_1(X_1, \ldots, X_{r_1})$, $g_2(X_{r_1+1}, \ldots X_{r_2}), \ldots, g_m(X_{r_{m-1}+1}, \ldots, X_{r_m})$ *stochastisch unabhängig* $(1 \leq r_1 < r_2 < \ldots < r_m \leq n)$.

11.27 Definition (Produktverteilung):
Es seien $P_1, \ldots, P_n$, $n \in \mathbb{N}$, *Wahrscheinlichkeitsmaße auf* $\mathfrak{B}^1$, *der* σ*–Algebra der Borelschen Mengen über* $\mathbb{R}$. *Dann heißt das gemäß Satz 11.23 eindeutig bestimmte Wahrscheinlichkeitsmaß* P *auf* $\mathfrak{B}^n$ *das* *Produktmaß* *der Wahrscheinlichkeitsmaße* $P_1, \ldots, P_n$.

$$\textit{Schreibweise:} \quad P := \bigotimes_{i=1}^{n} P_i = P_1 \otimes \ldots \otimes P_n.$$

Es seien $P, P_1, \ldots, P_n$ wie in Definition 11.27 gewählt und $F, F_1, \ldots, F_n$ die zugehörigen Verteilungsfunktionen, dann gilt:

$$P\left(\bigtimes_{i=1}^{n} (-\infty, x_i]\right) = \prod_{i=1}^{n} P_i((-\infty, x_i]) = \prod_{i=1}^{n} F_i(x_i)\,,$$

d.h.

$$F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_i(x_i), \qquad (x_1, \ldots, x_n) \in \mathbb{R}^n\,.$$

11.3 Transformation von Zufallsvariablen (von stetigen Verteilungen)

Zunächst soll folgendes einführendes Beispiel betrachtet werden:

11.28 Beispiel (Transformation einer Dichte, I):
Es seien $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $X \colon \Omega \to \mathbb{R}$ eine Zufallsvariable mit Dichte

$$f_X(x) := \begin{cases} x, & 0 \leq x \leq 1 \\ 2 - x, & 1 \leq x \leq 2 \\ 0, & \text{sonst.} \end{cases}$$

Es soll die Dichte der Zufallsvariablen $Y := -2X + 1$ berechnet werden.
1. Schritt: Bestimmung der Verteilungsfunktion von X:

$$F_X(x) = \int\limits_{-\infty}^{x} f_X(t)\, dt = \begin{cases} 0, & x < 0 \\ \frac{1}{2}x^2, & 0 \leq x \leq 1 \\ -\frac{1}{2}x^2 + 2x - 1, & 1 \leq x \leq 2 \\ 1, & x > 2. \end{cases}$$

2. Schritt: Bestimmung der Verteilungsfunktion von Y:

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) = P(-2X+1 \leq y) = P\left(X \geq \frac{1}{2}(1-y)\right) \\
&= 1 - P\left(X < \frac{1}{2}(1-y)\right) = 1 - F_X\left(\frac{1}{2}(1-y) - 0\right) \\
&= 1 - F_X\left(\frac{1}{2}(1-y)\right) = \begin{cases} 0, & y < -3 \\ \frac{1}{8}(y+3)^2, & -3 \leq y \leq -1 \\ \frac{1}{8}(-y^2+2y+7), & -1 \leq y \leq 1 \\ 1, & y > 1. \end{cases}
\end{aligned}$$

3. Schritt: Bestimmung der Dichte von Y:
Aus $F_y(y)$ folgt durch Differentiation

$$f_Y(y) = \begin{cases} \frac{1}{4}(y+3), & -3 \leq y \leq -1 \\ -\frac{1}{4}y + \frac{1}{4} = \frac{1}{4}(1-y), & -1 \leq y \leq 1 \\ 0, & \text{sonst.} \end{cases}$$

Transformationen von Zufallsvariablen kommen nicht nur innermathematisch sondern auch in vielen praktischen Anwendungen vor.

11.29 Satz (Transformationssatz für Dichten):
Es seien $X := (X_1, \ldots, X_m)$, $m \in \mathbb{N}$, ein m–dimensionaler Zufallsvektor mit der Dichte f_X und $G : \mathbb{R}^m \to \mathbb{R}^m$,

$$G\colon (x_1, \ldots, x_m) \to (G_1(x_1, \ldots, x_m), \ldots, G_m(x_1, \ldots, x_m))$$

eine Abbildung mit folgenden Eigenschaften:

(i) G ist injektiv und stetig differenzierbar auf $\mathbb{R}^m$,

(ii) die Funktionaldeterminante

$$\det\left(\frac{\partial G_i}{\partial x_j}\right)_{1 \leq i,j \leq m} = \begin{vmatrix} \dfrac{\partial G_1(x_1, \ldots, x_m)}{\partial x_1} & \cdots & \dfrac{\partial G_1(x_1, \ldots, x_m)}{\partial x_m} \\ \vdots & & \vdots \\ \dfrac{\partial G_m(x_1, \ldots, x_m)}{\partial x_1} & \cdots & \dfrac{\partial G_m(x_1, \ldots, x_m)}{\partial x_m} \end{vmatrix}$$

ist entweder positiv oder negativ auf $\mathbb{R}^m$.
Dann gilt:

$$f_Y(y_1, \ldots, y_m) = \begin{cases} \dfrac{f_X(G^{-1}((y_1, \ldots, y_m)))}{\left|\det\left(\frac{\partial G_i}{\partial x_j}(G^{-1}(y_1, \ldots, y_m))\right)\right|} & \textit{für } (y_1, \ldots, y_m) \in G(\mathbb{R}^m) \\ 0 & \textit{sonst} \end{cases}$$

ist die Dichte des Zufallsvektors $Y := G(X)$.

Beweis: Heuser, ,,Substitutionsregel für das Riemann-Integral”, Kapitel 205.2.

11.30 Beispiel (Transformation einer Dichte, II):
Die Zufallsgröße X habe die Dichte

$$f_X(x) := \begin{cases} x, & 0 \le x \le 1 \\ 2-x, & 1 \le x \le 2 \\ 0, & \text{sonst.} \end{cases}$$

Zu berechnen ist die Dichte von $Y := -2X + 1$ (vgl. Beispiel 11.28). Hier ist $m = 1$ und $G\colon \mathbb{R} \to \mathbb{R}$ mit $G(X) := -2X + 1$. G ist offensichtlich injektiv, messbar und stetig differenzierbar mit $G'(x) = -2 < 0 \ \forall\ x$.

$$G(x) = -2x + 1 := y \Longleftrightarrow x = \frac{1}{2}(1-y) = G^{-1}(y), \qquad \left|G'(G^{-1}(y))\right| = |-2| = 2.$$

Folglich gilt nach obigem Satz:

$$\begin{aligned} f_Y(y) &= \frac{f_X(G^{-1}(y))}{|-2|} = \frac{1}{2} f_X\left(\frac{1}{2}(1-y)\right) = \begin{cases} \frac{1}{2}\cdot\frac{1}{2}(1-y), & 0 \le \frac{1}{2}(1-y) \le 1 \\ \frac{1}{2}(2-\frac{1}{2}(1-y)), & 1 \le \frac{1}{2}(1-y) \le 2 \\ 0, & \text{sonst} \end{cases} \\ &= \begin{cases} \frac{1}{4}(1-y), & -1 \le y \le 1 \\ \frac{1}{4}(y+3), & -3 \le y \le -1 \\ 0, & \text{sonst.} \end{cases} \end{aligned}$$

11.31 Satz (Summe stetig verteilter Zufallsvektoren, Faltungssatz):
Es sei $X := (X_1, X_2)$ ein stetig verteilter Zufallsvektor mit Dichte f_X. Dann besitzt die Zufallsvariable $Y := X_1 + X_2$ eine Dichte der Form

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(x, y-x)\ dx = \int_{-\infty}^{\infty} f_X(x-y, y)\ dy, \quad y \in \mathbb{R}.$$

Sind speziell X_1, X_2 unabhängig, so ist

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(x) \cdot f_{X_2}(y-x)\ dx, \quad y \in \mathbb{R}.$$

Beweis:
Es wird $G(x_1, x_2) := (x_1, x_1 + x_2)$, $x_1, x_2 \in \mathbb{R}$ gesetzt. G ist injektiv, messbar und auf $\mathbb{R}^2$ stetig differenzierbar mit

$$\det\begin{pmatrix} \frac{\partial G_1}{\partial x_1} & \frac{\partial G_1}{\partial x_2} \\ \frac{\partial G_2}{\partial x_1} & \frac{\partial G_2}{\partial x_2} \end{pmatrix} = \det\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = 1.$$

$$\begin{aligned} & (y_1, y_2) = G(x_1, x_2) = (x_1, x_1 + x_2) \\ \iff & (x_1, x_2) = (y_1, y_2 - y_1) = G^{-1}(y_1, y_2), \qquad y_1, y_2 \in \mathbb{R}. \end{aligned}$$

Nach dem Transformationssatz hat der Zufallsvektor $Z := G(X) = (X_1, X_1 + X_2)$ eine Dichte der Form

$$f_Z(y_1, y_2) = f_X(y_1, y_2 - y_1), \qquad y_1, y_2 \in \mathbb{R}.$$

Die Verteilung von $X_1 + X_2$ ist nun die zweite Randverteilung von Z, die sich durch Integration von f_Z nach y_1 ergibt. Damit erhält man die erste Aussage des Satzes. Die zweite Aussage folgt mit Satz 11.21 a). ■

11.32 Beispiel:
Es seien X_1, X_2 beide $\mathcal{N}(0,1)$–verteilt, d.h.

$$f_{X_1}(x_1) := \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x_1^2}, \quad x_1 \in \mathbb{R},$$

$$f_{X_2}(x_2) := \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x_2^2}, \quad x_2 \in \mathbb{R}.$$

Sind X_1 und X_2 stochastisch unabhängig, so gilt für ihre gemeinsame Wahrscheinlichkeitsdichte f_X:

$$f_X(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2), \quad (x_1, x_2) \in \mathbb{R}^2.$$

$Y := X_1 + X_2$ hat demnach die Dichte

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_X(x, y - x)\, dx = \int_{-\infty}^{\infty} f_{X_1}(x) \cdot f_{X_2}(y - x)\, dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(y-x)^2}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2}} \cdot e^{-\frac{1}{2}\cdot\frac{y^2}{2}} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} \cdot e^{-\frac{1}{2}(\frac{y^2}{2} - 2xy + 2x^2)}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2}} \cdot e^{-\frac{1}{2}\cdot\frac{y^2}{2}} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} \cdot e^{-\frac{1}{2}\cdot 2(\frac{y^2}{4} - xy + x^2)}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2}} \cdot e^{-\frac{1}{2}\cdot\frac{y^2}{2}} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} \cdot e^{-\frac{1}{2}\cdot\frac{(x-y/2)^2}{1/2}}\, dx, \ y \in \mathbb{R}. \end{aligned}$$

Mit

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-(x-\frac{y}{2})^2}\, dx \stackrel{(*)}{=} \frac{1}{\sqrt{\pi}} \underbrace{\int_{-\infty}^{\infty} e^{-t^2}\, dt}_{\sqrt{\pi}} = 1$$

((∗) Substitution: $t = x - \frac{y}{2}$, $dx = dt$) folgt

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2}} \cdot e^{-\frac{1}{2}\frac{y^2}{2}},\ y \in \mathbb{R},$$

also ist Y $\mathcal{N}\big(0, \sqrt{2}\big)$–verteilt.

Anwendungen

11.33 Beispiel:
Zwei verschieden starke Bleche werden aufeinander geschweißt. Es mögen $X_1 \geq 0$ und $X_2 \geq 0$ die Dickenabweichungen der beiden Bleche vom Sollwert bezeichnen. Damit weicht die Gesamtdicke um (maximal) $X_1 + X_2$ vom Sollwert ab. Es seien X_1 und X_2 stochastisch unabhängig und $\mathcal{N}(0,1)$–verteilt. Dann ist $Y := X_1 + X_2$ nach Beispiel 11.32 $\mathcal{N}\big(0, \sqrt{2}\big)$ verteilt.

11.34 Beispiel (Wartezeit an einem Postschalter):
Die aufeinanderfolgenden Bedienzeiten seien stochastisch unabhängig und exponentiell verteilt mit dem Parameter $\lambda > 0$. Zum Zeitpunkt der Ankunft befinden sich genau n Kunden im System.

Wie ist die Wartezeit des gerade ankommenden Kunden verteilt?

Zunächst wird festgestellt, dass die Restbedienzeit des gerade im Bediener befindlichen Kunden aufgrund der Gedächtnislosigkeit der Exponentialverteilung ebenfalls exponentiell verteilt ist. Denn es gilt:

$$\begin{aligned} P(X_1 > t + x \mid X_1 > t) &= \frac{P(X_1 > t + x, X_1 > t)}{P(X_1 > t)} = \frac{P(X_1 > t + x)}{P(X_1 > t)} \\ &= \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x} = P(X_1 > x),\ t > 0,\ x > 0. \end{aligned}$$

Zur Bestimmung der Verteilung der Wartezeit eines ankommenden Kunden wird der folgende Satz betrachtet:

11.35 Satz:
Es seien $X_1, \ldots, X_n$ stochastisch unabhängige Zufallsvariablen mit $X_i \stackrel{d}{=} Exp(\lambda)$, $i = 1, \ldots, n$. Dann ist $Y = X_1 + \cdots + X_n$ Erlang(n, λ)–verteilt, d.h. es ist

$$f_Y(y) = \begin{cases} \dfrac{\lambda^n y^{n-1} e^{-\lambda y}}{(n-1)!}, & y \geq 0 \\ 0, & y < 0. \end{cases}$$

Beweis (mit vollständiger Induktion):
Für $n := 1$ ist die Behauptung evident. Es wird von $n-1$ auf n geschlossen:

$$\begin{aligned} f_n(y) &= \int\limits_0^y f_1(y-x) f_{n-1}(x)\, dx \\ &= \frac{1}{(n-2)!} \int\limits_0^y \lambda e^{-\lambda(y-x)} \lambda^{n-1} x^{n-2} e^{-\lambda x}\, dx \\ &= \frac{\lambda^n e^{-\lambda y}}{(n-2)!} \int\limits_0^y x^{n-2}\, dx = \frac{\lambda^n e^{-\lambda y} y^{n-1}}{(n-1)!},\ y \geq 0. \end{aligned}$$

■

Zusammenfassend ergibt sich für Beispiel 11.34 das Ergebnis, dass die Wartezeit (inklusive Bedienung) Erlang(n,λ)–verteilt ist, wenn bei der Ankunft n Kunden im System verweilen.

11.4 Der Satz von Fubini und seine Anwendungen

Nicht immer ist man bei der Transformation von Zufallsvariablen an der Verteilung von $G(x)$ interessiert. Häufig genügt die Kenntnis einer gewissen Kenngröße von $G(x)$ wie dem Erwartungswert oder der Varianz. Eine besondere Bedeutung haben dabei Produkte und Summen von Zufallsvariablen.
Im Falle diskreter und stochastisch unabhängiger Zufallsvariablen X, Y ergibt sich mit den Vereinbarungen $P_{ij} := P((X,Y) = (x_i, y_j))$ sowie $P_{i\cdot} := P(X = x_i)$ und $P_{\cdot j} := P(Y = y_j)$

$$\begin{aligned} \mathrm{E}[X \cdot Y] &= \sum_{i,j} x_i y_j P_{ij} = \sum_{i,j} x_i y_j \cdot P_{i\cdot} \cdot P_{\cdot j} \\ &= \sum_i x_i P_{i\cdot} \cdot \sum_j y_j P_{\cdot j} = \mathrm{E}[X] \cdot \mathrm{E}[Y]\,. \end{aligned}$$

Im Falle stetiger Zufallsvariablen lässt sich analog schließen:

$$\begin{aligned} \mathrm{E}[X \cdot Y] &= \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} x \cdot y \cdot f_X(x) \cdot f_Y(y)\, dx\, dy \\ &= \int\limits_{-\infty}^{+\infty} x \cdot f_X(x)\, dx \cdot \int\limits_{-\infty}^{+\infty} y \cdot f_Y(y)\, dy = \mathrm{E}[X] \cdot \mathrm{E}[Y]\,. \end{aligned}$$

Der allgemeine Fall kann mit Hilfe des Satzes von Fubini gelöst werden. Er beinhaltet hinreichende Bedingungen für die Vertauschbarkeit zweier Integrationen. Zunächst müssen dafür ein paar Vorbereitungen getroffen werden:

Es seien $(\Omega_i, \mathfrak{F}_i, \mu_i)$, $i = 1, 2$, zwei Maßräume, $\mu := \mu_1 \bigotimes \mu_2$ das Produktmaß von μ_1 und μ_2, und $X\colon \Omega_1 \times \Omega_2 \longrightarrow \overline{\mathbb{R}}$ eine $\mathfrak{F}_1 \bigotimes \mathfrak{F}_2$–$\overline{\mathfrak{B}}$–messbare numerische Funktion. In diesem Fall stellt sich die Frage, inwieweit man das Integral $\int X d(\mu_1 \bigotimes \mu_2)$ auch in der iterierten Form

$$\int \Big[X(\omega_1, \omega_2) d\mu_2 \Big] d\mu_1$$

darstellen kann. Damit dies überhaupt eine sinnvolle Vorgehensweise wird, muss $X(\omega_1, \omega_2)$ bei festgehaltenem ω_1 als Funktion von ω_2 $\mathfrak{F}_2$–$\overline{\mathfrak{B}}$–messbar und das μ–Integral $\int X(\omega_1, \omega_2) d\mu_2$, als Funktion von ω_1, $\mathfrak{F}_1$–$\overline{\mathfrak{B}}$–messbar sein. Die Beantwortung dieser Frage erfordert den Begriff der Schnittbildung bei Mengen und Funktionen.

11.36 Definition (Schnitt von Mengen):
Ist $A \subseteq \Omega_1 \times \Omega_2$ eine beliebige Menge, so nennt man $A_{\omega_1} := \{\omega_2 \in \Omega_2 \mid (\omega_1, \omega_2) \in A\}$, $\omega_1 \in \Omega_1$, den Schnitt von A an der Stelle ω_1. Entsprechend wird A_{ω_2} definiert.

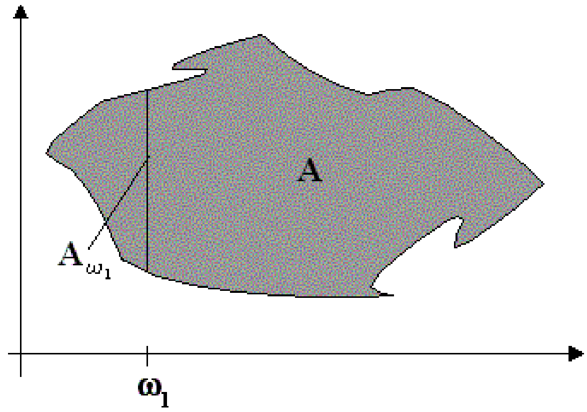


Abbildung 11.7: Veranschaulichung der Definition des Schnittes von Mengen

Die Schnittbildung von Mengen ist operationstreu in folgendem Sinn:

$$(\overline{A})_{\omega_1} = \overline{(A_{\omega_1})}, \quad \left(\bigcup_{i=1}^{\infty} A_i\right)_{\omega_1} = \bigcup_{i=1}^{\infty} (A_i)_{\omega_1}, \quad \left(\bigcap_{i=1}^{\infty} A_i\right)_{\omega_1} = \bigcap_{i=1}^{\infty} (A_i)_{\omega_1}$$

Ist $A := A_1 \times A_2,\ A_i \subset \Omega_i,\ i = 1, 2,$ so gilt:

$$(*) \qquad (A)_{\omega_1} = (A_1 \times A_2)_{\omega_1} = \begin{cases} A_2 \ , & \text{falls } \omega_1 \in A_1 \\ \varnothing \ , & \text{sonst.} \end{cases}$$

11.37 Satz:
Für $A \in \mathfrak{F}_1 \otimes \mathfrak{F}_2$ gilt $A_{\omega_1} \in \mathfrak{F}_2$.

Beweis:
Das Mengensystem $\mathfrak{F} := \{A \subseteq \Omega_1 \times \Omega_2 \mid A_{\omega_1} \in \mathfrak{F}_2\}$ definiert eine σ–Algebra. Wegen $(*)$ umfaßt $\mathfrak{F}$ die Rechteckmengen $A_1 \times A_2$ mit $A_i \in \mathfrak{F}_i,\ i = 1, 2$, und somit auch die von diesen Mengen erzeugte σ–Algebra $\mathfrak{F}_1 \otimes \mathfrak{F}_2$. ∎

11.38 Definition (Schnitt von Funktionen):
Ist X eine Abbildung $X: \Omega_1 \times \Omega_2 \longrightarrow \overline{\mathbb{R}}$, so nennt man die Funktion $X_{\omega_1}: \Omega_2 \longrightarrow \overline{\mathbb{R}}$, $\omega_1 \in \Omega_1$, mit $X_{\omega_1}(\omega_2) := X(\omega_1, \omega_2)$ den Schnitt von X an der Stelle $\omega_1 \in \Omega_1$. Entsprechend definiert man wieder X_{ω_2}.

Die Schnittbildung von Funktionen ist linear und monoton. Insbesondere gilt:

$$\left(\sum_{i=1}^{n} \alpha_i X_i\right)_{\omega_1} = \sum_{i=1}^{n} \alpha_i (X_i)_{\omega_1}.$$

Aus $X_n \nearrow X$ folgt $(X_n)_{\omega_1} \nearrow X_{\omega_1}$.

11.39 Satz:
Ist X $\mathfrak{F}_1 \otimes \mathfrak{F}_2$–$\overline{\mathfrak{B}}$–messbar, so ist X_{ω_1} $\mathfrak{F}_2$–$\overline{\mathfrak{B}}$–messbar.

Beweis:
Wegen $X^{-1}(B) \in \mathfrak{F}_1 \otimes \mathfrak{F}_2$ für alle $B \in \overline{\mathfrak{B}}$ gilt:

$$\begin{aligned}(X_{\omega_1})^{-1}(B) =& \{\omega_2 \in \Omega_2 \mid X_{\omega_1}(\omega_2) \in B\} \\ =& \{\omega_2 \in \Omega_2 \mid X(\omega_1, \omega_2) \in B\} = (X^{-1}(B))_{\omega_1} \in \mathfrak{F}_2 \ .\end{aligned}$$ ■

Der Zusammenhang zwischen Schnittbildung und Funktionen ist gegeben durch:

$$(I_A)_{\omega_1} = I_{(A_{\omega_1})} \quad \text{für } A \subseteq \Omega_1 \times \Omega_2.$$

11.40 Satz (Produktmaßsatz):
Es seien $(\Omega_i, \mathfrak{F}_i, \mu_i)$, $i = 1, 2$, zwei Maßräume, μ_i, $i = 1, 2$, σ–endlich und $A \in \mathfrak{F} := \mathfrak{F}_1 \otimes \mathfrak{F}_2$. Dann ist $\mu_2(A_{\omega_1})$ als Funktion von ω_1 $\mathfrak{F}_1$–$\overline{\mathfrak{B}}$–messbar und es gilt

$$(\mu_1 \otimes \mu_2)(A) = \int \mu_2(A_{\omega_1}) d\mu_1.$$

Beweis:
Wir zeigen der Reihe nach:

(1) $\mu_2(A_{\omega_1})$ ist $\mathfrak{F}_1$–$\overline{\mathfrak{B}}$–messbar,
wobei wir zunächst μ_2 als endlich (1. Fall) und dann als σ-endlich (2. Fall) voraussetzen,

(2) $(\mu_1 \otimes \mu_2)(A) = \int \mu_2(A_{\omega_1}) d\mu_1$.

Zu 1.: 1. Fall: $\mu_2(\Omega_2) < \infty$.
Setze $X_A(\omega_1) := \mu_2(A_{\omega_1})$ für $A \in \mathfrak{F}$. Betrachte

$$\mathfrak{F}' := \{A \in \mathfrak{F} \mid X_A \text{ ist } \mathfrak{F}_1\text{–}\overline{\mathfrak{B}}\text{–messbar}\} \subseteq \mathfrak{F}$$

und weise dieses als Dynkin-System nach. Setze anschließend $\mathcal{E} := \{A_1 \times A_2 \mid A_1 \in \mathfrak{F}_1 \text{ und } A_2 \in \mathfrak{F}_2\}$. Wegen $X_{A_1 \times A_2} = \mu_2(A_2) I_{A_1}$ für $A_1 \times A_2 \in \mathcal{E}$ ist $X_{A_1 \times A_2}$ $\mathfrak{F}_1$–$\overline{\mathfrak{B}}$–messbar und es gilt $\mathcal{E} \subseteq \mathfrak{F}' \subseteq \mathfrak{F}$. Da $\mathcal{E}$ durchschnittsstabil ist, folgt mit Satz 5.12, dass das von $\mathcal{E}$ erzeugte Dynkin–System mit der von $\mathcal{E}$ erzeugten σ–Algebra zusammenfällt, d.h. $\mathfrak{D}(\mathcal{E}) = \sigma(\mathcal{E})$. Damit ist für jedes endliche μ_2 die $\mathfrak{F}_1$–$\overline{\mathfrak{B}}$–Messbarkeit von $X_A(\omega_1) = \mu_2(A_{\omega_1})$ nachgewiesen.
Es bleibt zu zeigen, dass $\mathfrak{F}'$ ein Dynkin-System ist. Dazu

(a) $\Omega \in \mathfrak{F}'$:
Da $X_\Omega = \mu_2(\Omega_2)$ $\mathfrak{F}_1$–$\overline{\mathfrak{B}}$–messbar ist, gilt $\Omega \in \mathfrak{F}'$.

(b) Für $A, B \in \mathfrak{F}'$ mit $A \subseteq B$ gilt: $B \setminus A \in \mathfrak{F}'$.
Für jedes $\omega_1 \in \Omega_1$ gilt:

$$X_{B\setminus A}(\omega_1) = \mu_2((B \setminus A)_{\omega_1}) = \mu_2(B_{\omega_1} \setminus A_{\omega_1}).$$

Da μ_2 als endlich vorausgesetzt ist, gilt

$$X_{B\setminus A}(\omega_1) = \mu_2(B_{\omega_1}) - \mu_2(A_{\omega_1})$$

und damit $X_{B\setminus A} = X_B - X_A$, also $B \setminus A \in \mathfrak{F}'$.

(c) Für jede Folge $(A_n)_{n\in\mathbb{N}}$ paarweise disjunkter Mengen aus $\mathfrak{F}'$ ist auch $\bigcup\limits_{n\in\mathbb{N}} A_n \in \mathfrak{F}'$:
Für alle $n \in \mathbb{N}$ stellen deshalb die X_{A_n} $\mathfrak{F}_1$–$\overline{\mathfrak{B}}$–messbare Funktionen dar. Dann ist nach Satz 8.13 auch $\sum\limits_{n\in\mathbb{N}} X_{A_n}$ $\mathfrak{F}_1$–$\overline{\mathfrak{B}}$–messbar. Wegen $X_{\bigcup\limits_{n\in\mathbb{N}} A_n} = \sum\limits_{n\in\mathbb{N}} X_{A_n}$ folgt $\bigcup\limits_{n\in\mathbb{N}} A_n \in \mathfrak{F}'$.

Damit ist $\mathfrak{F}'$ als Dynkin–System nachgewiesen und folglich ist $X_A(\omega_1) = \mu_2(A_{\omega_1})$ $\mathfrak{F}_1$–$\overline{\mathfrak{B}}$–messbar für jedes endliche μ_2.

2. Fall: μ_2 σ–endlich.
Dieser Fall kann auf den ersten Fall zurückgeführt werden:
Es seien $(C_n)_{n\in\mathbb{N}}$ eine Folge von Elementen aus $\mathfrak{F}_2$ mit $C_1 \subseteq C_2 \subseteq C_3 \subseteq \ldots$, $\bigcup\limits_{n\in\mathbb{N}} C_n = \Omega_2$ und $\mu_2(C_n) < \infty \quad \forall\ n \in \mathbb{N}$. Für $n \in \mathbb{N}$ sei $\nu_n \colon \mathfrak{F}_2 \longrightarrow \overline{\mathbb{R}}$ definiert durch $\nu_n(A_2) := \mu_2(A_2 \cap C_n)$ für $A_2 \in \mathfrak{F}_2$. Dann ist ν_n ein endliches Maß auf $\mathfrak{F}_2$. Für beliebiges $A \in \mathfrak{F}$ wird $X_{A,n} \colon \Omega_1 \longrightarrow \overline{\mathbb{R}}$ mit $X_{A,n}(\omega_1) := \nu_n(A_{\omega_1})$, $\omega_1 \in \Omega_1$ definiert. Nach dem 1. Fall sind die Funktionen $X_{A,n}$ $\mathfrak{F}_1$–$\overline{\mathfrak{B}}$–messbar, weshalb auch $\sup\limits_{n\in\mathbb{N}} X_{A,n}$ $\mathfrak{F}_1$–$\overline{\mathfrak{B}}$–messbar ist.
Unter Ausnutzung der Monotonie erhält man:

$$\sup_{n\in\mathbb{N}} X_{A,n}(\omega_1) = \sup_{n\in\mathbb{N}} \nu_n(A_{\omega_1}) = \sup_{n\in\mathbb{N}} \mu_2(A_{\omega_1} \cap C_n) = \mu_2(A_{\omega_1}).$$

Damit ist X_A $\mathfrak{F}_1$–$\overline{\mathfrak{B}}$–messbar.

Zu 2.: $\mu(A) = (\mu_1 \otimes \mu_2)(A) = \int \mu_2(A_{\omega_1}) d\mu_1$
Definiere $\mu' \colon \mathfrak{F} \longrightarrow \mathbb{R}$ mit $\mu'(A) := \int X_A d\mu_1 \quad \forall\ A \in \mathfrak{F}$ und zeige, dass μ' ein Maß auf $\mathfrak{F}$ ist. Betrachte anschließend die Menge der kartesischen Produkte $\mathcal{E} := \{A_1 \times A_2 \mid A_1 \in \mathfrak{F}_1$ und $A_2 \in \mathfrak{F}_2\}$. Nach Satz 11.23 ist das Produktmaß $\mu = \mu_1 \otimes \mu_2$ durch seine Werte auf $\mathcal{E}$ bereits eindeutig festgelegt und es gilt für alle $A_1 \times A_2 \in \mathcal{E}$:

$$\mu'(A_1 \times A_2) = \int X_{A_1 \times A_2} d\mu_1 = \int \mu_2(A_2) I_{A_1}\, d\mu_1 = \mu_2(A_2) \cdot \mu_1(A_1).$$

Daher ist $\mu = \mu'$.
Es bleibt zu zeigen: μ' ist Maß auf $\mathfrak{F}$.

(a) μ' ist nichtnegativ:
Gilt nach Definition.

(b) $\mu^{'}$ ist σ–additiv:
Für eine Folge $(A_n)_{n\in\mathbb{N}}$ paarweise fremder Mengen aus $\mathfrak{F}$ gilt aufgrund von Satz 9.43:

$$\mu^{'}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \int X_{\bigcup\limits_{n\in\mathbb{N}} A_n} d\mu_1 = \int \left(\sum_{n\in\mathbb{N}} X_{A_n}\right) d\mu_1 = \sum_{n\in\mathbb{N}} \int X_{A_n} d\mu_1 = \sum_{n\in\mathbb{N}} \mu^{'}(A_n).$$

■

11.41 Satz (Satz von Fubini):
Es seien $(\Omega_i, F_i, \mu_i), i = 1,2$, zwei Maßräume, μ_i, $i = 1,2$, σ–endlich und $X : \Omega_1 \times \Omega_2 \longrightarrow \overline{\mathbb{R}}$ $\mathfrak{F}_1 \otimes \mathfrak{F}_2$–$\overline{\mathfrak{B}}$–messbar. Gilt

a) $X \geq 0$ oder

b) $\int |X|\, d(\mu_1 \otimes \mu_2) < \infty$,

dann ist

$$\int X\, d(\mu_1 \otimes \mu_2) = \int \left[\int X_{\omega_1} d\mu_2\right] d\mu_1.$$

(Siehe auch Lebensdaten von Fubini im Anhang D.)

Beweis:
Der Beweis folgt wieder dem Prinzip der sogenannten algebraischen Induktion:

(i) Zeige die Behauptung für Indikatorvariablen $X = I_A$, $A \in \mathfrak{F}_1 \otimes \mathfrak{F}_2$.

(ii) Zeige die Behauptung für Elementarfunktionen $X = \sum\limits_{k=1}^{n} \alpha_k I_{A_k}$.

(iii) Zeige die Behauptung für nichtnegative, messbare numerische Funktionen X.

(iv) Zeige die Behauptung für allgemeine meßbare numerische Funktionen.

Zu (i): Die Behauptung folgt unmittelbar aus Satz 11.40.

Zu (ii): Folgt aus (i) und der Linearität des Integrals.

Zu (iii): Folgt aus dem Satz von der monotonen Konvergenz (Satz 9.42).

Zu (iv): Im allgemeinen Fall sei wieder $X = X^+ - X^-$. Aufgrund von Voraussetzung b) gelten

$$\int X^+ d(\mu_1 \otimes \mu_2) < \infty \text{ und } \int X^- d(\mu_1 \otimes \mu_2) < \infty.$$

Damit folgen

$$\int \left[\int X^+_{\omega_1} d\mu_2\right] d\mu_1 < \infty \text{ und } \int \left[\int X^-_{\omega_1} d\mu_2\right] d\mu_1 < \infty.$$

Insgesamt ergibt sich also

$$\begin{aligned}\int X\, d(\mu_1 \otimes \mu_2) &= \int X^+ d(\mu_1 \otimes \mu_2) - \int X^- d(\mu_1 \otimes \mu_2) \\ &= \int \int X^+_{\omega_1} d\mu_2\, d\mu_1 - \int \int X^-_{\omega_1} d\mu_2\, d\mu_1 \\ &= \int \int X_{\omega_1} d\mu_2\, d\mu_1 .\end{aligned}$$

■

Als eine Anwendung des Satzes von Fubini wird der Multiplikationssatz für Erwartungswerte betrachtet:

11.42 Satz (Multiplikationssatz für Erwartungswerte):
Es seien X und Y zwei stochastisch unabhängige Zufallsgrößen, die eine der beiden Bedingungen

a) $(X \geq 0$ *und* $Y \geq 0)$ *oder*

b) $\mathrm{E}[|X|] < \infty$ *und* $\mathrm{E}[|Y|] < \infty$

erfüllen. Dann gilt: $\mathrm{E}[X \cdot Y] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$.

Beweis:
Ist a) erfüllt, so gilt aufgrund der stochastischen Unabhängigkeit von X und Y und dem Satz von Fubini:

$$\begin{aligned}\mathrm{E}[X \cdot Y] &= \int xy\, d(P_X \otimes P_Y) \\ &= \int x \left(\int y\, dP_Y \right) dP_X = \int x\, dP_X \cdot \int y\, dP_Y = \mathrm{E}[X] \cdot \mathrm{E}[Y] .\end{aligned}$$

Ist b) erfüllt, dann muss man auch noch $\mathrm{E}[|X \cdot Y|] < \infty$ nachweisen, um den Satz von Fubini anwenden zu können. Da aber wegen $|X| \geq 0$ und $|Y| \geq 0$ gilt: $\mathrm{E}[|X| \cdot |Y|] = \mathrm{E}[|X|] \cdot \mathrm{E}[|Y|]$ folgt aus Voraussetzung b) $\mathrm{E}[|X \cdot Y|] = \mathrm{E}[|X|] \cdot \mathrm{E}[|Y|] < \infty$. ■

11.43 Definition (Kovarianz, unkorreliert):
Es sei (X, Y) ein Zufallsvektor. Falls X und Y endliche Momente zweiter Ordnung besitzen, bezeichnet man

$$COV[X,Y] := \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] = \mathrm{E}[X \cdot Y] - \mathrm{E}[X] \cdot \mathrm{E}[Y]$$

als <u>Kovarianz</u> von X und Y.
Ist $COV[X,Y] = 0$ bezeichnet man X und Y als <u>unkorreliert</u>. Für unabhängige Zufallsvariablen X und Y ist stets $COV[X,Y] = 0$. Aus $COV[X,Y] = 0$ folgt im Allgemeinen jedoch nicht, dass X und Y unabhängig sind.

11.44 Beispiel:
$\mathrm{E}[X \cdot Y] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$ impliziert nicht, dass X und Y unabhängig sind. Es seien $Z \overset{\mathrm{d}}{=} U_{[-\frac{\pi}{2}, \frac{\pi}{2}]}$, $X := \sin(Z)$ und $Y := \cos(Z)$. Hieraus folgt:

$$X \cdot Y = \sin(Z) \cdot \cos(Z) = \frac{1}{2} \sin(2Z).$$

Hieraus folgen

$$\mathrm{E}[X] = \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin(t) dt = 0, \quad \mathrm{E}[Y] = \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos(t) dt = \frac{2}{\pi},$$

$$\mathrm{E}[X \cdot Y] = \frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin(2t) dt = 0, \text{ also } \mathrm{E}[X \cdot Y] = \mathrm{E}[X] \cdot \mathrm{E}[Y] = 0.$$

Literatur zu Kapitel 11

Folgende Bücher werden als begleitende Literatur zu diesem Kapitel des Skriptes empfohlen:

- H. Bauer:
  Wahrscheinlichkeitstheorie,
  5. Auflage, Walter de Gruyter, Berlin, 2002.
  ISBN: 3110172364

- W. Behnen, G. Neuhaus:
  Grundkurs Stochastik,
  3. Auflage, Teubner-Verlag, Stuttgart, 1995.
  ISBN: 3930737698

- P. Billingsley:
  Probability and Measure,
  2nd edition, John Wiley and Sons, New York, 1986.
  ISBN: 0471007102

- P. Gänssler und W. Stute:
  Wahrscheinlichkeitstheorie,
  Springer–Verlag, Berlin, 1977.
  ISBN: 3540084185

- E. Henze:
  Einführung in die Maßtheorie,
  Bibl. Institut, Mannheim, 1971.
  ISBN: 341100505X

- H. Heuser:
  Lehrbuch der Analysis. Teil 2 Mathematische Leitfäden,
  Teubner, Stuttgart, 2002.
  ISBN: 3519522322

- R. Mathar/D. Pfeiffer:
  Stochastik für Informatiker,
  Teubner, 1990.
  ISBN: 3519022400

- P. P. Spies:
  Grundlagen stochastischer Modelle,
  Hanser, München, 1982.
  ISBN: 3446137114

Kapitel 12

Schwaches und starkes Gesetz der großen Zahlen

Oftmals interessiert man sich für die Eigenschaften geeignet normierter Summen von Zufallsvariablen. Beim Würfelspiel zum Beispiel werden die für ein bestimmtes Ereignis günstigen Versuchsausgänge zusammengezählt und anschließend durch die Anzahl der Versuche dividiert, um zu einer Schätzung für die entsprechende Wahrscheinlichkeit zu gelangen. In diesem Fall stellt sich die Frage, unter welchen Bedingungen und mit welcher Konvergenzgeschwindigkeit sich Ausdrücke dieser Form stabilisieren. Eine Antwort darauf geben das schwache und starke Gesetz der großen Zahlen.

Schlüsselwörter: Markovsche Ungleichung, Tschebyscheffsche Ungleichung, Stochastische Konvergenz, i.i.d. Folge, Schwaches Gesetz der großen Zahlen, Satz von Bernoulli, fast sichere Konvergenz, Kolmogorovsche Ungleichung, Starkes Gesetz der großen Zahlen.

12.1 Das schwache Gesetz der großen Zahlen

Beim axiomatischen Aufbau der Wahrscheinlichkeitstheorie haben wir uns an den Eigenschaften der relativen Häufigkeiten orientiert. Von besonderer Bedeutung war dabei die Beobachtung gewesen, dass sich die relativen Häufigkeiten für große Stichprobenumfänge stabilisieren. Wir werden nun umgekehrt vorgehen, in dem wir den abstrakten Wahrscheinlichkeitsbegriff von A. N. Kolmogorov zugrundelegen und zeigen, dass sich die relativen Häufigkeiten $h_n(A)$ eines Ereignisses A, die wir als spezielle Zufallsvariablen auffassen, für $n \to \infty$ konvergieren. Dabei können verschiedene Konvergenzbegriffe zugrundegelegt werden. Wir beginnen mit dem Begriff der stochastischen Konvergenz.

12.1 Definition (stochastische Konvergenz):
Es seien $X, X_1, X_2, \ldots$ reellwertige Zufallsvariablen jeweils über demselben Wahrscheinlichkeitsraum $(\Omega, \mathfrak{F}, P)$. Man sagt, dass die Folge $(X_n)_{n=1}^{\infty}$ <u>stochastisch gegen X konvergiert</u>, kurz $X_n \xrightarrow{P} X$, falls für alle $\varepsilon > 0$ gilt:

$$P(|X_n - X| \geq \varepsilon) \to 0.$$

12.2 Beispiel:
Es seien $(\Omega, \mathfrak{F}, P)$ ein Wahrscheinlichkeitsraum und $(X_n)_{n=1}^{\infty}$ eine Folge von reellwertigen Zufallsvariablen über $(\Omega, \mathfrak{F}, P)$ mit $P(X_n = 1) := \frac{1}{n}$ und $P(X_n = 0) := 1 - \frac{1}{n}$ für $n = 1, 2, \ldots$. Dann gilt

$$P(|X_n| \geq \varepsilon) = \begin{cases} P(X_n = 1) = \frac{1}{n} & , \; 0 < \varepsilon \leq 1 \\ 0 & , \quad \varepsilon > 1. \end{cases}$$

Als Grenzwert ergibt sich

$$\lim_{n \to \infty} P(|X_n| \geq \varepsilon) = 0,$$

also $X_n \xrightarrow{P} 0$.

In der Regel kann die stochastische Konvergenz mit Hilfe der Ungleichungen von Markov und Tschebyscheff nachgewiesen werden.

12.3 Satz (Markovsche Ungleichung):
Es sei X eine reellwertige Zufallsvariable über $(\Omega, \mathfrak{F}, P)$ mit $\mathrm{E}[|X|^r] < \infty$ für $r > 0$. Dann gilt für alle $\varepsilon > 0$:

$$P(|X| \geq \varepsilon) \leq \frac{1}{\varepsilon^r} \mathrm{E}[|X|^r] \, .$$

(Siehe auch Lebensdaten von Markov im Anhang D.)

Beweis:
Es gilt

$$\begin{aligned} E[| X |^r] &= \int\limits_{\Omega} | X |^r \, dP = \int\limits_{\Omega} | X |^r \, I_{\{|X|<\varepsilon\}} dP + \int\limits_{\Omega} | X |^r \, I_{\{|X|\geq\varepsilon\}} dP \\ &\geq \int\limits_{\Omega} | X |^r \, I_{\{|X|\geq\varepsilon\}} dP \geq \varepsilon^r \int\limits_{\Omega} I_{\{|X|\geq\varepsilon\}} dP = \varepsilon^r P(|X| \geq \varepsilon). \end{aligned}$$

■

12.4 Satz (Tschebyscheffsche Ungleichung):
Es sei Y eine reellwertige Zufallsvariable mit $\mathrm{Var}[Y] < \infty$. Dann gilt für alle $\varepsilon > 0$:

$$P(|Y - \mathrm{E}[Y]| \geq \varepsilon) \leq \frac{\mathrm{Var}[Y]}{\varepsilon^2}.$$

(Siehe auch Lebensdaten von Tschebyscheff im Anhang D.)

Beweis:
Die Aussage folgt mit Satz 12.3 mit $X := Y - \mathrm{E}[Y]$ und $r := 2$. ■

Die nachfolgenden Beispiele zeigen, dass die Markovsche Ungleichung für numerische Zwecke ungeeignet ist.

12.5 Beispiel:

1. Es sei $X \stackrel{d}{=} \mathrm{Exp}(\lambda)$. Damit ergibt sich: $P(|X| \geq \varepsilon) \leq \frac{1}{\varepsilon}\mathrm{E}[|X|] = \frac{1}{\varepsilon \cdot \lambda}$.
   Für $\lambda := 1$ und $\varepsilon := 2$ erhält man damit $P(|X| \geq 2) \leq \frac{1}{2}$.
   Andererseits gilt aber $P(|X| \geq 2) = P(X \geq 2) = 1 - F_X(2) = e^{-2} \approx 0.14$.
2. Es sei $X \stackrel{d}{=} N(0, \sigma)$ und $\varepsilon := 3\sigma$ gewählt. Bei Anwendung der Tschebyscheffschen Ungleichung ergibt sich:
   $$P(|X| \geq 3\sigma) \leq \frac{1}{(3\sigma)^2}\mathrm{Var}[X] = \frac{1}{9} \approx 0.11.$$
   Den Tafeln der Normalverteilung entnimmt man jedoch:
   $$P(|X| \geq 3\sigma) \leq 0.01.$$

12.6 Definition (identisch verteilt, i.i.d. Folge):

a) Die Zufallsvariablen $X_1, X_2, \ldots$ heißen <u>identisch verteilt</u>, wenn $X_i \stackrel{d}{=} X_j$ für alle $i, j \in \mathbb{N}$ gilt.

b) Eine Folge $(X_n)_{n=1}^{\infty}$ von Zufallsvariablen, die stochastisch unabhängig und identisch verteilt sind, heißt eine <u>i.i.d. Folge</u> (i.i.d. $\equiv$ independent and identically distributed).

Mit diesen Vorbetrachtungen lassen sich die verschiedenen Versionen des schwaches der Gesetzes der großen Zahlen formulieren und beweisen.

12.7 Satz (Schwaches Gesetz der großen Zahlen):
Es sei $(X_i)_{i \in \mathbb{N}}$ eine Folge von identisch verteilten reellwertigen Zufallsvariablen, die paarweise unkorreliert sind d.h. $\mathrm{E}[X_i X_j] = \mathrm{E}[X_i] \cdot \mathrm{E}[X_j] \quad \forall\, i, j \in \mathbb{N}$ mit $i \neq j$. Außerdem sei $\mathrm{Var}[X_1] < \infty$. Dann gilt für alle $\varepsilon > 0$:

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

Insbesondere gilt

$$\lim_{n \to \infty}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \xrightarrow{P} \mathrm{E}[X_1].$$

Beweis:
Der Satz ist eine unmittelbare Konsequenz aus der Tschebyscheffschen Ungleichung und den folgenden Beziehungen:

$$\begin{aligned}
\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] &= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[X_i] = \frac{1}{n}(n\cdot \mathrm{E}[X_1]) = \mathrm{E}[X_1]\,,\\
\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_i] = \frac{1}{n^2}(n\cdot \mathrm{Var}[X_1]) = \frac{\mathrm{Var}[X_1]}{n}.
\end{aligned}$$

■

Anwendung: Bernoulli–Folgen
Es sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von stochastisch unabhängigen und $B(1,p)$–verteilten Zufallsvariablen über $(\Omega,\mathfrak{F},P)$, d.h. $P(X_i = 1) := p$ und $P(X_i = 0) := 1-p$ für alle $i \in \mathbb{N}$. Für $n\in\mathbb{N}$ wird $H_n := \sum\limits_{i=1}^{n} X_i$ gesetzt. Die Zufallsvariable H_n entspricht der absoluten Häufigkeit des Ereignisses $A := \{X_i = 1\}$ in einer Versuchsreihe der Länge n. Entsprechend definiert $h_n := \frac{1}{n}H_n$ die relative Häufigkeit von A.

12.8 Satz (Satz von Bernoulli):
Es sei $(X_i)_{i\in\mathbb{N}}$ eine i.i.d. Folge mit $P(X_i = 1) := p$ und $P(X_i = 0) := 1-p$. Dann gilt für alle $\varepsilon > 0$:

$$P\left(\left|\frac{H_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{p\cdot(1-p)}{n\cdot\varepsilon^2} \quad \textit{bzw.} \quad \frac{H_n}{n} \xrightarrow{P} p.$$

(Siehe auch Lebensdaten von Bernoulli im Anhang D.)

Beweis:
Der Satz ergibt sich durch Anwendung des schwachen Gesetzes der großen Zahlen 12.7 mit $H_n := \sum_{i=1}^{n} X_i, E[X_1] := p$ und $\mathrm{Var}[X_1] = E[(X_1)^2] - (E[X_1])^2 := p - p^2 = p(1-p)$. ■

Die Tschebyscheffsche Ungleichung lässt genügend Spielraum für geeignete Modifikationen des schwachen Gesetzes der großen Zahlen:

12.9 Satz (Schwaches Gesetz der großen Zahlen, verallgemeinerte Version):
Es bezeichne $(X_i)_{i\in\mathbb{N}}$ eine Folge von reellwertigen Zufallsvariablen über $(\Omega,\mathfrak{F},P)$. Außerdem sei mindestens eine der nachstehenden Bedingungen erfüllt:

a) $\mathrm{Var}\left[\sum\limits_{i=1}^{n} X_i\right] = o(n^2)$ *für* $n\to\infty$.

b) $COV[X_i, X_j] = 0 \quad \forall\, i,j\in\mathbb{N}$ *mit* $i\neq j$ *und* $\sum\limits_{i=1}^{n}\mathrm{Var}[X_i] = o(n^2)$ *für* $n\to\infty$.

c) $COV[X_i, X_j] = 0 \quad \forall\, i,j\in\mathbb{N}$ *mit* $i\neq j$, $F_{X_i} = F_{X_j} \;\; \forall\, i,j\in\mathbb{N}$ *und* $\mathrm{Var}[X_1] < \infty$.

d) $(X_i)_{i\in\mathbb{N}}$ *ist eine i.i.d. Folge und* $\mathrm{Var}[X_1] < \infty$.

Dann gilt:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \frac{1}{n}\sum_{i=1}^{n} E[X_i].$$

Bemerkung:
Es ist in obigem Satz mit $o(n^2)$ das wie folgt definierte Landau-Symbol gemeint:

$$f(n) = o(g(n)) \quad :\Longleftrightarrow \quad \lim_{n\to\infty} \frac{f(n)}{g(n)} = 0.$$

Beweis:
Die Voraussetzungen für c) und d) wurden bereits in Satz 12.7 behandelt. Da die Forderung a) offensichtlich schwächer ist als die Forderung b), genügt es, diesen Fall zu betrachten. Lässt sich die Aussage für den Fall, dass die Voraussetzungen unter a) erfüllt sind beweisen, so gilt sie auch sofort für den Fall, dass die Voraussetzungen unter b) erfüllt sind. Es gilt:

$$E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] \quad \text{und} \quad \operatorname{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\operatorname{Var}\left[\sum_{i=1}^{n} X_i\right]$$

Damit folgt:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E[X_i]\right| \geq \varepsilon\right) \leq \frac{\operatorname{Var}\left[\sum_{i=1}^{n} X_i\right]}{\varepsilon^2 \cdot n^2} = \frac{o(n^2)}{\varepsilon^2 \cdot n^2} \to 0 \quad \text{für} \quad n \to \infty.$$

■

12.10 Bemerkung:
Die stochastische Konvergenz macht nur eine Aussage über die Folge der Wahrscheinlichkeiten $P(|X_n - X| \geq \varepsilon), n = 1, 2, \ldots$. Über die Konvergenz der zugrundeliegenden Zufallsvariablen $X_n(\omega)$ als Funktionen von ω wird keine Aussage getroffen. Mit Fragen dieser Art beschäftigt sich der nächste Abschnitt.

12.2 Das starke Gesetz der großen Zahlen

Im vorherigen Abschnitt konnte mittels der stochastischen Konvergenz das schwache Gesetz der großen Zahlen aufgestellt werden. Im Folgenden wird dieser Konvergenzbegriff verschärft und das starke Gesetz der großen Zahlen formuliert.

12.11 Definition (*P*–fast–sicher):
Man sagt, eine Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen über einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathfrak{F}, P)$ strebt P–fast–sicher (P–f.s.) gegen die Zufallsvariable X, falls

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1 \qquad \textit{kurz} \qquad X_n(\omega) \longrightarrow X(\omega) \;\; P\text{–}f.s.$$

gilt, d.h. es gibt eine Nullmenge $N \in \mathfrak{F}$ mit $P(N) = 0$, so dass $\lim_{n\to\infty} X_n(\omega) = X(\omega) \;\; \forall\, \omega \in \overline{N}$.

12.12 Satz (Zusammenhang zwischen stochastischer und P–fast–sicherer Konvergenz):

a) *Konvergiert die Folge $(X_n)_{n\in\mathbb{N}}$ P–f.s. gegen X, so auch stochastisch, d.h. es gilt:*

$$X_n \longrightarrow X \quad P\text{–}f.s. \quad \Longrightarrow \quad X_n \xrightarrow{P} X \text{ für } n \to \infty.$$

b) *Konvergiert die Folge $\left\{\sup_{m\geq n} |X_m - X|\right\}_{n\in\mathbb{N}}$ stochastisch gegen 0, so konvergiert die Folge $(X_n)_{n\in\mathbb{N}}$ P–fast–sicher gegen X, d.h. es gilt:*

$$\left\{\sup_{m\geq n} |X_m - X|\right\} \xrightarrow{P} 0 \quad \Longrightarrow \quad X_n \longrightarrow X \quad P\text{–}f.s. \text{ für } n \to \infty.$$

Beweis:

a) Für alle $\varepsilon > 0$ gilt:

$$\begin{aligned}
0 &\leq \limsup_{n\to\infty} P(|X_n - X| \geq \varepsilon) \\
&\leq \limsup_{n\to\infty} P\left(\bigcup_{m=n}^{\infty} \{|X_m - X| \geq \varepsilon\}\right) \\
&= P\left(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} \{|X_m - X| \geq \varepsilon\}\right) \\
&= P(|X_n - X| \geq \varepsilon \text{ für unendlich viele } n) \\
&\leq 1 - P\left(\left\{\lim_{n\to\infty} X_n = X\right\}\right) \;=\; 1-1 \;=\; 0 \quad (\text{da } X_n \longrightarrow X \ P\text{–f.s.}).
\end{aligned}$$

b) Es wird $Y_n := \sup\limits_{m\geq n} |X_m - X| \;\; \forall\, n \in \mathbb{N}$ gesetzt. Offensichtlich ist

$$\begin{aligned}
Y_n &= \sup\{|X_n - X|, |X_{n+1} - X|, \dots\} \\
&\geq \sup\{|X_{n+1} - X|, |X_{n+2} - X|, \dots\} \;=\; Y_{n+1},
\end{aligned}$$

d.h. die Folge $(Y_n)_{n\in\mathbb{N}}$ ist monoton fallend. Da $(Y_n)_{n\in\mathbb{N}}$ nach unten durch 0 beschränkt ist, konvergiert $(Y_n)_{n\in\mathbb{N}}$ f.s. gegen eine Zufallsvariable $Y \geq 0$. Zusammen mit Teil a) folgt, dass $(Y_n)_{n\in\mathbb{N}}$ dann auch stochastisch gegen Y konvergiert. Wenn sich jetzt zeigen lässt, dass bei stochastischer Konvergenz der Grenzwert Y f.s. eindeutig bestimmt ist, muss aufgrund der Voraussetzung $Y = 0$ f.s. gelten und damit X_n f.s. gegen X. Sei deshalb Y' eine weitere Zufallsvariable, für die $\lim\limits_{n\to\infty} P(|Y_n - Y'| \geq \varepsilon) = 0$ gilt. Dann würde aber folgen

$$\begin{aligned}
P(\left|Y - Y'\right| \geq \varepsilon) &= P(\left|(Y - Y_n) - (Y' - Y_n)\right| \geq \varepsilon) \\
&\leq P\left(\left\{|Y - Y_n| \geq \frac{\varepsilon}{2}\right\} \cup \left\{\left|Y' - Y_n\right| \geq \frac{\varepsilon}{2}\right\}\right) \\
&\leq P\left(\left\{|Y - Y_n| \geq \frac{\varepsilon}{2}\right\}\right) + P\left(\left\{\left|Y' - Y_n\right| \geq \frac{\varepsilon}{2}\right\}\right) \longrightarrow 0 \text{ für } n \to \infty.
\end{aligned}$$

Aus $P(|Y| \geq \varepsilon) = 0 \;\; \forall\, \varepsilon > 0$ folgt

$$P(|Y| \neq 0) = P\left(\bigcup_{k\in\mathbb{N}} |Y| \geq \frac{1}{k}\right) = \lim_{k\to\infty} P\left(|Y| \geq \frac{1}{k}\right) = 0.$$

Wenn aber $Y_n \longrightarrow Y$ punktweise und $Y = 0$ P-f.s., so gilt $Y_n \longrightarrow 0$ P-f.s. und somit $X_n \longrightarrow X$ P-f.s. ■

12.13 Satz (Kolmogorovsche Ungleichung):
Es seien $X_1, \dots, X_n$ stochastisch unabhängige Zufallsgrößen mit den Mittelwerten $\mathrm{E}[X_i]$ *und den Varianzen* $\mathrm{Var}[X_i]$ *für $i = 1, \dots, n$. Die zugehörigen Partialsummen seien $S_m := \sum\limits_{i=1}^{m} X_i$, $m = 1, \dots, n$. Dann gilt für jedes $\varepsilon > 0$*

$$P\left(\max_{1 \le m \le n} |S_m - \mathrm{E}[S_m]| \ge \varepsilon\right) \le \frac{\mathrm{Var}[S_n]}{\varepsilon^2}.$$

(Siehe auch Lebensdaten von Kolmogorov im Anhang D.)

Beweis:
O.B.d.A. kann angenommen werden, dass $\mathrm{E}[X_m] = 0$ für $m = 1, \dots, n$ gilt. Es wird

$$A := \left\{\omega \in \Omega \mid \max_{1 \le m \le n} |S_m| \ge \varepsilon\right\} = \bigcup_{m=1}^{n} \{\omega \in \Omega \mid |S_m| \ge \varepsilon\} = \bigcup_{m=1}^{n} A_m,$$

mit

$$A_m := \{\omega \in \Omega \mid |S_1| < \varepsilon, \dots, |S_{m-1}| < \varepsilon, |S_m| \ge \varepsilon\} \qquad (m = 1, \dots, n)$$

gesetzt. Man beachte, dass die Mengen $A_1, A_2, \dots, A_n$ paarweise disjunkt sind.
Aus $S_n = S_k + S_n - S_k$ folgt

$$S_n^2 = S_k^2 + 2(S_n - S_k)S_k + (S_n - S_k)^2.$$

Damit wird

$$\begin{aligned}
\mathrm{Var}[S_n] &= \mathrm{E}\big[S_n^2\big] - (\mathrm{E}[S_n])^2 = \mathrm{E}\big[S_n^2\big] = \int_\Omega S_n^2 \, dP \\
&\ge \int_A S_n^2 \, dP = \sum_{k=1}^{n} \int_{A_k} S_n^2 \, dP \\
&= \sum_{k=1}^{n} \int_{A_k} \left(S_k^2 + 2(S_n - S_k)S_k + (S_n - S_k)^2\right) dP.
\end{aligned}$$

Es wird zunächst gezeigt, dass die Zufallsvariablen

$$[S_n - S_k] = X_{k+1} + \ldots + X_n \qquad \text{und} \qquad I_{A_k} \cdot S_k$$

stochastisch unabhängig sind. Dazu wird Satz 11.26 angewandt mit

$$\begin{aligned}
g_1 :& \quad \mathbb{R}^{n-k} \longrightarrow \mathbb{R} \quad , \quad (x_{k+1}, \dots, x_n) \longrightarrow \sum_{i=k-1}^{n} x_i \\
g_2 :& \quad \mathbb{R}^{k} \longrightarrow \mathbb{R} \quad , \quad (x_1, \dots, x_k) \longrightarrow \left(\sum_{i=1}^{k} x_i\right) f(x_1, \dots, x_k),
\end{aligned}$$

wobei gilt

$$f(x_1, \dots, x_k) := \begin{cases} 1 \, , & |x_1| < \varepsilon, \dots, |x_{k-1}| < \varepsilon, \ |x_k| \ge \varepsilon \\ 0 \, , & \text{sonst.} \end{cases}$$

Damit folgt

$$\begin{aligned}
\int_{A_k} (S_n - S_k) S_k \, dP &= \mathrm{E}[(S_n - S_k) \cdot I_{A_k} \cdot S_k] \\
&= \mathrm{E}[S_n - S_k] \cdot \mathrm{E}[I_{A_k} \cdot S_k] \\
&= 0 \cdot \mathrm{E}[I_{A_k} \cdot S_k] \;=\; 0.
\end{aligned}$$

Somit ist

$$\begin{aligned}
\mathrm{Var}[S_n] &= \sum_{k=1}^{n} \int_{A_k} S_k^2 \, dP + \sum_{k=1}^{n} \int_{A_k} (S_n - S_k)^2 \, dP \\
&\geq \sum_{k=1}^{n} \int_{A_k} S_k^2 \, dP \;\geq\; \sum_{k=1}^{n} \varepsilon^2 \int_{A_k} dP \;=\; \sum_{k=1}^{n} \varepsilon^2 \cdot P(A_k) \;=\; \varepsilon^2 \cdot P(A)
\end{aligned}$$

bzw.

$$P\left(\max_{1 \leq m \leq n} |S_m| \geq \varepsilon \right) \leq \frac{\mathrm{Var}[S_n]}{\varepsilon^2}.$$

■

12.14 Satz (Starkes Gesetz der großen Zahlen, 1. Fassung):
Es sei $(X_i)_{i \in \mathbb{N}}$ eine Folge stochastisch unabhängiger Zufallsvariablen über einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathfrak{F}, P)$ mit

$$\mathrm{E}[X_i] := \mu \;\; \forall \, i \qquad \textit{und} \qquad \sum_{i=1}^{\infty} \frac{\mathrm{Var}[X_i]}{i^2} < \infty.$$

Dann gilt:

$$\lim_{n \to \infty} \frac{1}{n} S_n = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu \quad P\text{-}f.s.$$

Beweis:
Für den Beweis wird Satz 12.12 b) und die Kolmogorovsche Ungleichung benutzt. Es wird gezeigt, dass

$$P\left(\sup_{m \geq n} \left| \frac{1}{m} S_m - \mu \right| \geq \varepsilon \right) = P\left(\bigcup_{m=n}^{\infty} \left\{ \left| \frac{1}{m} S_m - \mu \right| \geq \varepsilon \right\} \right) \longrightarrow 0 \text{ für } n \to \infty$$

gilt. Da diese Folge von Wahrscheinlichkeiten monoton in n ist, genügt es, sich beim Konvergenznachweis auf eine Teilfolge zu beschränken, etwa auf $n := 2^{r-1} + 1, r \to \infty$.

$$\begin{aligned}
P\left(\bigcup_{m=2^{r-1}+1}^{\infty}\left\{\left|\frac{1}{m}S_m-\mu\right|\geq\varepsilon\right\}\right) &= P\left(\bigcup_{i=r}^{\infty}\bigcup_{2^{i-1}<m\leq 2^i}\left\{\left|\frac{1}{m}S_m-\mu\right|\geq\varepsilon\right\}\right)\\
&\leq \sum_{i=r}^{\infty}P\left(\bigcup_{2^{i-1}<m\leq 2^i}\left\{\left|\frac{1}{m}S_m-\mu\right|\geq\varepsilon\right\}\right)\\
&= \sum_{i=r}^{\infty}P\left(\bigcup_{2^{i-1}<m\leq 2^i}\{|S_m-m\mu|\geq m\varepsilon\}\right)\\
&\leq \sum_{i=r}^{\infty}P\left(\bigcup_{2^{i-1}<m\leq 2^i}\{|S_m-m\mu|\geq 2^{i-1}\cdot\varepsilon\}\right)\\
&\leq \sum_{i=r}^{\infty}P\left(\bigcup_{1\leq m\leq 2^i}\{|S_m-m\mu|\geq 2^{i-1}\cdot\varepsilon\}\right)\\
&= \sum_{i=r}^{\infty}P\left(\max_{1\leq m\leq 2^i}|S_m-\mathrm{E}[S_m]|\geq 2^{i-1}\varepsilon\right)\\
&\leq \sum_{i=r}^{\infty}\frac{\mathrm{Var}[S_{2^i}]}{\varepsilon^2\cdot 2^{2(i-1)}} = \frac{4}{\varepsilon^2}\sum_{i=r}^{\infty}\frac{1}{2^{2i}}\sum_{j=1}^{2^i}\mathrm{Var}[X_j]\,.
\end{aligned}$$

(Kolmogorovsche Ungleichung) .

Die rechte Seite konvergiert für $r\to\infty$ gegen 0, falls $\sum\limits_{i=1}^{\infty}\frac{1}{2^{2i}}\sum\limits_{j=1}^{2^i}\mathrm{Var}[X_j]<\infty$ gilt.

$$\begin{aligned}
\sum_{i=1}^{\infty}\frac{1}{2^{2i}}\sum_{j=1}^{2^i}\mathrm{Var}[X_j] &= \sum_{j=1}^{\infty}\mathrm{Var}[X_j]\cdot\sum_{2^i\geq j}\frac{1}{2^i}\cdot\frac{1}{2^i}\\
&\leq \sum_{j=1}^{\infty}\frac{\mathrm{Var}[X_j]}{j^2}\cdot\sum_{k=0}^{\infty}\frac{1}{2^{2k}}\\
&= \sum_{j=1}^{\infty}\frac{\mathrm{Var}[X_j]}{j^2}\cdot\frac{1}{1-\frac{1}{4}}\\
&= \frac{4}{3}\sum_{j=1}^{\infty}\frac{\mathrm{Var}[X_j]}{j^2} < \infty.
\end{aligned}$$

■

Ganz analog kann mit Hilfe der Kolmogorovschen Ungleichung der folgenden Satz bewiesen werden:

12.15 Satz (Starkes Gesetz der großen Zahlen, 2. Fassung):
Es sei $(X_i)_{i\in\mathbb{N}}$ *eine Folge von stochastisch unabhängigen Zufallsgrößen über* $(\Omega, \mathfrak{F}, P)$ *mit Mittelwerten* $\mu_i := \mathrm{E}[X_i]$ *und Varianzen* ${\sigma_i}^2 := \mathrm{Var}[X_i] \;\; \forall\, i \in \mathbb{N}$. *Ist* $\sum\limits_{i=1}^{\infty} \frac{{\sigma_i}^2}{i^2} < \infty$ *, so gilt*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow \frac{1}{n}\sum_{i=1}^{n} \mu_i \;\; P\text{–}f.s.$$

Im Fall stochastisch unabhängiger und identisch verteilter Zufallsvariablen $X_1, X_2, \ldots$ kann auf die Existenz der Varianz für die Gültigkeit des starken Gesetzes der roßen Zahlen sogar verzichtet werden (siehe Bauer, Kapitel 6):

12.16 Satz (Starkes Gesetz der großen Zahlen, 3. Fassung):
Es sei $(X_i)_{i\in\mathbb{N}}$ *eine i.i.d. Folge mit* $\mathrm{E}[|X_1|] < \infty$. *Dann gilt*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mathrm{E}[X_1] \quad P\text{–}f.s.$$

Das starke Gesetz der großen Zahlen besagt, dass unter geeigneten Regularitätsbedingungen die Mittelwertbildung bei Zufallsvariablen zu einem Ausgleich der zufälligen Schwankungen führt. Diese Tatsache gilt keineswegs allgemein, wie das nachfolgende Beispiel zeigt.

12.17 Beispiel:
Es sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von stochastisch unabhängigen und mit den Parametern $\lambda := 1$ und $\mu := 0$ identisch Cauchy-verteilten Zufallsvariablen, d.h.

$$f_{X_i}(x) := \frac{1}{\pi} \cdot \frac{1}{1+x^2}, \;\; x \in \mathbb{R}, \; i \in \mathbb{N}.$$

Man zeigt, dass das arithmetische Mittel der Zufallsvariablen $X_1, X_2, \ldots$ ebenfalls Cauchy-verteilt ist mit den Parametern $\lambda := 1$ und $\mu := 0$, weshalb die Mittelwertbildung zu keinem Ausgleich von Schwankungen führt. Andererseits ist aber wegen

$$\int_0^a \frac{x}{1+x^2} dx = \frac{1}{2} ln(1+a^2)$$

auch

$$\mathrm{E}[|X_1|] = \int_{\mathbb{R}} |x|\, f_{X_1}(x) dx = \frac{1}{\pi}\int_{\mathbb{R}} \frac{|x|}{1+x^2} dx = +\infty,$$

so dass die Voraussetzungen des starken Gestzes der großen Zahlen 12.16 auch nicht erfüllt sind.

Literatur zu Kapitel 12

Folgende Bücher werden als begleitende Literatur zu diesem Kapitel des Skriptes empfohlen:

- H. BAUER:
  Wahrscheinlichkeitstheorie,
  5. Auflage, Walter de Gruyter, Berlin, 2002.
  ISBN: 3110172364

- W. Behnen, G. Neuhaus:
  Grundkurs Stochastik,
  3. Auflage, Teubner-Verlag, Stuttgart, 1995.
  ISBN: 3930737698

- M. A. Berger:
  An Introduction to Probability and Stochastic Processes,
  Springer–Verlag, New York, 1992.
  ISBN: 3540977848

- P. Billingsley:
  Probability and Measure,
  2nd edition, John Wiley and Sons, New York, 1986.
  ISBN: 0471007102

- P. Gänssler und W. Stute:
  Wahrscheinlichkeitstheorie,
  Springer–Verlag, Berlin, 1977.
  ISBN: 3540084185

- H. O. Georgii:
  Stochastik,
  2. Auflage, de Gruyter, 2004.
  ISBN: 3110172356

- R. Mathar/D. Pfeiffer:
  Stochastik für Informatiker,
  Teubner, 1990.
  ISBN: 3519022400

- P. P. Spies:
  Grundlagen stochastischer Modelle,
  Hanser, München, 1982.
  ISBN: 3446137114

Anhang A

Tabelle der χ^2–Verteilung

$F(x)$	Anzahl der Freiheitsgrade									
	1	2	3	4	5	6	7	8	9	10
0.001	0.00	0.00	0.02	0.09	0.21	0.38	0.60	0.86	1.15	1.48
0.005	0.00	0.01	0.07	0.21	0.41	0.68	0.99	1.34	1.73	2.16
0.01	0.00	0.02	0.11	0.30	0.55	0.87	1.24	1.65	2.09	2.56
0.025	0.00	0.05	0.22	0.48	0.83	1.24	1.69	2.18	2.70	3.25
0.05	0.00	0.10	0.35	0.71	1.15	1.64	2.17	2.73	3.33	3.94
0.1	0.02	0.21	0.58	1.06	1.61	2.20	2.83	3.49	4.17	4.87
0.25	0.10	0.58	1.21	1.92	2.67	3.45	4.25	5.07	5.90	6.74
0.5	0.45	1.39	2.37	3.36	4.35	5.35	6.35	7.34	8.34	9.34
0.75	1.32	2.77	4.11	5.39	6.63	7.84	9.04	10.22	11.39	12.55
0.9	2.71	4.61	6.25	7.78	9.24	10.64	12.02	13.36	14.68	15.99
0.95	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31
0.975	5.02	7.38	9.35	11.14	12.83	14.45	16.01	17.53	19.02	20.48
0.99	6.63	9.21	11.35	13.28	15.09	16.81	18.48	20.09	21.67	23.21
0.995	7.88	10.69	12.84	14.86	16.75	18.55	20.28	21.96	23.59	25.19
0.999	10.83	13.82	16.27	18.47	20.52	22.46	24.32	26.13	27.88	29.59

$F(x)$	Anzahl der Freiheitsgrade									
	11	12	13	14	15	16	17	18	19	20
0.001	1.83	2.21	2.62	3.04	3.48	3.94	4.42	4.90	5.41	5.92
0.005	2.60	3.07	3.57	4.07	4.60	5.14	5.70	6.26	6.84	7.43
0.01	3.05	3.57	4.11	4.66	5.23	5.81	6.41	7.01	7.63	8.26
0.025	3.82	4.40	5.01	5.63	6.26	6.91	7.56	8.23	8.91	9.59
0.05	4.57	5.23	5.89	6.57	7.26	7.96	8.67	9.39	10.12	10.85
0.1	5.58	6.30	7.04	7.79	8.55	9.31	10.09	10.86	11.65	12.44
0.25	7.58	8.44	9.30	10.17	11.04	11.91	12.79	13.68	14.56	15.45
0.5	10.34	11.34	12.34	13.34	14.34	15.34	16.34	17.34	18.34	19.34
0.75	13.70	14.85	15.98	17.12	18.25	19.37	20.49	21.60	22.72	23.83
0.9	17.28	18.55	19.81	21.06	22.31	23.54	24.77	25.99	27.20	28.41
0.95	19.68	21.03	22.36	23.68	25.00	26.30	27.59	28.87	30.14	31.41
0.975	21.92	23.34	24.74	26.12	27.49	28.85	30.19	31.53	32.85	34.17
0.99	24.73	26.22	27.69	29.14	30.58	32.00	33.41	34.81	36.19	37.57
0.995	26.76	28.30	29.82	31.32	32.80	34.27	35.72	37.16	38.58	40.00
0.999	31.26	32.91	34.53	36.12	37.70	39.25	40.79	42.31	43.82	45.32

Beispiel: Bei 3 Freiheitsgraden ist $F = 0.99$ für $x = 11.35$.

$F(x)$	Anzahl der Freiheitsgrade									
	21	22	23	24	25	26	27	28	29	30
0.001	6,4	7,0	7,5	8,1	8,7	9,2	9,8	10,4	11,0	11,6
0.005	8,0	8,6	9,3	9,9	10,5	11,2	11,8	12,5	13,1	13,8
0,01	8,9	9,5	10,2	10,9	11,5	12,2	12,9	13,6	14,3	15,0
0,025	10,3	11,0	11,7	12,4	13,1	13,8	14,6	15,3	16,0	16,8
0,05	11,6	12,3	13,1	13,8	14,6	15,4	16,2	16,9	17,7	18,5
0,1	13,2	14,0	14,8	15,7	16,5	17,3	18,1	18,9	19,8	20,6
0,25	6,3	17,2	18,1	19,0	19,9	20,8	21,7	22,7	23,6	24,5
0,5	20,3	21,3	22,3	23,3	24,3	25,3	26,3	27,3	28,3	29,3
0,75	24,9	26,0	27,1	28,2	29,3	30,4	31,5	32,6	33,7	34,8
0,9	29,6	30,8	32,0	33,2	34,4	35,6	36,7	37,9	39,1	40,3
0,95	32,7	33,9	35,2	36,4	37,7	38,9	40,1	41,3	42,6	43,8
0,975	35,5	36,8	38,1	39,4	40,6	41,9	43,2	44,5	45,7	47,0
0,99	38,9	40,3	41,6	43,0	44,3	45,6	47,0	48,3	49,6	50,9
0,995	41,4	42,8	44,2	45,6	46,9	48,3	49,6	51,0	52,3	53,7
0,999	46,8	48,3	49,7	51,2	52,6	54,1	55,5	56,9	58,3	59,7

$F(x)$	Anzahl der Freiheitsgrade						
	40	50	60	70	80	90	100	> 100 (Näherung)
0,001	17,9	24,7	31,7	39,0	46,5	54,2	61,9	$(h-3,09)^2/2$
0,005	20,7	28,0	35,5	43,3	51,2	59,2	67,3	$(h-2,58)^2/2$
0,01	22,2	29,7	37,5	45,4	53,5	61,8	70,1	$(h-2,33)^2/2$
0,025	24,4	32,4	40,5	48,8	57,2	65,6	74,2	$(h-1,96)^2/2$
0,05	26,5	34,8	43,2	51,7	60,4	69,1	77,9	$(h-1,64)^2/2$
0,1	29,1	37,7	46,5	55,3	64,3	73,3	82,4	$(h-1,28)^2/2$
0,25	33,7	42,9	52,3	61,7	71,1	80,6	90,1	$(h-0,67)^2/2$
0,5	39,3	49,3	59,3	69,3	79,3	89,3	99,3	$h^2/2$
0,75	45,6	56,3	67,0	77,6	88,1	98,6	109,1	$(h+0,67)^2/2$
0,9	51,8	63,2	74,4	85,5	96,6	107,6	118,5	$(h+1,28)^2/2$
0,95	55,8	67,5	79,1	90,5	101,9	113,1	124,3	$(h+1,64)^2/2$
0,975	59,3	71,4	83,3	95,0	106,6	118,1	129,6	$(h+1,96)^2/2$
0,99	63,7	76,2	88,4	100,4	112,3	124,1	135,8	$(h+2,33)^2/2$
0,995	66,8	79,5	92,0	104,2	116,3	128,3	140,2	$(h+2,58)^2/2$
0,999	73,4	86,7	99,6	112,3	124,8	137,2	149,4	$(h+3,09)^2/2$

In der letzten Spalte ist $h = \sqrt{2m-1}$
(m = Anzahl der Freiheitsgerade)

Anhang B

Zeichenerklärungen

$\mathbb{N}$	Menge der natürlichen Zahlen
$\mathbb{Q}$	Menge der rationalen Zahlen
$\mathbb{R}$	Menge der reellen Zahlen
$\mathbb{C}$	Menge der komplexen Zahlen
$\mathbb{M}$	Menge der maßdefinierenden Funktionen auf $\mathbb{R}$, die in $(-\infty, 0)$ verschwinden
$\mathbb{B}$	Menge der nichtnegativen reellen Funktionen, die auf jedem Intervall der Form $[0, t]$ beschränkt sind
$\mathbb{I}^n$	Menge der links offenen und rechts abgeschlossenen Intervalle im $\mathbb{R}^n$, $n \in \mathbb{N}$
$\mathfrak{B}^n$	$:= \sigma(\mathbb{I}^n)$ "σ–Algebra der Borelschen Mengen des $\mathbb{R}^n$"
$\overline{\mathbb{R}}$	$:= \mathbb{R} \cup \{-\infty, +\infty\}$
$\overline{\mathfrak{B}}$	$:= \{B, B \cup \{-\infty\}, B \cup \{+\infty\}, B \cup \{-\infty, +\infty\} \vert B \in \mathfrak{B}\}$
$\mathfrak{P}(M)$	Potenzmenge von M
$(a, b]$	$:= \{x \vert a < x \leq b\}$ "links offenes, rechts abgeschlossenes Intervall"
$n!$	$:= n \dots (n-1) \cdot \dots \cdot 2 \cdot 1$ "Fakultät von n"
$(N)_n$	$:= \frac{N!}{n!} = N \cdot (N-1) \cdot \dots \cdot (N-n+1)$ "n–te untere Faktorielle von N"
$\binom{n}{k}$	$:= \dfrac{n!}{k! \cdot (n-k)!}$ "n über k"
$F(a-0)$	meint den linksseitigen Limes von $F(a)$
$\uparrow$	konvergiert von unten gegen
$X \stackrel{d}{=} \mathrm{Exp}(\lambda)$	X ist exponential–verteilt
$X \stackrel{d}{=} Y$	X und Y sind identisch verteilt
$\Re(x)$	Realteil der komplexen Zahl x
$\Im(x)$	Imaginärteil der komplexen Zahl x
$O(n)$, $o(n)$	seien die Landau-Symbole.

Anhang C

Literatur

Folgende Bücher werden als begleitende Literatur zu diesem Skript empfohlen:

- H. BAUER:
  Maß- und Integrationstheorie,
  Walter de Gruyter, Berlin, 1990.
  ISBN: 3110127725
  Preis: 26.95 €
  Kurzbeschreibung:
  ,,Viele Gebiete der Mathematik und ihrer Anwendungen [...] erfordern solide Kenntnisse aus der Maß- und Integrationstheorie. Das Lehrbuch [...] führt den Leser [...] schnell, verlässlich und präzise zu den wichtigsten Ergebnissen der Maß- und Integrationstheorie hin. [...] Zahlreiche Beispiele erläutern die Bedeutung der erzielten Ergebnisse.[...] Übungsaufgaben laden den Leser zum vertieften Eindringen in den behandelten Stoff ein."

- H. BAUER:
  Wahrscheinlichkeitstheorie,
  5. Auflage, Walter de Gruyter, Berlin, 2002.
  ISBN: 3110172364
  Preis: 36.95 €
  Kurzbeschreibung:
  ,,Das vorliegende Buch soll dem Studierenden als Wegführer in die Wahrscheinlichkeitstheorie dienen. Der Leser soll dabei mit den wichtigsten Ideen, Methoden und Resultaten dieser sich heute schnell entwickelnden und verzweigenden mathematischen Theorie bekanntgemacht werden. [...] Da heutzutage die Wahrscheinlichkeitstheorie unlöslich mit der Maß- und Integrationstheorie verbunden ist, verfolgt das Buch zugleich aber auch ein zweites Ziel, nämlich den Leser mit den Grundzügen der Maßtheorie vertraut zu machen. [...]"

- W. BEHNEN, G. NEUHAUS:
  Grundkurs Stochastik,
  3. Auflage, Teubner-Verlag, Stuttgart, 1995.
  ISBN: 3930737698
  Preis: 24.00 €

Kurzbeschreibung:
,,Eine integrierte Einführung in die Wahrscheinlichkeitstheorie und Mathematische Statistik für Mathematiker, Wirtschaftsmathematiker, Informatiker und Physiker.
Es enthält: Wahrscheinlichkeitsmodelle, Anwendungspostulat und statistische Tests; mehrstufige Zufallsexperimente und grundlegende diskrete Modelle; Wahrscheinlichkeitsmodelle über euklidischen Räumen; Koppelung von allgemeinen Zufallsexperimenten (Satz von Fubini); Parameterschätzung (auch in approximativen Modellen); Konfidenzbereiche für Modellparameter; das Testen von Hypothesen."

- M. A. Berger:
  An Introduction to Probability and Stochastic Processes,
  Springer–Verlag, New York, 1992.
  ISBN: 3540977848
  Kurzbeschreibung:
  ,,This is a textbook which will provide students with a straightforward introduction to the mathematical theory of probability. It is written with the aim of presenting the central results and techniques of the subject in a complete and self-contained account. [...] Any student who has a familiarity with calculus and basic algebra will be able to use this text and throughout there are a wide variety of exercises to illustrate and to develop ideas. [...]"

- O. Beyer, H. Hackel, V. Pieper, J. Tiedge:
  Wahrscheinlichkeitsrechnung und mathematische Statistik,
  7. Auflage, Teubner-Verlag, Stuttgart, 1995.
  ISBN: 3-8154-2075-X
  Kurzbeschreibung:
  ,,Die Reihe ,,Mathematik für Ingenieure, Naturwissenschaftler, Ökonomen und Landwirte" umfasst den [...] Lehrstoff für die Mathematikausbildung der genannten Disziplinen, bietet Möglichkeiten zur Vertiefung sowie Spezialisierung und unterstützt die Individualisierung des Studiums. [...] Das Lehrwerk ist nach modernen fachlichen und hochschulpädagogischen Prinzipien aufgebaut. [...]"

- P. Billingsley:
  Probability and Measure,
  2nd edition, John Wiley and Sons, New York, 1986.
  ISBN: 0471007102
  Preis: 102.90 €
  Kurzbeschreibung:
  ,,Intertwines measure theory and modern probability: probability problems generate an interest in measure theory and measure theory is then developed and applied to probability. Illustrates the connections probability theory has with applied mathematics on the one hand and with pure mathematics on the other."

- M. Fisz:
  Wahrscheinlichkeitsrechnung und mathematische Statistik,
  VEB, Deutscher Verlag der Wissenschaften 1989.
  ISBN: 3326000790

Kurzbeschreibung:
,,Dieses Buch ist in der Hauptsache für Mathematiker bestimmt; es dürfte aber auch [...] solchen Lesern zugänglich sein, die [...] über gewisse Kenntnisse in der höheren Mathematik verfügen und sich für die Anwendungen der Wahrscheinlichkeitsrechnung interessieren. Der Leser findet in diesem Buch eine Einführung in die moderne Wahrscheinlichkeitsrechnung und die moderne mathematische Statistik. [...] Das Buch enthält zahlreiche Anwendungsbeispiele. [...]"

- P. Gänssler und W. Stute:
  Wahrscheinlichkeitstheorie,
  Springer–Verlag, Berlin, 1977.
  ISBN: 3540084185
  Kurzbeschreibung:
  ,,Für das Verständnis des vorliegenden Textes sind [...] Grundkenntnisse aus einer Vorlesung ,,Einführung in die Wahrscheinlichkeitstheorie und Mathematische Statistik" wünschenswert. [...] Auf eine Diskusion diskreter Modelle ist deshalb bewusst verzichtet worden. Die [...] getroffene Stoffauswahl umfasst eine zweisemestrige Vorlesung über Wahrscheinlichkeitstheorie. Neben der Vermittlung klassischer Grundlagen liegt der methodische Schwerpunkt auf der Konstruktion stochastischer Modelle unter besonderer Berücksichtigung einiger für die Anwendungen in der Mathematischen Statistik wichtigen Resultate. [...] "

- H. O. Georgii:
  Stochastik,
  2. Auflage, de Gruyter, 2004.
  ISBN: 3110172356

- M. Greiner/G. Tinhofer:
  Stochastik für Studienanfänger der Informatik,
  Hanser, München, 1996.
  ISBN: 3446186360
  Kurzbeschreibung:
  ,,Dieses Lehrbuch bietet einen Grundstock an Lehrstoff aus Wahrscheinlichkeitstheorie und Statistik, wie er in der Informatik benötigt wird und verbindet diesen Lehrstoff mit der Begriffswelt, die Informatiker in ihrem Berufsalltag vorfinden. Hierbei wird besonderes Gewicht auf die Aspekte Methodik und Modellierung gelegt. Der Leser soll [...] zukünftig in der Lage sein, Fragen aus seinem Berufsalltag in ein geeignetes stochastisches Modell umzusetzen und die ermittelten Resultate anschließend im Rahmen der ursprünglichen Fragestellung zu interpretieren.[...] Mehr als hundert Beispiele, Aufgaben und deren Lösungen sowie ein Kompromiss zwischen mathematischer Strenge und ausgewogener textlicher Darstellung des Stoffes motivieren den Leser zur aktiven Teilnahme an der Entwicklung und Lösung von Problemen aus der Stochastik."

- E. Henze:
  Einführung in die Maßtheorie,
  Bibl. Institut, Mannheim, 1971.
  ISBN: 341100505X

Kurzbeschreibung:
,,Bei dieser Einführung in die Maß- und Integrationstheorie werden gleichzeitig die notwendigen Ergebnisse und Methoden für den Einstieg in die moderne Wahrscheinlichkeitstheorie bereitgestellt. Das Buch wendet sich in erster Linie an Studenten der Mathematik, der Informatik und der Physik, kann aber auch anderen interessierten Lesern von Nutzen sein."

- E. Henze:
Stochastik für Einsteiger,
Bibl. Institut, Mannheim, 1997.
ISBN: 3528368942
Kurzbeschreibung:
,,[...] Dieses Buch soll dem Leser einen Einstieg in die Stochastik, die Kunst des ,,geschickten Vermutens", vermitteln und ihn in die Lage versetzen [...] kritisch und kompetent mitreden zu können. Es enthält 160 Übungsaufgaben mit Lösungen. [...] Als Lehrbuch zwischen gymnasialem Mathematikunterricht und Universität wendet es sich unter anderem an: [...] Studienanfänger an Universitäten, Fachhochschulen und Berufsakademien; Quereinsteiger aus Industrie und Wirtschaft."

- H. Heuser:
Lehrbuch der Analysis. Teil 2 Mathematische Leitfäden,
Teubner, Stuttgart, 2002.
ISBN: 3519522322
Kurzbeschreibung:
,,Bei der Abfassung des zweiten Bandes [...] wollte ich die Theorie ausführlich und fasslich darstellen, ausgiebig motivieren und durch viele Beispiele und Übungen zum sicheren Besitz des Lesers machen. Außerdem wollte ich Brücken schlagen zu den Anwendungen analytischer Methoden in den allerverschiedensten Wissenschaften. [...] Dabei stehen diesmal im Vordergrund der Überlegungen Funktionen, deren Argumente und Werte Vektoren aus dem $\mathbb{R}^p$ oder sogar Elemente aus noch viel allgemeineren Räumen sind. [...]"

- K. Hinderer:
Grundbegriffe der Wahrscheinlichkeitstheorie,
Springer–Verlag, 1980.
ISBN: 3540073094
Kurzbeschreibung:
,,Das Buch [...] bietet eine solide, gut motivierte Darstellung mit einer Fülle konkreter Beispiele, ergänzt durch sorgfältig ausgesuchte Aufgaben nach jedem Paragraphen. Sowohl die historischen als auch die weiterführenden Bemerkungen geben eine gute Übersicht über Probleme und Fragestellungen aus der Wahrscheinlichkeitstheorie."

- G. Hübner:
Stochastik. Eine Einführung für Mathematiker, Informatiker und Ingenieure.,
4. Auflage, Vieweg Verlag, 2003.
ISBN: 3528254432
Preis: 22.50 €

Kurzbeschreibung:
,,Dieses Buch soll Informatiker, Ingenieure und Mathematiker in die Lage versetzen, konkrete Vorgänge mit Zufallseinfluss in den wesentlichen Aspekten zu verstehen, zu modellieren und daraus Prognosen und Entscheidungshilfen abzuleiten. [...] Das Buch [...] richtet sich [...] an [...] Informatiker, Ingenieure, Mathematiker und Mathematik–Lehrer, die sich grundlegende Kenntnisse in stochastischer Modellierung und erste Einblicke in Anwendungsbereiche verschaffen wollen. [...] Besonders auf die Belange der Informatik zugeschnitten ist die Einbeziehung von Modellen und Bewertungen für Bedienungsprobleme und Kommunikationsnetze auf elementarem Niveau. [...]"

- U. Krengel:
  Einführung in die Wahrscheinlichkeitstheorie und Statistik,
  6. Auflage, Vieweg, 2002.
  ISBN: 3528672595
  Preis: 22.90 €
  Kurzbeschreibung:
  ,,Dieses Buch wendet sich an alle, die [...] in die Ideenwelt der Stochastik eindringen möchten. Stochastik ist die Mathematik des Zufalls. [...] Die beiden Hauptgebiete der Stochastik sind Wahrscheinlichkeitstheorie und Statistik. In der Wahrscheinlichkeitstheorie untersucht man zufällige Prozesse mit festen als bekannt angenommenen steuernden Wahrscheinlichkeiten. [...] Darüber hinaus liefert die Wahrscheinlichkeitstheorie Grundlagen für die Statistik, in der aus beobachteten Daten Schlüsse über unbekannte Wahrscheinlichkeiten und über zweckmäßiges Verhalten gezogen werden sollen. [...]"

- K. Krickeberg/H. Ziezold:
  Stochastische Methoden,
  4. Auflage, Springer–Verlag, Berlin, 1995.
  ISBN: 3-540-57792-0
  Kurzbeschreibung:
  ,,Im Vordergrund [...] stehen die eigentlichen ,,stochastischen" Ideen und ihre praktischen Anwendungen, insbesondere in der Statistik, ohne dass mathematische strenge und Schönheit zu kurz kommen. Über die üblichen Grundlagen hinaus finden sich Kapitel über Simulation, nichtparametrische Statistik und Regression- und Varianzanalyse. [...] Besonderer Anziehungspunkt dieses Buches ist die ,,genetische" Entwicklung der verschiedenen Typen von Wahrscheinlichkeitsverteilungen, ausgehend von der hypergeometrischen Verteilung. [...]"

- J. Lehn/H. Wegmann:
  Einführung in die Statistik,
  4. Auflage, Teubner, 2004.
  ISBN: 3519320711
  Preis: 22.90 €
  Kurzbeschreibung:
  ,,Eine elementare Darstellung statistischer Schätz– und Testverfahren einschließlich der zugrundeliegenden Modellbildung für Mathematiker, Informatiker, Wirtschaftswissenschaftler, Naturwissenschaftler und Ingenieure.

Es enthält: Methoden der Beschreibenden Statistik; Zufallsvariablen und ihre Verteilungen; Gesetze der Großen Zahlen und ihre Eigenschaften; Tests bei Normalverteilungsannahmen; χ^2-Tests und Kontingenztafeln; verteilungsunabhängige Tests; einfache Varianzanalyse und Regression."

- R. Mathar/D. Pfeiffer:
  Stochastik für Informatiker,
  Teubner, 1990.
  ISBN: 3519022400
  Kurzbeschreibung:
  ,,Das vorliegende Buch [...] wendet sich vor allem an Informatikstudenten und Mathematikstudenten mit Nebenfach Informatik mit dem Ziel, stochastische Grundbegriffe unter besonderer Berücksichtigung Informatik–spezifischer Aspekte zu vermitteln. [...] Ziel des Buches ist es daher, eine einheitliche und möglichst geschlossene Übersicht über die zum Verständnis benötigten Grundlagen zu geben. [...] Trotz des überwiegenden Lehrbuchcharakters dieses Textes haben wir uns allerdings auch bemüht, neuere Entwicklungen, die z.T. bisher nur in Originalarbeiten vorliegen, mit einzubeziehen, um dort, wo es im Rahmen unseres Zugangs möglich ist, Anschluss an Fragestellungen der aktuellen Forschung zu erlangen. [...]"

- J. Pfanzagl:
  Elementare Wahrscheinlichkeitsrechnung,
  Gruyter, Berlin, 1988.
  ISBN: 3110114194
  Kurzbeschreibung:
  ,,Die vorliegende Einführung der Wahrscheinlichkeitsrechnung ist ,,elementar" in dem Sinne, dass weder Kenntnisse aus der Maßtheorie noch aus der Funktionentheorie vorausgesetzt werden. [...] Das Anliegen des Buches ist die Entwicklung anwendungsbezogenen stochastischen Denkens. Diesem Ziel dient eine verhältnismäßig große Anzahl von Beispielen, die [...] zeigen sollen, dass es sich bei der Wahrscheinlichkeitsrechnung um ein Teilgebiet der Mathematik handelt, das durch Anwendungen immer wieder neue Facetten erhält. [...]"

- P. P. Spies:
  Grundlagen stochastischer Modelle,
  Hanser, München, 1982.
  ISBN: 3446137114

Literatur speziell zu Kapitel 4

- J. Banks:
  Principles of Quality Control,
  John Wiley and Sons, New York, 1989.
  ISBN: 0471635510

- D.C. Montgomery:
  Introduction to Statistical Quality Control,

2nd edition, John Wiley and Sons, New York, 1991.
ISBN: 0471656313
Kurzbeschreibung:
,,This book is about the use of modern statistical methods for quality control and improvement. It provides comprehensive coverage of the subject from basic principles to state-of-art concepts and applications. The objective is to give the reader a sound understanding of the principles and the basis for applying them in a variety of both product and nonproduct situations. While statistical techniques are emphasized throughout, the book has a strong engineering and management orientation. [...] By presenting theory, and supporting the theory with clear and relevant examples, Montgomery helps the reader to understand the big picture of important concepts. [...]"

- H. Rinne und H.–J. Mittag:
  Statistische Methoden der Qualitätssicherung,
  3. Auflage, Carl Hanser Verlag, München, 1995.
  ISBN: 3446180060
  Kurzbeschreibung:
  ,,Dieses Buch ist bewusst anwendungsorientiert geschrieben und zeichnet sich durch eine besonders sorgfältige didaktische Gestaltung aus. Es enthält: Zahlreiche Abbildungen und Fotos; mehr als 100 Übungsaufgaben mit ausführlichen Lösungen; viele durchgerechnete Anwendungsbeispiele; verständnisfördernde, zusammenfassende Übersichten; kommentierte Literaturangaben."

- W. Uhlmann:
  Statistische Qualitätskontrolle,
  2. Auflage, Teubner–Verlag, Stuttgart, 1982.
  ISBN: 3519123061
  Kurzbeschreibung:
  ,,Ein Lehrbuch für Statistiker, Mathematiker, Ingenieure und Wirtschaftswissenschaftler. Es enthält: Wahrscheinlichkeitstheoretische Grundlagen; statistische Grundlagen; Eingangs- und Endkontrolle; kostenoptimale Prüfpläne; sequentielle Tests; Kontrollkarten; Kosten und Kontrollabstand; kontinuierliche Stichprobenpläne."

Literatur speziell zu Kapitel 10

- F. Beichelt:
  Zuverlässigkeits– und Instandhaltungstheorie,
  Teubner, Stuttgart, 1993.
  ISBN: 3519029855
  Kurzbeschreibung:
  ,,Das Buch ist eine moderne Einführung in die Zuverlässigkeits– und Instandhaltungstheorie auf der Grundlage stochastischer Modellbildung.[...] Zahlreiche numerische Beispiele erleichtern dainhaltliche Verständnis. Das Buch wendet sich an Praktiker und Studierende mathematisch-naturwissenschaftlich-technischer Fachrichtungen. [...]"

Anhang D

Historie

In der folgenden Auflistung werden einige für die Entwicklung der Stochastik bedeutende Mathematiker mit ihren Lebensdaten und kurzen Beschreibungen ihrer Wirkungsfelder aufgeführt. Die Liste erhebt keinen Anspruch auf Vollständigkeit. Insbesondere fehlen wichtige Mathematiker, die nicht direkt im Bereich der Stochastik gewirkt haben, durch ihre Arbeiten zur Maß- oder Integrationstheorie die moderne Stochastik aber erst ermöglichten.

- **Thomas Bayes**
  (∗ 1702 in London, England; † 17. April 1761 in Tunbridge Wells, Kent, England)

  Thomas Bayes studierte ab 1719 Theologie an der Universität in Edinburgh und beschäftige sich nebenbei mit Mathematik. 1733 wurde er Pfarrer der presbyterianischen Kapelle in Tunbridge Wells, 35 Meilen südöstlich von London. 1742 ernannte man Bayes zum Mitglied der Royal Society, obwohl er bis zu diesem Zeitpunkt noch keinerlei mathematische Arbeiten veröffentlich hatte. Insgesamt publizierte Bayes selbst nur zwei Arbeiten. Seine wichtigten Forschungsergebnisse, die unter anderem auch den später als ,,Formel von Bayes" benannten Satz enthielten, wurden erst aus seinem Nachlass bekannt.

- **Familie Bernoulli**
  Die schweizer Gelehrtenfamilie Bernoulli hat über mehrere Generationen hinweg sehr große Beiträge zur Mathematik geleistet.

  Im Folgenden werden nur die beiden Mitglieder der Bernoulli-Familie aufgeführt, die sich wesentlich mit stochastischen Fragestellungen beschäftigt haben. Insbesondere werden Jakob Bernoulli II, Johann Bernoulli I, Johann Bernoulli II, Johann Bernoulli III, Niklaus Bernoulli I und Niklaus Bernoulli II, die allesamt bedeutende Mathematiker waren, hier nicht näher erwähnt.

  - **Daniel Bernoulli**
    (∗ 8. Februar 1700 in Groningen; † 17. März 1782 in Basel)

    Daniel Bernoulli interessierte sich hauptsächlich für Anwendungen der Mathematik. Er entwickelte das Prinzip zur Lösung algebraischer Gleichungen mit Hilfe von rekurrenten Reihen (,,Methode von Bernoulli") und untersuchte Kettenbrüche. Außerdem lieferte er wichtige Beiträge zur Wahrscheinlichkeitstheorie, die später teilweise von Laplace in seine Theorie aufgenommen wurden.

- **Jakob Bernoulli I**
  (∗ 27. Dezember 1654 in Basel; † 16. August 1705 in Basel)

  Jakob Bernoulli I ist der erste Gelehrte in der Familie der Bernoullis und überhaupt der erste bekannte Schweizer Mathematiker. Er befasste sich überwiegend mit analytischen Fragestellungen (er stand u.a. mit Leibniz, der gerade eine Infinitesimalmethoden aufgestellt hatte, in Kontakt), sowie mit stochastischen Problemen. Seine Arbeit baute auf den Ergebnissen von Huygens über das Glücksspiel auf. In einer, erst nach seinem Tode durch seinen Neffen Niklaus Bernoulli I veröffentlichten Arbeit, stellte Jakob Bernoulli I bereits das Gesetz der großen Zahlen auf und verallgemeinerte viele kombinatorische Ansätze von Huygens.

- **Emile Borel**
  (∗ 7. Januar 1871 in Saint-Affrique; † 3. Februar 1956 in Paris)

  Borel beschäftige sich zunächst mit Funktionentheorie. Nach seiner Tätigkeit als Forschungsbeirat im Kriegsministerium von 1914–1918 übernahm er den Lehrstuhl für Wahrscheinlichkeitsrechnung und mathematische Physik an der Sorbonne. Während seiner Arbeit in der Funktionentheorie prägte Borel den Begriff des Maßes und der überabzählbaren Überdeckung. Ab 1905 befasste sich Borel mit den Nutzungsmöglichkeiten der Maßtheorie in der Wahrscheinlichkeitstheorie. Außerdem ist Borel Mitbegründer der Spieltheorie und bewies das Minimax-Theorem für 3 Spieler.

- **Constantin Caratheodory**
  (∗ 13. September 1873 Berlin; † 2. Februar 1950 in München)

  Caratheodory stammt aus einer angesehenen griechischen Familie. Er arbeitete bis 1900 als Ingenieur im Dienste Englands an der Aufstauung des Nils mit und studierte im Anschluss daran in Deutschland u.a. bei Schwarz, Frobenius, Plank, Hilbert und Minkowski Mathematik. 1909 erhielt er seinen ersten Ruf als Professor an die TH Hannover. Im Folgenden lehrte er an verschiedenen deutschen und griechischen Universitäten. Caratheodory beschäftigte sich hauptsächlich mit Variationsrechnung, partiellen Differentialgleichungen, Maß- und Integrationstheorie sowie der Theorie reeller Funktionen. Die von ihm bewiesenen Maßfortsetzungssätze gehören heute zu den Kernelementen der Maßtheorie und die mittlerweile übliche axiomatische Einführung des Maßbegriffs geht in vielen Teilen auf ihn zurück.

- **Guido Fubini**
  (∗ 19. Januar 1879 in Venedig; † 6. Juni 1943 in New York)

  Zu den wichtigsten Arbeiten Fubinis gehört der 1907 von ihm bewiesene und später nach ihm benannte Satz. Darüber hinaus befasste sich Fubini mit projektiver Differentialgeometrie sowie der Theorie diskontinuierlicher Gruppen und automorpher Funktionen.

- **Andrej Nikolajewitsch Kolmogorov**
  (∗ 25. April 1903 in Tambow; 20. Oktober 1987 in Moskau)

  Kolmogorov gilt als einer der bedeutensten Mathematiker der Gegenwart. Er befasste sich vorwiegend mit Wahrscheinlichkeitstheorie, mathematischer Statistik und Logik, Maß- und Integrationstheorie, Funktionnalanalysis sowie Informations- und Algorithmentheorie. Nebenbei entwarf er Lehrpläne und Schulbücher für den Mathematikunterricht und prägte so zu großen Teilen den Mathematikunterricht in der Sowjetunion.

Mit seiner Arbeit ,,Grundbegriffe der Wahrscheinlichkeitsrechnung" aus dem Jahre 1933 löste er das 6. Problem der berühmten 23 von Hilbert gestellten mathematischen Probleme.

- **Pierre Simon Marquis de Laplace**
  (∗ 28. März 1749 in Beaumont-en-Auge; † 5. März 1827 in Paris)

  Laplace befasste sich vor allem mit partiellen Differential– und Differenzengleichungen. Seine Entwicklung der Laplace-Transformation diente ihm dazu, Naturerscheinungen analytisch zu erfassen. Neben Arbeiten zu physikalischen Themen befasste er sich auch mit Themen der Wahrscheinlichkeitsrechnung. Seine 1812 erschienene Theorie beinhaltete bereits den mathematischen Begriff der Wahrscheinlichkeit und den des Erwartungswertes. Zudem greift Laplace in seiner Arbeit das von J. Bernoulli gefundene Gesetz der großen Zahlen auf.

  Auf Laplace geht auch die Idee zurück, dass das Geschehen in einem physikaischen System exakt vorherbestimmbar sei, wenn nur alle Anfangszustände bekannt sind (,,Laplacescher Determinismus").

  Laplace war neben seiner Tätigkeit als Forscher ab 1794 Vorsitzender der Kommission für Maße und Gewichte und unter Napoleon Bonaparte Minister des Inneren.

- **Henri Lebesgue**
  (∗ 28. Juni 1875 in Beauvais (Frankreich); † 26. Juli 1941 in Paris)

  Lebesgue erkannte, dass viele zu seiner Zeit gültigen Theorien für eine Reihe von Fragestellungen unzureichend waren. 1902 verallgemeinerte er den Riemannschen Integralbegriff zu dem wesentlich leistungsfähigeren Lebesgueschen Integralkalkül. Lebesgues Resultate wurden zunächst nur zögernd aufgenommen, stellen heute aber die Grundlage für die moderne Analysis dar.

- **Andrej Andrejewitch Markov**
  (∗ 14. Juni 1856 in Gouvernement Rjasan; † 20. Juli 1922 in Petrograd)

  Markov studierte von 1874–1878 unter anderem bei Tschebyscheff und beschäftigte sich zunächst hauptsätlich mit Fragestellungen der Zahlen– und Funktionentheorie. Später befasste er sich überwiegend mit Wahrscheinlichkeitsrechnung. Dabei legte er wichtige Grundlagen zur Entwicklung der Theorie der stochastischen Prozesse. Außerdem entwickelte Markov die Theorie der später nach ihm benannten Markovschen Prozesse bzw. Ketten.

- **Pafnuti Lwowitch Tschebyscheff**
  (∗ 16. Mai 1821 in Okatowo; † 8. Dezember 1894 in Petersburg)

  Tschebyscheff befasste sich zunächst überwiegend mit Zahlentheorie. Unter anderem wirkte er an der Herausgabe der zahlentheoretischen Manusskripte Eulers mit. Später beschäftige er sich dann überwiegend mit wahrscheinlichkeitstheoretischen Fragestellungen. Insbesondere erarbeitete er die Gesetzmäßigkeiten von Summen unabhängiger Summanden. Er verdeutlichte die Wichtigkeit solcher Begriffe wie Zufallsgröße oder Erwartungswert, verallgemeinerte das Gesetz der großen Zahlen und vereinfachte dessen Beweis erheblich.

- **Bernhard Georg Friedrich Riemann**
  (∗ 17. September 1826 in Breselenz bei Dannenberg; † 20. Juli 1866 in Selasca in Italien)

  Riemann studierte ab 1846 an der Universität in Göttingen zunächst Theologie und Philosophie, wechselte dann aber bald zur Mathematik. In seiner 1854 vorgelegten Habilitationsschrift führte Riemann das später nach ihm benannte Riemann-Integral ein. Neben der Integrationstheorie befasste er sich mit vielen weiteren mathematischen Gebieten. So forschte Riemann u.a. auf dem Gebiet der partiellen Differentialgleichungen, sowie in der Zahlentheorie und der nichteuklidischen Geometrie. Die Ideen Riemanns sind bis heute von großer Bedeutung: Die Riemannsche Hypothese über die Nullstellen der ζ-Funktion wird in sehr vielen Sätzen der Zahlentheorie verwendet. Beweisen konnte man die Riemannsche Hypothese allerdings bis heute nicht.

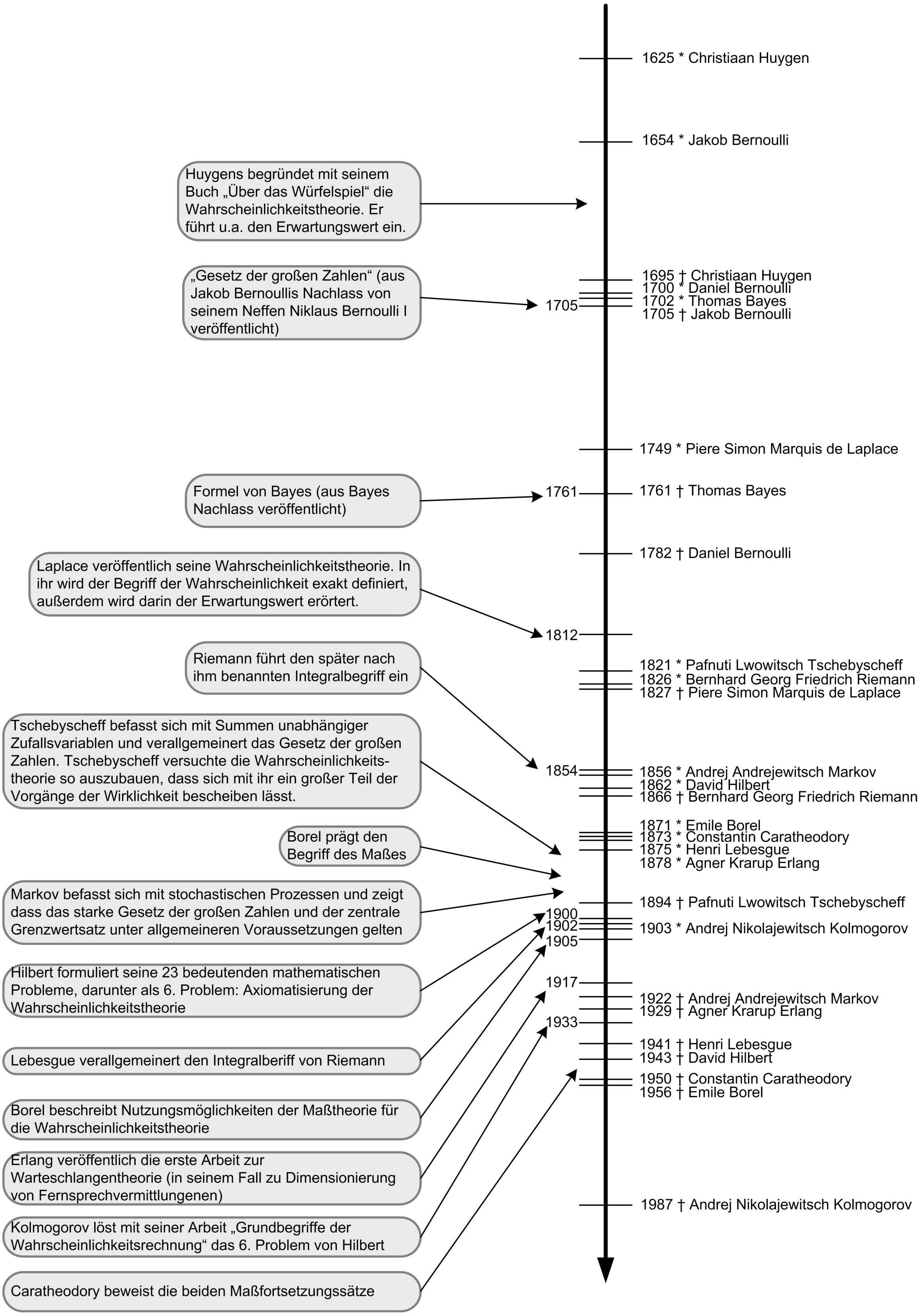
Weitere Informationen und diverse Biographien finden sich unter:

- S. Gottwald, H.-J. Ilgauds, K.-H. Schlote:
  Lexikon bedeutender Mathematiker,
  Verlag Harri Deutsch, Thun, 1990.
  ISBN: 3-8171-1164-9

- Turnbell Server, Biographies
  http://www-groups.dcs.st-and.ac.uk/

- Mathematik.ch: Bedeutende Mathematiker
  http://www.mathematik.ch/mathematiker/

- Wikipedia (Kategorie: Mathematiker)
  http://de.wikipedia.org/wiki/Kategorie:Mathematiker


1625 * Christiaan Huygen
1654 * Jakob Bernoulli
Huygens begründet mit seinem Buch „Über das Würfelspiel“ die Wahrscheinlichkeitstheorie. Er führt u.a. den Erwartungswert ein.
„Gesetz der großen Zahlen“ (aus Jakob Bernoullis Nachlass von seinem Neffen Niklaus Bernoulli I veröffentlicht)
1705
1695 † Christiaan Huygen
1700 * Daniel Bernoulli
1702 * Thomas Bayes
1705 † Jakob Bernoulli
1749 * Piere Simon Marquis de Laplace
Formel von Bayes (aus Bayes Nachlass veröffentlicht)
1761
1761 † Thomas Bayes
1782 † Daniel Bernoulli
Laplace veröffentlich seine Wahrscheinlichkeitstheorie. In ihr wird der Begriff der Wahrscheinlichkeit exakt definiert, außerdem wird darin der Erwartungswert erörtert.
1812
Riemann führt den später nach ihm benannten Integralbegriff ein
1821 * Pafnuti Lwowitsch Tschebyscheff
1826 * Bernhard Georg Friedrich Riemann
1827 † Piere Simon Marquis de Laplace
Tschebyscheff befasst sich mit Summen unabhängiger Zufallsvariablen und verallgemeinert das Gesetz der großen Zahlen. Tschebyscheff versuchte die Wahrscheinlichkeits-theorie so auszubauen, dass sich mit ihr ein großer Teil der Vorgänge der Wirklichkeit bescheiben lässt.
1854
1856 * Andrej Andrejewitsch Markov
1862 * David Hilbert
1866 † Bernhard Georg Friedrich Riemann
Borel prägt den Begriff des Maßes
1871 * Emile Borel
1873 * Constantin Caratheodory
1875 * Henri Lebesgue
1878 * Agner Krarup Erlang
Markov befasst sich mit stochastischen Prozessen und zeigt dass das starke Gesetz der großen Zahlen und der zentrale Grenzwertsatz unter allgemeineren Voraussetzungen gelten
1894 † Pafnuti Lwowitsch Tschebyscheff
1900
1902
1905
1903 * Andrej Nikolajewitsch Kolmogorov
Hilbert formuliert seine 23 bedeutenden mathematischen Probleme, darunter als 6. Problem: Axiomatisierung der Wahrscheinlichkeitstheorie
1917
1922 † Andrej Andrejewitsch Markov
1929 † Agner Krarup Erlang
1933
Lebesgue verallgemeinert den Integralberiff von Riemann
1941 † Henri Lebesgue
1943 † David Hilbert
1950 † Constantin Caratheodory
1956 † Emile Borel
Borel beschreibt Nutzungsmöglichkeiten der Maßtheorie für die Wahrscheinlichkeitstheorie
Erlang veröffentlich die erste Arbeit zur Warteschlangentheorie (in seinem Fall zu Dimensionierung von Fernsprechvermittlungenen)
1987 † Andrej Nikolajewitsch Kolmogorov
Kolmogorov löst mit seiner Arbeit „Grundbegriffe der Wahrscheinlichkeitsrechnung“ das 6. Problem von Hilbert
Caratheodory beweist die beiden Maßfortsetzungssätze

Stichwortverzeichnis

$(n-c)$-Stichprobenplan, 60
μ^*–messbar, 99
σ–Algebra der Borelschen Mengen, 85
σ–endlich, 101
Überlebensfunktion, 174
äußeres Maß, 98

a–posteriori–Verteilung, 48
a–priori–Verteilung, 48
additiv, 90
Algebra, 79
  Produkt–σ–, 187
Alternativ–Hypothese, 61
Annahmezahl, 60
AOQ, 70
AOQL, 70
Attribut–Prüfung, 60
Ausfallrate, 174
Axiomensystem, 16

Bayes, Thomas, 243
bedingte Überlebenswahrscheinlichkeit, 175
bedingte Restlebensdauer, 175
Bernoulli, Daniel, 243
Bernoulli, Jakob I, 244
Betafunktion, 121
Betaverteilung 1. Art, 121
Bild, 128
Bildmaß, 33, 128
Binomialverteilung, 28, 30
  Erwartungswert, 37
  Varianz, 37
Borel, Emile, 244
Borel–Maß, 103
Borelsche Mengen, 85–86

Caratheodory, Constantin, 244
Cauchy–Verteilung, 117, 118
χ^2–Verteilung, 120

Decreasing Failure Rate, 177
Dichte, 112, 193
Dirac–Maß, 91
diskret, 13
Durchschlupf
  maximaler mittlerer, 70
  mittlerer, 70
durchschnittsstabil, 76
Dynkin–System, 81

Einfachstichprobenplan, 60
Einschränkung, 94
Elementarereignisse, 13
Elementarfunktion, 137
endlich, 90
Ereignis, 13
  Eigenschaften, 14
  Rechenregeln, 13
Ergebnismenge, 13
Ergebnisraum, 13
Erlang–Verteilung, 120
Erwartungswert, 136–149
  diskret, 35
Erweiterung, 94
erzeugende Funktion, 38
erzeugenden System, 83
Erzeugendensystem, 83
Erzeuger, 83
erzeugte σ-Algebra, 198
Exponentialverteilung, 113
  Varianz, 164

fast–überall, 168
fast–sicher, 223
Formel von Bayes, 48
Formel von Sylvester–Poincaré, 16
Fortsetzung, 94
Fortsetzungssatz
  erster, 94

Fubini, Guido, 244
Funktionen
    numerische, 129

Günther, Verfahren von, 63
Gütefunktion, 61
Gammaverteilung, 118
Gaußsche Glockenkurve, 115
Gleichverteilung, 23, 112
Gut–Schlecht–Prüfung, 60

Häufigkeit
    absolute, 15
    bedingte relative, 45
    relative, 15
hypergeometrische Verteilung, 26

i.i.d. Folge, 221
identisch verteilt, 221
Increasing Failure Rate, 177
Indifferenzpunkt, 68
Indikatorvariable, 127
induziertes äußeres Maß, 98
Inhalt, 90
integrabel, 137, 145
Integral, 138, 168
    bestimmte, 138, 143
    Eigenschaften, 143
integrierbar, 137, 145, 146

Kolmogorov, Andrej Nikolajewitsch, 244
Konvergenz
    stochastisch, 220
Korrespondenzsatz, 109
Kovarianz, 216

Laplace, Pierre Simon Marquis de, 245
Laplaceverteilung, 23
Lebensdauer, 174
Lebensdauerfunktion, 174
Lebesgue, Henri, 245
Lebesgue-Maß, 111
linksschief, 165
Logarithmische Normalverteilung, 117
logarithmische Normalverteilung, 117

Mächtigkeit, 24
Maß, 90
    Produkt-, 206
maßdefinierende Funktion, 105, 106, 191
Markov, Andrej Andrejewitch, 245
maximaler mittlerer Durchschlupf, 70
Mengenfunktionen, 89–103
Mengensysteme, 75–86
messbar, 127, 129–132
messbare Funktionen, 132
Messraum, 127
mittlere Lebensdauer, 174
mittlerer Durchschlupf, 70
Moment, k-tes, 36
Momente, 165
    k–te, 165
    k–te absolute, 165
    k–te zentrale, 165
    k–te zentrale absolute, 165
Multiplikationssatz, 52

ncStichprobenplan, 63
Negativ–Teil, 145
nichtnegativ, 90
Normaldarstellung, 137
Normalverteilung, 115
n–te untere Faktorielle, 24
Null–Hypothese, 61
Nullmenge, 167

Operationscharakteristik, 62, 63
Operationspfade, 181
operationstreu, 126

paarweise stochastisch unabhängig, 55
Poissonverteilung, 31
    Erwartungswert, 38
    Varianz, 38
Positiv–Teil, 145
Prämaß, 90
Produkt–σ–Algebra, 187
Produktmaß, 206

quasiintegrabel, 145

Randverteilung, 192
Randverteilungsdichte, 197
Randverteilungsfunktion, 192
Rechteckverteilung, 112
rechtsschief, 165

Restriktion, 94
Riemann, Bernhard Georg Friedrich, 246
Ring, 79, 84

Satz
  der monotonen Konvergenz, 169
  der totalen Wahrscheinlichkeit, 47
  Faltungs-, 208
  Transformationssatz für Dichten, 207
Satz von Bernoulli, 222
Schiefe, 165
Schnitt, 212
schnittstabil, 76
Semiring, 76, 84
separabel, 84
σ–additiv, 90
σ–Algebra, 14, 79
σ–Algebra, 81, 127
σ–subadditiv, 90
Spur–σ–Algebra, 45
Standard–Normalverteilung, 115
Standardabweichung, 165
Standardnormalverteilung, 115
statistische Tests, 60
Steilheit, 68
Stetigkeit von oben, 93
Stetigkeit von unten, 93
Stichprobenmenge, 13
Stichprobenplan
  $(n-c)$, 67
Stichprobensystem der Firma Philips, 68
Stochastik, 12
stochastisch unabhängig, 54, 198
Streuung, 165
subadditiv, 90

Test zum Signifikanzniveau α, 62
Tschebyscheff, Pafnuti Lwowitch, 245

unabhängig
  paarweise stochastisch, 55
  stochastisch, 54
  vollständig, 56
Ungleichung, Kolmogorovsche, 225
Ungleichung, Markovsche, 220
Ungleichung, Tschebyscheffsche, 221
unkorreliert, 216
Urbild, 126
Urnenmodell, 23–28
  von Polya, 53
  Ziehen mit Zurücklegen, 27
  Ziehen ohne Zurücklegen, 24

Varianz, 36, 164
Varianz der Lebensdauer, 174
Variationskoeffizienten, 165
vereinigungsstabil, 76
Verfahren von Günther, 63
Verteilung, 33
Verteilungsfunktion, 105, 106, 191, 197
  mehrdimensionale, 188
vollständig stochastisch unabhängig, 56

Wahrscheinlichkeit, 15
  bedingte, 45
  Satz von der totalen, 47
Wahrscheinlichkeitsmaß, 16, 90, 106, 128
Wahrscheinlichkeitsraum, 16
  diskret, 22
  Laplace'scher, 23
Wahrscheinlichkeitsvektor, 22
Wahrscheinlichkeitsverteilung
  Beta, 121
  binomiale, 28–30
  Cauchy, 117
  χ^2, 120
  Erlang, 120
  Exponential, 113
    Varianz, 164
  Gamma, 118
  Gleichverteilung, 23, 112
  hypergeometrische, 26, 29
  Laplace'sche, 23
  Log. Normal, 117
  Poisson, 31
  Rechteck, 112
  Standardnormal, 115
  Weibull, 115
Weibull–Verteilung, 115

zentrales Moment, 36
Zufallsexperiment, 12
Zufallsgröße, 128
  reelle, 128
Zufallsvariable, 33, 128
  diskret, 33

diskret reellwertig, 33
Zufallsvektor, 128
Zuverlässigkeitsfunktion, 174